UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Interactive Evolutionary Generative Design in Augmented Reality

Permalink
https://escholarship.org/uc/item/6zm3q6zb

Author
Bailey, Elnaz Tafrihi

Publication Date
2024

Supplemental Material
https://escholarship.org/uc/item/6zm3q6zb#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Interactive Evolutionary Generative Design in Augmented Reality

by

Elnaz Bailey

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Architecture

and the Designated Emphasis

in

New Media

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Luisa Caldas, Chair
Associate Professor Simon Schleicher
Associate Professor Bjoern Hartmann

Spring 2024

Abstract

Interactive Evolutionary Generative Design in Augmented Reality

by

Elnaz Bailey

Doctor of Philosophy in Architecture

Designated Emphasis in New Media

University of California, Berkeley

Professor Luisa Caldas, Chair

This dissertation explores immersive technologies in the architectural domain, aiming to answer an increased demand for new remote design platforms that promote interactive collaboration. It explores the existing space of architectural design processes, along with the design applications of genetic algorithms and participatory design. In response to existing gaps, this dissertation introduces InsightXR, an Augmented Reality (AR) platform that facilitates collaboration during the architectural design process and provides individualized feedback to designers while exploring generative design. The use case of InsightXR presented in this research focuses on developing massing studies during early stages of architectural design using a new application of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to the field of Operative Design. As a method and a design platform, however, InsightXR can be applied to many other design problems and disciplinary domains.

The proposed workflow for massing studies involves utilizing a site model to generate a population using the Operative Generative NSGA-II algorithm. During each generation, users are presented with the three most distinct geometries, based on shape similarity calculated by mesh similarity methods, and asked to provide a qualitative fitness value to be used in the proposed algorithm. To understand the benefits and limitations of this new interactive generative design process, three experiments are developed in this dissertation. The first experiment presents a novel Operative Generative NSGA-II using algorithm experiments by investigating Floor Area Ratio (FAR), Non-Passive Zone (NPZ), Roofs and Best Oriented Surfaces (RBOS), and Usable Open Space (UOS) as objective functions. The second experiment explores the integration of the Operative Generative NSGA-II with InsightXR's application using RBOS, UOS, and Area Violation (AV) as objective functions. Finally, the third experiment is designed in two stages. In

the first stage, a human subject experiment is performed with experts and non-experts in architectural design using AV, NPZ, and UOS as objective functions. In the second stage, a follow-up experiment uses only the algorithm without including users, utilizing the first stage initial populations as seeds. The resulting building metrics and genetic algorithm performance are analyzed across the various experiments, exploring the benefits of the platform as well as looking at the effects on the results from Expert versus Non-Expert participants and Human-Directed versus Algorithm-Only executions.

In Experiment 3, both user groups' feedback conflicted with the other objective functions, demonstrating the relevance of introducing user participation in generative design processes. This was due to the tendency of the algorithm to generate massing options with smaller footprints and taller typologies to improve UOS and NPZ values, which most users described as their least preferred building type. A high percentage of both experts and non-experts rated NPZ as the most important objective function, demonstrating a valorization of design aspects that contribute to climate change mitigation. The NPZ values had a higher impact on experts compared to non-experts, signaling the importance that experts currently attribute to those performance criteria. Comparative results between the two user groups also highlighted the relevance of expertise in the generative design process. While providing complex spatial information to experts helped improve their decision-making process, in contrast, it did not significantly improve the decision-making process for non-experts.

As an overall conclusion, while the AI system alone could often generate better-performing solutions in a quantitative-only optimization space, it tended to create solutions that were qualitatively unsatisfactory to many users. The absence of user preference in generative design processes has arguably been a major obstacle for the adoption of generative design in real world situations. Introducing user preference ratings as an additional objective function often generated a relative reduction in some quantitative metrics, but it simultaneously increased the likelihood that solutions were qualitatively appreciated by users, and thus more likely to be adopted. Furthermore, when expert users of our AR-driven participatory design interface were asked to rate alternative designs purely based on aesthetic preferences, results were typically inferior to when users were additionally given some simplified quantitative information to inform their rating preferences. This suggests that the successful integration of informed user preference with AI-based support for complex multi-dimensional design tasks may prove a successful and promising avenue for future design methods research.

To my husband,

my mother, my father, and my sister

who supported me throughout this journey

Table of Contents

LIST OF ACRONYMS ..... **IV**

ACKNOWLEDGMENTS ..... **V**

CHAPTER 1: INTRODUCTION **1**

1.1 PROJECT INTRODUCTION 1

CHAPTER 2: LITERATURE REVIEW **5**

2.1 GENERATIVE DESIGN 5
2.1.1 Architectural Design Process 5
2.1.2 Computational Design in Architecture 9
2.1.3 Generative Design in Architecture 11
2.1.4 Operative Design 11
2.1.4 Optimization 14
2.1.5 Role of Humans in Generative Design and Optimization 22
2.2 PARTICIPATORY DESIGN IN ARCHITECTURE 28
2.2.1 Early-Stage Architectural Design 28
2.2.2 Design Review Process in Architecture 28
2.2.3 Participatory Design in Architecture 31
2.2.4 User Feedback in Crowdsourcing 38
2.2.5 Participatory and Collaborative Interactive Evolutionary Generative Design 40
2.3 USER ASSESSMENT IN VIRTUAL AND AUGMENTED REALITY 42
2.3.1 Augmented and Virtual Reality in Architectural Design 42
2.3.2 The Role of User Feedback Collection in AR and VR 47
2.3.3 Physiological Assessment Methods 50
2.3.4 User Feedback and Participatory Design 51
2.3.5 Eye Tracking 53
2.3.6 Limitations 54
2.3.7 Eye tracking in AR and VR 55
2.3.8 Eye Tracking Metrics 56
2.3.9 AR and VR Feedback Collection and Visualization in Architecture 58

CHAPTER 3: METHODS **63**

3.1 INSIGHTXR EARLY-STAGE AR DESIGN 64
3.2 INSIGHTXR EARLY-STAGE VR PROTOTYPE 67
3.3 INSIGHTXR FINAL DESIGN 69
3.4 RESULTS VISUALIZATION 73

CHAPTER 4: EXPERIMENTS **74**

4.1 EXPERIMENT 1: OPERATIVE GENERATIVE DESIGN USING NON-DOMINATED SORTING GENETIC ALGORITHM II (NSGA-II) 74
4.1.1 Introduction 74
4.1.2 Method 75
4.1.3 Results 88
4.1.4 Discussion 99
4.1.5 Experiment Conclusions 106
4.2 EXPERIMENT 2: INSIGHTXR: OPERATIVE GENERATIVE DESIGN IN AUGMENTED REALITY 107
4.2.1 Introduction 107
4.2.2 Methods 108
4.2.3 Results 111
4.2.4 Discussion 120
4.2.5 Experiment Conclusions 121
4.3 EXPERIMENT 3: HUMAN SUBJECT EXPERIMENTS 122
4.3.1 Introduction 122
4.3.2 Experiment 1 126
4.3.3 Experiment 2: Algorithm Experiments Without Users (NSGA-II Alone) 135
4.3.4 Results 136
4.3.5 Discussion 171
4.3.6 Experiment Conclusions 178
CHAPTER 5: CONCLUSION AND FUTURE WORK 181
5.1 DISCUSSION 181
5.2 CONCLUSIONS 184
5.3 FUTURE WORK 187
BIBLIOGRAPHY 189
APPENDIX A. HIGHLY RATED SOLUTIONS BY EXPERTS AND NON-EXPERTS. 210
A.1 EXPERTS HIGHLY RATED SOLUTIONS 210
A.2 NON-EXPERT HIGHLY RATED SOLUTIONS 236

List of Acronyms

AR	Augmented Reality
AV	Area Violation
CD	Computational Design
CSSL	Co-training Algorithm for Semi-Supervised Learning
FAR	Floor Area Ratio
GA	Genetic Algorithm
HV	Hypervolume
IGA	Interactive Genetic Algorithm
MOO	Multi-objective Optimization
NPZ	Non-Passive Zone
NSGA-II	Non-Dominated Sorting Genetic Algorithm-II
PSO	Particle Swarm Optimization
RBFN	Radial Basis Function Network
RBOS	Roofs and Best Oriented Surfaces
SA	Simulated Annealing
SGA	Standard Genetic Algorithm
SHGC	Solar Heat Gain Coefficient
SP	Space Planning
UOS	Usable Open Space
VR	Virtual Reality
VT	Visual Transmittance
WWR	Window-to-Wall Ratio

Acknowledgments

Studying at UC Berkeley was a dream come true. This journey exposed me to different points of view and ingrained in me the importance of freedom of speech, advocating for people, and designing for people as an architect. I am forever grateful for having the opportunity to meet many brilliant students and faculty members during my PhD at Berkeley.

This journey would not have been possible without the continuous support of my PhD advisor, Professor Luisa Caldas. She enabled me to be curious and innovative and allowed me to keep exploring throughout my PhD, which I'm deeply grateful for. She always provided me with the tools necessary to succeed in my research. I also had the opportunity to work with Professor Caldas on the Virtual Bauer Wurster Project during the COVID-19 pandemic, which gave me the inspiration and drive to continue my PhD.

I am grateful to Professor Simon Schleicher for his continuous engagement and support of my research. Professor Schleicher has always sparked my excitement in architectural design and digital fabrication, and I'm forever grateful for the opportunity to meet and work with him.

During my PhD, I first met Professor Björn Hartmann at the Jacobs Institute for Design Innovation. Professor Hartmann later became a member of my PhD committee, for which I'm very grateful. He provided continuous mentorship during my research, from ideation to the final steps.

A special thanks to Professor Eric Paulos for his support and for inspiring me during my studies at UC Berkeley, and to Professor Coye Cheshire for helping me ground my visions in my research and mentoring me during my PhD.

The generous support from the University of California Berkeley's Chief Technology Officer (CTO) enabled me to continue working on my research as a Berkeley Changemaker Technology Innovation Grant recipient. An XLab grant during the final stages of my PhD made the human subject experiments possible, and I am deeply appreciative of that valuable support. I was fortunate to meet the fantastic faculty and staff at the Berkeley Center for New Media (BCNM), and I am very grateful for their numerous grants and fellowships, which enabled my PhD research to come to fruition.

My peers and PhD colleagues at the XR Lab, including Haripriya Sathyanarayanan, Yunting Zhao, Xin Zhou, Xinwei Zhuang, Botond Bognar, and Luis Santos, provided fundamental support to make this journey possible. Priya has been an amazing friend and has always been there for me throughout my PhD. A special thanks to Xin, for his friendship and support since the early days of InsightXR.

Writing this dissertation became a reality thanks to weekly writing sessions with Sabrina Soracco and Haripriya Sathyanarayanan where we mutually received and offered critical feedback, and I'm deeply thankful for the supportive environment of this group.

Finally, I would like to thank my family. My husband Phil has been there since the beginning of my PhD. He helped me be brave in order to embark on this journey. He created a loving home throughout this process, even with my long hours of research. My mother and father have been my biggest supporters for as long as I can remember. They believed in me, helped me move to the US to continue my education, and endured all the hardships of this process. My sister, Azar, was always my teacher and a great example of perseverance. I am grateful for all their sacrifices that made my PhD journey possible.

Chapter 1: Introduction

1.1 Project Introduction

Immersive remote collaboration has been widely explored in academic environments and within the workplace [192], [175]. Unprecedented times during the COVID-19 pandemic rapidly increased the demand for remote collaboration, altering architectural practice.

Design teaching in architecture is typically problem-based, where architecture students participate in a search for solutions under the instructor's guidance [137]. In addition, desk crits and design reviews play an important part in the architecture students' learning process. Desk crits are where students confront challenges and gain design knowledge, and critiquing is how architects (and architectural design students) receive feedback for their designs [205], [172]. Due to the pandemic, all studio experiences within American architecture schools and architecture firms had to transition to remote.

The increased demand for remote design and collaboration, as well as the potential of immersive technologies in the architectural design domain, are the motivations behind this dissertation.

In response to a gap in the state-of-the-art for decentralized and remote architectural design, this dissertation introduces InsightXR, an Augmented Reality (AR) platform that allows users to collaborate in the design process and provide individualized feedback while exploring generative design. InsightXR focuses on massing studies during the early stages of architectural design using a new application of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) for the field of Operative Design [209]. This research proposes a novel interactive generative design process leveraging users' feedback to improve early-stage massing design in the field of architecture.

User feedback, be it from peers, stakeholders, or end users, plays a vital role in the architectural design process and education, and finding ways to improve this process while empowering users is an essential step toward an engaging and inclusive design. Implementing generative design in a participatory approach can facilitate the introduction of novel design ideas and creative work that could not have been conceived of by the designers or design reviewers in a standard design process. Integrating immersive technology such as augmented and virtual reality into a user's participation can help their immersion into the design and better promote understanding of a project's scale and overall geometry.

In addition to the benefits of interacting with 3D designs remotely, the proposed approach empowers experts (designers) and non-experts (non-designers) by allowing the designer to develop the initial designs and asking the user to steer the optimization process to reach a desired design. In this framework, both the process of participation and the resulting outcome are of equal importance as the designers will have access to the generated database of different design outcomes produced using InsightXR [47]. Based on the data about the users' preferences, they can decide on a final design with full access to the users' thought process.

Generative design integrates a semi-autonomous approach to exploring design spaces allowing a greater freedom to the design in a shorter amount of time [258], and provides the solution to the designer for further evaluation and decision-making [164]. The two different approaches for involving users in a generative system are as follows: 1. A fully automated design process with the user's involvement only at the beginning and end, and 2. A system that solicits user inputs intermittently during the generative process [105]. Interactive generative design is defined as a generative system that allows for user involvement at the beginning, during, and at the end of the generative design process. Such a system benefits from human understanding in decision-making, and their expertise in different fields of design [25].

Participation in the design process can help a community reach democratic decisions regarding designs in their community, improve the quality of proposed designs and avoid early design mistakes by inclusion of people who are affected by those decisions [213]. Designers usually don't seek feedback until they have already completed an initial design and associated visualization, and the inclusion of user's feedback is traditionally ambiguous with no direct results [199]. Community participants in the design process would be able to share their unique wisdom and experiences about the built environment and help designers reach improved decisions throughout the different stages of design. Users in the community can feel empowered by this participation, and they can feel ownership and control of their built environment [17]. In addition to including non-experts, receiving feedback from peers and experts in the field of design can help with improving design solutions.

The proposed research aims to address the following objectives:

1. Propose a novel Operative Generative NSGA-II algorithm for early-stage architectural design.
2. Allow the user's qualitative feedback to guide generative design.
3. Design an AR platform that enables user participation in early-stage architectural design processes.

The methods and experiments proposed in this research will try to answer the following overall question:

- How can we propose a novel interactive generative design process leveraging users' feedback to improve early-stage massing design in the field of architecture?

Below are additional research questions in each subject category:

1. Generative design

A. How can we guide a generative design process leveraging user feedback?

B. How can an interactive algorithm be implemented into the generative design process?

C. How can architects' massing development design steps be translated into generative design?

2. Participatory design in architecture

A. What are the missing features in existing design tools that architects use for remote design review?

B. How does the use of augmented reality impact collaboration and user engagement?

C. What are the differences between the feedback provided to the designer from experts (i.e., instructors/peer reviewers) and non-experts (i.e., non-designers)?

3. User Experience Assessment in AR

A. How can we integrate user assessment into the 3D design process? How does that impact the optimization process?

B. How can we utilize existing AR technology to facilitate user and peer participation in design?

The final design of InsightXR uses a Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to generate design alternatives, using Operative Design for architects to start from during early massing studies [13]. The Operative Design process is based on the volumetric operations discussed by Mari and Yoo and uses an operative NSGA-II algorithm [148], [13]. The subset of design operations actually implemented for the GA focus on displacement and subtraction by using expansion and carving operations as the main verbs applied to their proposed operative generative design [148], [13].

In InsightXR, users are divided into two categories: experts and non-experts. The application prompts users to choose between designer (expert) and reviewer (non-expert) roles. If the designer role is selected, a user can choose between the following three alternatives:

1) Create a new generative design project; 2) Participate in an existing generative design project as a reviewer; 3) Provide feedback on single massing projects.

Users can also participate as reviewers in the generative design process, which will categorize them as non-experts. The main difference between designer and reviewer in this process is the ability of designers to upload a model to receive feedback on 3D designs and create a new generative design project.

To test and understand the benefits and limitations of this proposed interactive generative design process, three experiments are conducted, as described in the Experiments chapter. The first experiment introduces a novel Operative Generative NSGA-II using algorithm experiments by investigating Floor Area Ratio (FAR), Non-Passive Zone (NPZ), Roofs and Best Oriented Surfaces (RBOS), and Usable Open Space (UOS) as objective functions. The second experiment introduces the integration of the Operative Generative NSGA-II with InsightXR's application using RBOS, UOS, and AV as objective functions. Finally, a human subject experiment is

conducted with experts and non-experts in architectural design using AV, NPZ, UOS, and User Rating as objective functions. A follow-up algorithm-based experiment is discussed in this dissertation to allow a comparison of the algorithm with and without the user's impacts on the generative design process. The results of these experiments will be discussed in chapter 4, under Experiments.

This dissertation includes two previously published papers in chapter 4. The first publication is incorporated in chapter 4.1, and the second publication is incorporated in chapter 4.2. The titles of the publications are as follows:

- Elnaz Tafrihi Bailey and Luisa Caldas. 2023. Operative generative design using non-dominated sorting genetic algorithm II (NSGA-II). Automation in Construction 155: 105026. https://doi.org/10.1016/j.autcon.2023.105026
- Elnaz Bailey and Luisa Caldas. 2023. InsightXR: Operative Generative Design in Augmented Reality. In Generative Art 2023.

The work included in this dissertation was developed with support from several grants, including the UC Berkeley's Changemaker Technology Innovation Grant (2020), Amazon Web Services (AWS) Cloud Credits for Research (2022), Xlab Research Grant (2024), the Berkeley Center for New Media (BCNM) Summer Research Award (2023), UC Berkeley's Doctoral Completion Fellowship (2022), and BCNM's Undergraduate Research Fellowship (2022). The initial results of this research have been presented during the Fall 2020 Technology Innovation Forum at UC Berkeley.

Chapter 2: Literature Review

2.1 Generative Design

2.1.1 Architectural Design Process

The Royal Institute of British Architects (RIBA) defines the architectural design process in different countries, including the US and the UK [259]. As discussed by RIBA, some countries include a Pre-Design stage as an architect's initial design stage, whereas many countries, such as the US, don't have a Pre-design stage. Architects' services in the US start from the Schematic Design (SD) stage (SD is equivalent to the Concept Design (CD) stage in the UK), where a client approaches them with a detailed program, a budget, and a site [260]. This process can differ somewhat depending on the project type. Clients might work with architects in a Pre-design stage for programming and feasibility analysis. In this paradigm, the architect and owner will work together to understand the potential for a buildable area on the land, potential program distribution on the site, and preliminary site analyses. The architects might even come up with a rough representation of what can be placed on the site (massings).

During the pre-design phase, decisions such as sustainable design goals are also discussed with the team and stakeholders (e.g., LEED accreditation goals). An example of a comprehensive certification for buildings in sustainability is the Living Building Certification [252]. Zero Energy Certification discusses the requirements for the entire building's energy to be generated by on-site renewable energy, which is judged based on actual and non-simulated results [253]. California also has required performance goals for energy consumption that need to be achieved under Title 24, Part 11, California Green Building Standards Code (CALGreen) [261]. Design teams are required to submit documentation regarding these California code goals. In addition to energy consumption, structural plans must also be submitted for review to receive a permit for construction. These requirements are minimum requirements and are less strict than goals such as zero-net goals [261].

Regardless of access to these studies, architects during the SD phase (after the Pre-design phase) dedicate time to further confirm the feasibility of the requested programs on the site and come up with their analysis of the site and massing design options. The final results from the schematic design phase are initial architectural drawings: floor plans, sections, elevations, structural drawings, HVAC drawings, and landscape design [131]. SD is a crucial stage in the architectural design process since many decisions made during early SD, such as overall massing design strategies, can be set in stone throughout the project, as making fundamental modifications to the project after this stage can be challenging. In RIBA's description, some core tasks during the SD phase include creating proposals that align with site details and the project brief and defining goals that include spatial requirements [259]. Design reviews are used to discuss the design progress with the stakeholders, often leading to design brief changes during this process. Design proposals will be adjusted based on expert discussions with consultants such as structural engineers and mechanical engineers involved in the project. The project budgeting evolves during this process, which is an important factor in the design process. This stage sets the architectural concept for a project going forward. In the US, some project types require the submission of massing design and

shading studies to better understand the relationship of the proposed building with its surroundings. Many examples of generative design that we will discuss later are done during the SD phase, mainly later SD, which will exemplify the importance of this stage in the design process.

Design Development (DD) is the stage after SD, in which more detailed drawings, specifications, and information regarding materials along with structural, electrical, and mechanical systems in the project are produced [131]. Even though materials and other building details are defined in the DD phase, many optimization processes make assumptions about these materials by assigning the minimum requirements, such as ASHRAE 90.1 [262], to achieve energy-efficient buildings and run their energy analysis. These studies usually compare design options, which can be done during the SD phase.

The next phase of the design is Construction Documents (CD), where the architects prepare finalized and detailed drawings, specifications, and the final cost estimates of the project [131]. The subsequent phases are bidding (depending on the project) and Construction Administration (CA), where the architect works on approval of shop drawings and potential change orders, with visits to the construction site to make sure the general contractor is following the drawings. Although the American Institute of Architects (AIA) considers these 5 phases as the overall phases of design, depending on the project's complexity, some projects can be done in more phases. In some phases, schematic design may be divided into the conceptual design phase and schematic design. The schematic design phase is usually done by a small group of designers, engineers, and other experts. A conventional project's SD phase usually takes 1 to 2 months, while a complex project can take up to 12 months.

In the educational setting, architecture studios follow similar design steps. These steps of design will be introduced based on a studio offered at the UC Berkeley College of Environmental Design (CED) titled Integrated Studio [106]. This is the fifth studio in the 3-year program of the graduate architecture program, which resembles professional practice the most since it covers designing an entire building by integrating structural and lighting elements and applying daylighting strategies to the design [106]. The students are expected to produce architectural drawings such as floor plans and sections. To understand the studio design process, we had a discussion with Professor Hirschman, one of the instructors for the Integrated Studio at CED. In this studio, the students are given a detailed program to design a recreation center in Oakland, California. The program resembles one that a client provided to the architects in a professional practice. In the studio, students can combine and overlap the programs within the building [131]. The design phases for this class are divided into the following, which are discussed in Table 2.1.

Table 2.1. Design phases in integrated studio [106].

Design Phase	**Duration**
Pre-Design: Site documentation, analysis, and model; Research Focus Massing Studies; Programming Studies; Precedent Collection/Analysis/Massing.	One month
Schematic Design	One month
Documentation Design Development	One month and a half

During the Pre-design phase, there are multiple desk crits, followed by an internal review at the end of this phase. The massing studies take two weeks since the students are more experienced at this stage of their education [106]. Other design phases explored in this studio include schematic design, design development, and documentation [106]. Since the massing exercises are conducted in the pre-design phase in this studio, the next steps will focus on refining the design and details of the program and construction. In the first exercise, the students were asked to create a studio-wide model, and each student group created several initial models that fit into the studio's site model [106].



Figure 2.1. Group crit on initial "scrap massing models" exercise. ARCH 203: Integrated Design Studio, Fall 2021. Instructor: Sarah Hirschman. Department of Architecture, UC Berkeley [106].

This exercise was their first attempt to propose a massing on the site. Next, during the massing studies exercise, the students were asked to investigate three of their models further by applying the following rules to these designs: 1. Addition: created by combining multiple solids; 2. Subtraction: where the massing is created using removal from a solid; 3. Accumulation: surface combinations [106]. During the crits for the massing exercise, the students were asked to select a massing option that they found intriguing, and they commented on the design chosen [106]. The designer was then asked to provide any concepts behind their massing that they would like to share with the class [106].



Figure 2.2. Massing exercises results: Row 1: Three initial ideas; Row 2: Modified massings using Addition; Row 3: Modified massings using Subtraction, 4th row: Modified massings using Accumulation. From ARCH 203: Integrated Design Studio, Fall 2021. By Jun Tanabe and Maggie Johnson, modified by author [106].

The importance of massing in the design process was discussed in Ferris's 1922 New York Times article [263], where they discussed new zoning laws in New York City and what that entailed regarding the city's future image. Ferris talked about the impact of these new regulations on city open spaces and aspects such as allowable building areas, use cases, and heights in different areas [263]. In this article, Ferris proposed steps and a representation of what a massing study would look like under these regulations, mainly using a subtractive methodology [263]. The starting point was populating the land with proposed spaces in 3D, and then parts of the mass were subtracted to create "light courts" [263]. Next, angled planes were introduced that created structural forms and cut a series of setbacks into the massing to reach the maximum volume allowed [263]. This stage introduced limitations to the tower component of the design for its height and volume. The last step considered more structural aspects of the design, ready to be articulated by an individual designer [263]. This is an early example of how different aspects of the site context and design elements must be considered together in proposing massing strategies in architectural design using a subtraction method.

Massing studies by Ferris and the increased attention to zoning regulations became important topics for the architectural design process [142]. Architects must consider details such as Setbacks, plot area ratio, maximum building height, and other essential attributes that the architect must handle to fulfill the law [142]. At the early design stage, architects design a volume that becomes a reference for applying zoning requirements [142]. This representation should respond to the site itself, its context, the program (defined by the client), and the overall aesthetic goals of the design team and the client. Finding this representative volume is a complex task, and it is crucial as it solidifies the overall building geometry during the early stages of design. In this stage, called the massing study, many theoretical possibilities of shapes, sizes, and volumes in the plot are studied [142].

In California, the Building Code [264] proposes limitations on height and area. Per these regulations, the height above grade depends on the occupancy type and type of construction [264]. The allowable area can also be determined using the building code, where occupancy classification and construction type impact the allowable area. In addition to the building codes, architects need to refer to the zoning code based on the site's location, which might limit these numbers further. This stage, as discussed in the educational setting, includes physical prototyping and digital drawings, including diagrams, to help the team with decision-making. At this point, there is a prime opportunity to include stakeholders and peers to receive feedback and collaboratively work on improved design solutions. Generative design at these early stages can help designers find novel solutions to different objectives.

2.1.2 Computational Design in Architecture

The earliest introductions of computation to the design field were associated with the design methods movements in the 1960s, where researchers envisioned ambitious computer roles in the design process [116]. During these movements, objectivity was the basis for design, marking the 1960s as the “design science decade” [116].

The acceptance of computers' role in the design process changed over the next decades as the design research community moved away from them [116]. Early computers were designed to focus on novel methods of working, but the focus later turned into tools in the design process. Computers, as only tools, had no role in the creative process and were only used to support a designer’s manual tasks [116]. In more recent years, the role of computers has evolved to be viewed as a design support medium [116]. The idea of computers as support mediums versus tools was discussed in Janssen’s dissertation based on Schmitt’s perspective on computers' role [116]. In Schmitt’s writing, as a mere tool, computers would eliminate human activities due to their high efficiency, while as a support medium, the computer acts as an interactive “counterpart” [191].

An early example of computation in architecture, as discussed by Caetano et al. [27], was developed in Armour and Buffa’s [8] methodology from 1963 in the design of optimized facility locations. They developed a program that searched the “relative location patterns” and found the optimum solution by producing facility layout diagrams [8]. A pivotal point in the use of computers in design was Ivan Sutherland’s Sketchpad project [220], a graphical communication method that allows the user to set mathematical constraints to a drawing created from a sketch

[220]. Early work in computational design in architecture dates back to the 1970s, with publications including Eastman's Spatial Synthesis in Computer-Aided Building Design, Applied Science [76]. Other influential publications in architecture were Mitchel's "Computer-Aided Architectural Design" [156], which was published in 1977, and his paper called "The theoretical foundation of computer-aided architectural design" in 1975 [155]. Computer-aided design (CAD) became available to designers in the 1980s, with the first commercial release of AutoCAD in 1982, which facilitated the integration of performance objectives into design [27].

In the last few decades, computational design has gone far beyond drafting tools by integrating topics such as building simulation and fabrication methods combined with evolutionary algorithms into the design process [27]. Different common approaches to computational design include parametric and generative design. An important aspect when discussing computational design, and specifically the generative design approach, is how it slots into the design process.

In 1994, Mitchell [157] discussed his point of view towards the application of computers in design. Mitchell explained that the first paradigm in computer-aided design is designing to problem-solving, which dates back to the 1960s when design problems could be similar to other intellectual problems [157]. Such problems can be solved by specifying a "domain of possible solutions" [157]. Design research in this context aims to establish good problem formulations for programming the generator and find computationally efficient strategies [157]. Mitchell further discussed that from this perspective, design problems are either "well-defined" or "ill-defined" [157]. When dealing with well-defined problems, there is no ambiguity about the boundaries of the possible solutions. In such a well-defined problem, once a solution is found, it is accepted as the solution [157]. However, many problems might be "fuzzy" or "ill-defined", where a designer may be unsure if they considered all possible design options [157]. Based on Mitchell, "well-defined computationally easy" problems are the best candidates for an automated process [157]. Mitchell suggested separating the design process so that the designer gets the ill-defined problem, and the machine works on the well-defined part [157]. An example of breaking down a problem into ill-defined and well-defined is seen in Ivan Sutherland's Sketchpad project [8]. In 1995, Maher et al. [146] also discussed how, in the design process, both the search for the solution and the goals are ill-defined. During design explorations, both the goals and the design options evolve together, which also needs to be reflected in the optimization process [146].

Generative design in architecture has been mainly applied in two categories: 1. performance-based architectural design in areas such as building performance optimization, including daylighting, energy performance, as well as structural optimization, and 2. space planning (SP). Many examples of generative design in performance-based architecture that will be discussed are closer to the well-defined spectrum of problem-solving, where the interest is on optimizing the building using energy and daylighting simulation. Such approaches are best fit for the later stages of design (later schematic design process). In contrast, the early design stages, including pre-design and early schematic design, are associated with ill-defined problem-solving. To understand this concept better, we will discuss the architectural design process in further detail and discuss examples of generative design and how they fit within each design stage.

2.1.3 Generative Design in Architecture

Mitchell, in "The Theoretical Foundation of Computer-aided Architectural Design" [155], published in 1975, already included a section on generative systems and problem-solving in design. Mitchell characterized problem-solving as "a process of searching through alternative states of the representation to discover a state that meets certain specified criteria" [155]. Mitchell also introduced an early example of generative systems based on Aristotle's discussion of the designs for potential cities using a generative system [155].

Generative design uses algorithms to produce a solution space as design options based on goals (objectives) and constraints, which designers can predefine. Computers are used to run these algorithms because they can perform numerically complex problems and then visualize the results for humans to evaluate them and make final decisions [105]. The integration of generative systems into the design process allows for the creation of novel ideas that were not possible using other design methods. An example of generative design in architecture was introduced by Digital Blue Foam (DBF) [265]. This web-based tool allows for generative design as well as analysis of space using DBF's Artificial Intelligence (AI) based engine [265]. DBF creates urban scenarios for comparison and research [265].

Regarding the generation of three-dimensional forms in design, Janssen [116], in their dissertation, describes three overall approaches:

1. Parametric approach: Forms are created using parameters.
2. The combinatorial approach: Forms are created by combining components.
3. The substitution approach: Forms are generated by a repetitive substitution of components. There might be a few rules, but solutions can be unpredictable:
    a. Two subgroups: 1. "Grid-based substitution techniques: Cellular automata"; 2. "Shape-based substitution techniques: Fractals, Shape grammars, L-Systems" [116].

As explained earlier, generative design has primarily been applied in performance-based and space layout uses. Performance-based architectural design has been used in building performance optimization, including daylighting, energy performance, and structural optimization.

2.1.4 Operative Design

Massing in the architectural design process needs to respond to the site's context, the program, and the overall aesthetic goals of the design team and the client. Finding this representative volume is a complex task, and it is crucial as it solidifies the overall building geometry during the early stages of design [142]. At this stage, many theoretical possibilities of shapes, sizes, and volumes in a proposed site are studied [142]. Building codes can propose limitations on height and area, where the height above grade can depend on the occupancy type and type of construction [264]. The allowable area can also be determined using the building codes, where occupancy classification and construction type impact the allowable area [264]. In addition to the building codes, architects

need to refer to the zoning code based on the site's location, which might limit these numbers further. This stage includes physical prototyping, digital drawings, and diagrams to help the team make decisions.

In spatial design, surfaces are a fundamental component of contemporary architecture [148]. Surfaces can be modified using different operations, including wrapping, folding, and weaving [148]. Surface operations in architecture have been discussed in publications, including Folding Architecture: Spatial, Structural, and Organizational Architecture and Supersurfaces: Generating Forms for Architecture, Products and Fashion [148], [234], [235].

Using operative design in volumes is another method to start the massing design process [148]. In this dissertation, the volumetric operations are inspired by Mari and Yoo's collection of operative terms and examples, where design operations represent an evolution of volumes and can result in different design programs [148].

These operations can be based on different actions: Add, Displace, or Subtract, which can be further characterized by the following verbs [148]:

- Add: expand, extrude, inflate, branch, merge, nest, offset.
- Displace: bend, skew, split, twist, interlock, intersect, lift, lodge, overlap, rotate, shift.
- Subtract: carve, compress, fracture, grade, notch, pinch, shear, taper, embed, extract, inscribe, puncture.

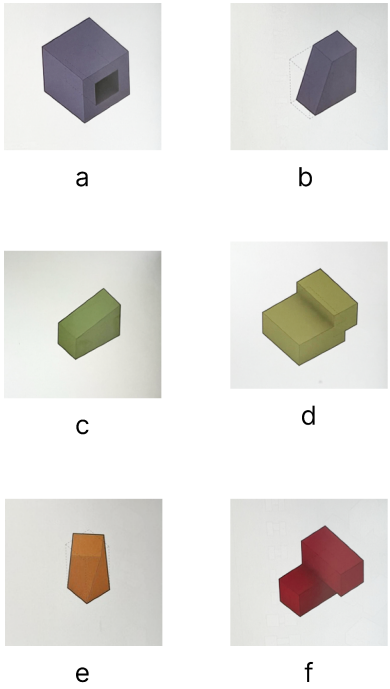


Figure 2.3. Examples of Actions: Row 1: Subtraction (a) Carving, (b) Shear; Row 2: Addition: Expansion (c) and Merge (d); Row 3: Displacement: Twist (e) and Overlap (f). Figures by Mari and Yoo, modified by author [148].

Figure 2.4 shows some of the early explorations of these volumetric operations. The massings proposed in this research can be the results of one operation or a combination of operations that will be called hybrids [148]. The same operations can be applied multiple times to a massing [148].

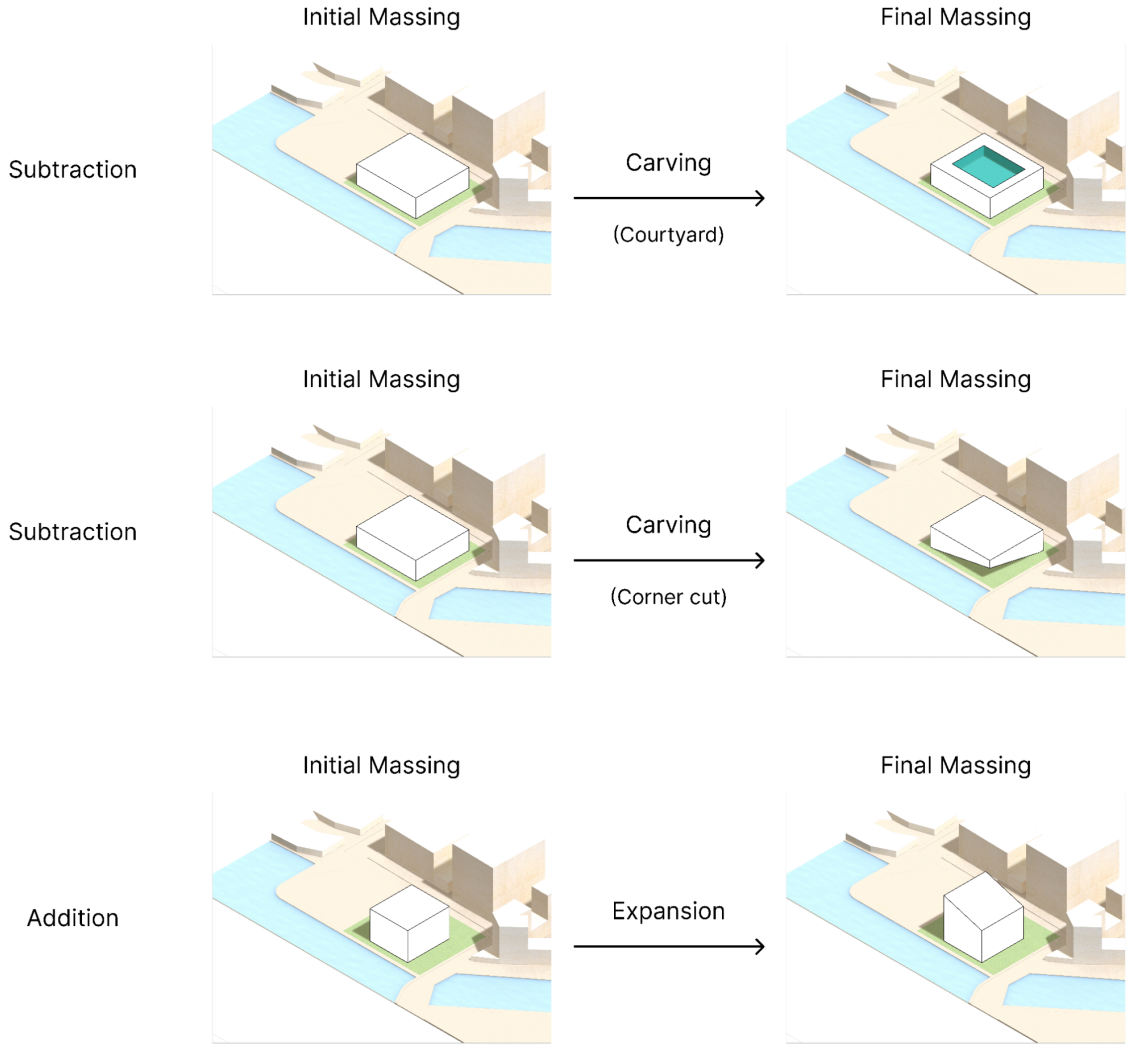


Figure 2.4. Design operations explorations in Rhinoceros based on Mari and Yoo [148]: Row 1: Carving (courtyard); Row 2: Carving (corner cut); Row 3: Expansion.

Operative design can be found in different architectural projects, including House N by Sou Fujimoto Architects (nest), The Twist Museum by Bjarke Ingels Group (twist), VIΛ 57 West by Bjarke Ingels Group (expand+carve), and Turning Torso by Santiago Calatrava Architects and Engineers (rotate+overlap). Figure 2.5 shows some of these examples.



(a) (b) (c)

Figure 2.5. Operative design examples: (a) House N by Sou Fujimoto Architects in Oita, Japan (Photo by Amy Hay Mew Hwang, Flickr CC, edited by author) is an example of nesting as an operative design move. (b) The Twist Museum by Bjarke Ingels Group in Jevnaker, Norway (Photo by Laurian Ghinitoiu) is an example of twisting as an operative design move. (c) Turning Torso by Santiago Calatrava Architects and Engineers in Malmö, Sweden (Photo by Mirko Junge) is an example of rotating and overlapping as operative design moves.

In parametric design, Osintseva et al. proposed using design operations to generate 3D design through parametric design in Grasshopper and allowing users to impact the resulting geometries [173]. They used residential buildings as a typology in their research. The proposed actions were: 1. Open block, 2. Setbacks edges, 3. Setbacks corners, 4. Reduce height, 5. Rooftop terraces, 6. Towers, 7. Yard buildings, and 8. Random breaks [173]. These actions were developed as a Grasshopper component and included manipulations that could be applied once or combined. Users could interact with the design process by changing parameters or placing attractor points or lines that applied the actions in certain parts of the geometry [173].

2.1.4 Optimization

Optimization is an integrated process in complex design decision-making [144]. Using optimization methods, users can identify variables associated with their problems and find desired solutions [144]. The objectives used in the optimization process can be quantitative or qualitative, and the variables can be subject to constraints or limits [144]. The aim of optimization is usually to minimize (or maximize) the defined function of optimization and can be mathematically described as [237]:

Minimize (or maximize):

$$F(x_1, x_2, \ldots, x_n)$$

Where:

$$G(x_1, x_2, \ldots, x_n) \geq 0 \;\; x_i \in S_i \quad [237]$$

In this definition, F represents the objective function of the optimization, G is the constraint function, and the variables x_i can have a constraint such that they belong to the set S_i [237]. Depending on their definition, optimizations can be categorized as constrained or unconstrained [237]. In the optimization process, the objectives can be single or multiple, which introduces the

topic of multi-objective optimizations, and the objectives can be a combination of quantitative and qualitative objectives [237]. The goal of single-objective optimization is to find a global minimum solution. In contrast, multi-objective optimizations introduce the topic of the Pareto front, where solutions are considered Pareto optimal if they are not dominated by other solutions based on the defined variables [237]. Pareto optimality can be defined as if a point x in the solution space exists where no other points are lower in value in at least one objective compared to that point [150]. To solve optimization problems, heuristics are proposed methods that provide solutions close to the global minimum in hard problems [82]. Heuristics do not guarantee a true optimal solution but have a high probability of reaching a close to optimal solution [82]. Heuristics are based on the optimization problem, so their development is complex [174]. Metaheuristics solves the issue of heuristics with the ability to adapt to add "problem-specific knowledge" into the algorithms [174].

When discussing generative design in performance-based architecture in the literature, the designers try to find solutions that help minimize or maximize objectives such as energy use, including heating, cooling, and lighting; reaching desired daylight factor levels; or optimizing structural components.

Most examples provided in architecture's generative design literature regarding performance-based design are associated with later steps of the SD phase, as they explore variables such as Window-to-Wall ratio (WWR) (which means the massing has been already defined) or early stages of heating, ventilating, air conditioning (HVAC) design. HVAC design usually involves some knowledge of floor plans and spatial design developed after the massing studies. Daylighting analysis and optimization require access to the massing design and involve decisions, including glazing systems; therefore, they happen after the massing has been sufficiently defined in the SD phase.

That being said, in practice, designers and engineers can make initial models with general knowledge about the use case of spaces based on the detailed program provided by the owner and the areas for each function in the building. These initial models are roughly estimated before the overall design is finalized. As discussed before, they also use material specifications based on standards as the starting point for these models. Such studies might happen during the early SD phase. In further steps of the design, such as DD, the design is more set in stone, so the benefit of such optimization processes is significantly less, as an optimized recommendation is likely not realizable. Setting high-performance goals early in the design phase allows for finding more cost-effective solutions early [179].

In structural design, optimization can be applied for objectives such as the sizing of structural elements or shape and topology optimization [182]. Design variables associated with the sizing of structural elements such as trusses include sizing variables (e.g., Cross-section area) related to the finite-element model [182]. Shape optimization of structural elements can be done using methods such as changing the parameters associated with the overall boundary form of the structure [182]. Other methods related to structural optimization include topology optimizations, where the goal is to find an optimal layout of the structural elements [182].

Space planning (SP) research in architecture explores the placement of spatial elements while working with certain constraints, such as adjacencies and distance between spaces [33]. In SP,

architects try to find creative solutions that best fit their objectives and the design narratives [33]. The roots of space planning optimization are tied closely to Sutherland's Sketchpad and the evolution of CAD [33]. An early publication discussing space planning optimization and CAD was by Krejcirik in 1969, focusing on industrial buildings [134]. SP planning optimization was further evolved during the 1970s and 80s with publications by researchers such as Mitchell and Liggett [33]. Liggett described the evaluation methods used in automated SP as belonging to three major categories: 1: Single objectives, such as costs; 2: Graphs for theoretical approaches to solving adjacency problems in SP optimizations where “nodes” are the activities and the “edges” are the adjacency requirements; 3: The feasibility and usefulness of the found solution such that it can respond to all constraints and required relationships in the design [141].

In space planning optimizations, one of the major focuses is on the adjacencies for the floor plan design, which is a later stage SD concern in the design process after the massing design. An example of such optimizations was explored in Project Discover [164], where they focused on only arrangements within the building boundaries (using the building boundaries as a constraint). When applied to a new construction, this constraint means that the building’s form was already decided, which places this in a later stage SD phase in addition to the focus on floor plans in later stages of design. Another approach is considering the adjacencies, which start as an abstract diagram in an architectural design. This method can be placed in the pre-design phase during program studies without a final decision about the form. This method of space planning has been explored in projects such as “Evolving Floor Plans” by Joel Simon [200], where “speculative” plans were created using the rooms and circulation as inputs without concerns for practical aspects such as constructability [200]. This project uses an ant colony optimization method.

Optimization Methods

Based on a literature review by Nguyen et al. [169] on simulation-based optimization methods applied in building performance optimizations, some commonly used search and optimization categories include direct search families and different metaheuristics and hybrid methods. Based on our review of SP examples in the field of architecture, the algorithms applied mainly fall within these categories as well, and these methods will be discussed in further detail in this section. The table below shows some of these methods' main weaknesses and strengths, as discussed by Nguyen et al. [169].

Table 2.2. Classification of some of the most-used algorithms [169].

Optimization	Advantage	Disadvantage	Sample Algorithms	Design Phase Applied
Direct search: Gradient-deterministic search	Fast convergence: a stationary point can be guaranteed	Sensitive to discontinuities in the objective function Sensitive to multi-modal function	Bounded BFGS, Levenberg Marquardt algorithm, Discrete Armijo Gradient algorithm, CONLIN method [169]	Later stage Schematic Design

Direct search: Gradient-free search	Derivative-free Useful despite discontinuities in the objective function	Inclination towards local minima. Might not be able to give the global minimum. Issues with "non-smooth functions".	Hooke–Jeeves algorithms, Simplex algorithms [169]	Later stage Schematic Design
Metaheuristics: Stochastic population-based	Less likely to get trapped in a local minimum	A large number of objective function evaluations. Global minimum might not be achieved.	Evolutionary algorithms: GA, Genetic programming, Evolutionary programming Particle swarm optimization (PSO), Ant colony algorithm [169]	Pre-design and early Schematic design, Late-stage Schematic Design, Early Design Development
Metaheuristics: Trajectory search	Best for discrete optimization problems (continuous variables can also be used), such as traveling salesman problems (TSP)	Effective in discrete search spaces. Can't identify if the solution found is optimal. Repeated annealing [169].	Simulated annealing, Tabu search [169]	Pre-design and early Schematic design, Later stage Schematic Design, Early Design Development
Hybrid	Hybrid methods combine the advantages and disadvantages of the methods above, aiming to solve the disadvantages.		PSO-HJ, GA-GPS [169]	Later stage Schematic Design

Note: Table is created based on Nguyen et al. [169], with modifications by author.

With the emergence of visual programming languages such as Grasshopper [168] in Rhino and Dynamo in Revit [266], there are various methods for architects to access optimization during their design process without prior programming experience. An example of such a plugin for Rhino is Galapagos. This optimization plugin can run only single objective optimizations and includes two different solvers, the standard genetic algorithm (SGA) and the simulated annealing (SA) algorithm [187]. Octopus is an advanced plugin that uses the SPEA-2 Algorithm to perform multi-objective optimization, and the user can modify the optimization while running, which is different from how other optimization methods work in Grasshopper [233]. Goat is another Grasshopper plugin designed based on Galapagos and is based on gradient-free optimization [267]. In Dynamo and Revit, Project Refinery can be used for generative design [268]. Project Refinery uses the NSGA-II optimization method, and the solutions can be visualized as 3D geometries and a series of charts or tables [269].

Metaheuristics: Stochastic population-based family

To solve optimization problems, heuristics are methods that provide solutions close to the global minimum in otherwise computationally hard problems [82]. Heuristics do not guarantee a true optimal solution but an efficient way with a high possibility of achieving near an optimal solution

[174]. Heuristics are based on the optimization problem's specifics; therefore, their development is a complex process [174]. Metaheuristics solves the issue of heuristics with the ability to adapt to add "problem-specific knowledge" into an otherwise generalizable algorithm [174]. Metaheuristics can "perform robust searches" in the solution space and include procedures to escape local optimality or to build or destroy solutions [89].

Stochastic population-based algorithms generate random populations and improve these populations of candidate solutions [169]. Based on a review by Nguyen et al., common stochastic populations-based algorithms for building performance optimization included genetic algorithms (GA), particle swarm optimization (PSO), Hybrid algorithms, and evolutionary algorithms. These methods do not guarantee the best solution but rather good solutions in less time [169].

1. Evolutionary Algorithms

Evolutionary algorithms are common metaheuristic algorithms that apply the concept of Darwin's "survival of the fittest" in a population where the poorest individuals get eliminated from each generation [82]. These algorithms apply operators including "mutation (random changes) and cross over (switching elements from different solutions)" to create new individuals [82]. Evolutionary algorithms can be divided into three approaches: evolutionary strategies (ES) introduced by Rechenberg [183], evolutionary programming, and genetic algorithms [174]. Due to the relevance of genetic algorithms in generative design in architecture, they will be discussed in further detail.

2. Genetic Algorithms (GA)

Genetic algorithms (GA) are techniques used as the backbone of evolutionary algorithms [174]. As discussed by Goldberg [90], the first theories discussing an artificial genetic search belonged to John Holland in 1975 [107]. Goldberg [90] describes genetic algorithms as search mechanisms inspired by natural selection processes, where the survival of the "fittest" individual is combined with "randomized" search. Genetic algorithms work with generations of individuals. New generations are built using individuals created from the components of acceptably high-performing individuals from previous generations mixed with a small population of completely new individuals. In GAs, a solution is an individual, and a group of solutions in each generation is considered a population [31]. Every new population is considered a generation, and in each generation, the fitness of individuals relates to the "value of the objective function at that specific point" [31]. The basic operators in genetic algorithms are "reproduction, crossover, and mutation". Fitness value is a defining factor in selecting an individual for reproduction [31]. A genetic algorithm generates a random group of solutions and then evaluates them based on an objective function [31]. In the next step, the mentioned genetic operators are applied to the population in "correlation" to their fitness value [31]. This process creates a new generation of individuals, which can continue until the specified number of generations defined by the user are achieved [31].

Different variations of GA exist that added improvements to these algorithms. One of the variations of GA is microGA [31]. MicroGA is a small-population GA first introduced by Goldberg [90], where the algorithm starts with a randomly generated small population and then applies the genetic operators until we have a "nominal convergence" [31]. Then, the new

population is generated by using the best individuals from the converged population (elitism), and the rest will be generated randomly [45]. An advantage of microGAs compared to GAs is that the algorithm can perform local searches around the best solutions during the generations prior to convergence, which allows for finding local minima [31]. MicroGAs and GAs' ability to reach reasonable solutions efficiently can be suitable in an early design stage where the design objectives are still fuzzy and changing.

When discussing GAs, it is important to discuss the condition of the objectives to optimize. Single-objective optimizations can cause limitations to designers, as design decisions usually include consideration of tradeoffs between multiple objectives simultaneously [29]. Therefore, multi-objective optimizations have been introduced. In GAs, different approaches to multi-objective optimizations are: “Simple aggregation” or “weighting” of objectives, Population-based “non-Pareto” approaches, or “Pareto-based approaches” [29].

3. NSGA and NSGA-II

Solutions to multi-objective optimizations do not result in a single global solution, where it is rather important to identify a group of solutions that fit a definition of optimal solution [150]. In order to evolve a population, there are different approaches to optimize the trade-offs, including the Non-dominated Sorting Genetic Algorithm (NSGA), the Pareto-optimal selection method, tournament selection, and Pareto optimality of an individual with assigning the same ranking to all non-dominated individuals or with a ranking assigned to “equal the number of individuals by which it is dominated” [29]. Due to their importance in building performance optimization, NSGA, NSGA-II, and SPEA II will be discussed in further detail.

NSGA was first introduced by Srinivas and Deb [209] in 1994. NSGA is a variation of GA used in multi-objective optimizations with two main types, including NSGA and NSGA-II [124]. NSGA implements a non-dominated sorting method to emphasize the good solutions [209], combined with niching, which enables the GA to search around more than one neighbor and reduces “genetic drift” [124]. The niche method helps maintain stable subpopulations of well-performing solutions [209]. Crossover and mutation remain the same as the simple GA. Within the same population, all nondominated individuals are assigned a “dummy” fitness value, an equal value for all, to allow an equal chance for reproduction [209]. There are certain criticisms towards NSGA, as discussed by Deb et al. [56] in 2002, including: “computational complexity of non-dominated sorting, lack of elitism, and the requirement to assign a sharing parameter in the algorithm” [56]. They introduced NSGA-II as an alternative to resolve these issues.

4. Examples of Applications of GAs in Generative Design in Architecture

Among computational optimization methods applied to building performance, the most common method used is genetic algorithms based on a literature review by Evins [82]. Other popular methods are the direct search method and the particle swarm optimization [82]. The characteristics of GAs that make them popular in building performance optimization include the ability for parallel simulations, robustness in working with discontinuity in solution spaces, and population-based algorithmic options that allow for solving multi-objective optimizations [169].

Some early examples of the use of performance as an objective for generative design are Caldas' work in 1999 [28] that combined, for the first time, a Genetic Algorithm with DOE2.1E (the building energy simulation software developed at Lawrence Berkeley National Laboratory that was one of the first in the world to provide whole building energy simulation, including passive and active systems). Caldas' work later evolved into incorporating EnergyPlus [270], the current LBNL building energy simulation widely used in the industry, into GENE_ARCH [29], [30]. GENE_ARCH has been integrated in later research in building performance, including research by Wang et al. in 2016 [236]. They created a multivariate regression model that was used as input into GENE_ARCH in order to introduce a new optimization approach for selecting important window properties including Visual Transmittance (VT), Solar Heat Gain Coefficient (SHGC), and U-factor [236].

In 2002, Caldas and Norford [31] used microGA to optimize design solutions based on thermal and lighting simulations in DOE2.1E. They used a small population with only five individuals in a hundred generations [31]. They applied their method to an office building with defined thermal zones, where window sizing is used as a variable to optimize its lighting, heating, and cooling performance [31]. The application of a GA in this example relates to later stages of SD and early DD since the office building has defined zone dimensions and material properties, as well as considerations made about glazing. These steps are after the definition of massing strategies in SD. However, their use of GA with a small population can be an efficient approach to conduct optimizations at different stages of design.

In 2005, Coello and Pulido [45] applied microGAs to multi-objective structural optimization problems using different forms of elitism, including keeping an external population to save the non-dominated solutions. They applied their proposed method to structural optimization problems such as "optimizing a two-bar plane truss" and compared their method with NSGA-II [45]. Based on their results, their proposed microGA solution outperformed the NSGA-II method [45]. Structural optimization problems focused on further details of a proposed design usually apply to the later stages of the SD phase, leveraging a better understanding of a space's overall volume and form.

In 2006, Duarte et al. combined shape grammar and GAs in an urban scale, to create a system that can apply existing urban fabrics in creating novel contemporary urban developments [73]. This research used Zaouiat Lakhdar quarter in Marrakech Medina as the case study [73].

Evins et al. [83] applied multi-objective optimization to minimize the heating and cooling load in the design of a double-skin facade. They tested their method on three cases: a hotel room, an office, and a meeting room [83]. They used the NSGA-II algorithm with a population size of 20, a max generation number of 20, and a crossover probability of 0.9 with a mutation probability of 0.7 [83]. The variables discussed in this project are fan position, window open/closed conditions, control temperatures, and glazing layers [83]. Such decisions in the design process regarding HVAC systems and materials are related to the later SD phase and early DD phase in design.

In 2016, Caldas and Santos introduced Painting with light, a user-guided interactive evolutionary system in Rhinoceros that allows users to specify daylight levels and patterns spatially [32]. Painting with light utilizes brushing and linking methods to find solutions that respond to the

required daylight performance [32]. This research resulted in crucial findings in user involvement limitations, where defining unfeasible daylighting goals resulted in inefficient search methods for the evolutionary algorithm [32]. This result further shows the possibility of inconsistency and conflict between user's preferences and performance metrics defined for an optimization problem [32].

In 2017, Nagy et al. [164] discussed Project Discover, which explores building performance optimization combined with space planning. Their methodology was used in designing Autodesk's new office space, utilizing a multi-objective genetic algorithm (MOGA). The constraints in the optimization process included the building boundary, which was an existing building. Other constraints include program requirements and the occupancy [164]. They defined design metrics to use for the optimization process based on conversations between the designers and the users, and the results include adjacency preferences, workstyle preferences, the distribution of high-activity zones, productivity, daylighting metrics, and unobstructed views within the workspace [164].

Santos et al. in 2018 coupled Radiance with a genetic algorithm, in order to assess the potentials of less computationally expensive daylight metrics such as daylight factor (DF) to optimize Light-Redirecting Fenestration Systems (LRFS) including Lasy_S and LightLouver [190]. Their results show that metrics such as DF are unsuccessful in optimizing for glare, which requires detailed glare simulations for improved results [190].

Ciardiello, in 2020, applied a multi-objective approach using genetic algorithms to optimize the shape and the building envelope to minimize the energy consumption of their proposed building [42]. They applied their method to the design of a case study building where their optimization variables were the overall shape of the building and the WWR. They modeled the 32 variations of the massing building for the overall shape and used that in Energy Plus during optimization [42]. The analysis of the overall massing of the building in the optimization process is related to the early-stage SD phase. For the energy simulation process, they had to make assumptions regarding building materials related to the DD phase. This project also considers future retrofits of the projects, with variables including glazing materials, walls, and roof insulation, which is more related to the DD phase of the project.

In 2020, Wortmann and Natanian [240] explored the use of multi-objective optimization in the design of a zero-energy neighborhood, using different building typologies, including courtyard, slab, and tower. Their study compares HypE and NSGA-II with Radial Basis Function Multi-objective Optimization (RBFMOpt) [240]. HypE [10] is a hypervolume-based multi-objective evolutionary algorithm introduced in 2011 to tackle the issue of an increase in the number of objectives that NSGA-II might have difficulty with [10]. This project explored the trade-offs between urban density energy consumption caused by different typologies of neighborhoods. Based on their results, out of 41 design solutions on the Pareto front, 33 employ the courtyard typologies number 1 and 2, four applied typology number 4 (slabs), and four typologies number 6-9 [240]. HypE and NSGA-II combined with an RBF can be applied early in architectural design during SD or Pre-design, where conflicting subjective and objective criteria are commonly considered.

2.1.5 Role of Humans in Generative Design and Optimization

One view towards generative design was proposed by McCormack et al. in 2004, where generative design can be defined as finding “novel and complex” solutions using simple inputs [152]. There are two methods to find these novel solutions in a generative system: 1. A fully automated design process with the user's involvement only at the beginning and end, and 2. Allowing for user involvement intermittently during the generative process [105]. Mueller and Brown [25] discuss the benefits of applying large data in different fields where human understanding is important in decision-making. Providing possible solutions to humans can help them with future decision-making, which can be used as a guiding tool in fields such as design [25].

In the 1960s, Nigel Cross conducted research using two groups; one group consisted of architects, and another group replicated an intelligent computer to test early ideas of CAD systems [50]. Based on the design project the architects were given to design, they were able to write questions on the computer, and the other group of designers who were the computers then had to try and answer the questions as fast as possible [50]. This experiment was designed to help understand designers' needs when working with a CAD system. Cross described the designer's role in this process as stress-free, where the computer can do tasks such as drawings but with human guidance [50].

There are different ways to view the designer's role in the generative design process. McCormack et al. in 2004 [152] discussed how the designer plays a role in exploring the traditional design process, where the designer’s goals and the final product have a direct relationship. They discuss that by using generative design systems, designers can lose control over the generated designs using generative design systems due to their role as mainly setting the objectives and variables [152]. The unexpected outcomes of generative design can challenge a designer's decision-making [152]. In a generative system, the designer's creativity emerges through their control over the design parameters, the solution space, and the final design solution [152]. Despite this description, McCormack considers the role of humans (designers) crucial to the design process, as they are to the conventional design process [152].

Generative design processes can be fully automated or involve humans [105]. An example of an automated design process is “Emergent Design” by Testa et al. in 2001, where their software finds “emergent spatial systems” [226]. Using algorithms such as evolutionary algorithms allows for an automated search by a computer where the user's involvement can be mainly in setting the constraints and variables, running the optimization, and then making the final decisions on the presented solutions [105]. In other research, Herr and Fischer in 2001 defined generative design as a design process where the designer’s involvement is only mediated through the generative design system rather than a hands-on method of engagement with the final design [104].

Humans can be involved in the optimization process using their specific design field to make quick decisions at each stage [196]. Human ability in critical thinking can “outperform” computers, and therefore, including humans can be a valuable asset in the optimization process [196]. Furthermore, allowing users to visualize and critically analyze the design solution space can help them understand the relationships between variables and solutions, assist in crafting more focused optimizations, and inform future design decisions [37].

Anderson et al.'s [5] “Human-guided search” proposes an example of a system where tasks were divided between humans and the machine, where the machine used the hill-climbing method to find the minimum. Then, the human-defined regions of interest were where the computer had to do further local searches to avoid getting trapped in local minima [5]. They applied their method to the traveling salesman problem (TSP) [5].

Another example of such interactive optimization methods is explored in interactive evolutionary computation (IEC). According to Takagi’s [221] survey on interactive evolutionary computation, IEC is an optimization process that uses evolutionary algorithms combined with human subjective feedback. This subjective feedback includes the user's preference, intuition, and emotions [221]. In architecture, it is possible to integrate fitness functions such as energy performance. However, interactive evolutionary algorithms can be the most helpful solution if the focus is on topics such as user's subjective feedback in genetic algorithms, which were first introduced by Dawkins in 1986 in his book “The Blind Watchmaker” [55]. Dawkins introduced the concept of Biomorphs, which allows the user to select preferred candidates subjectively using an “artificial selection” [55]. Biomorph has inspired other interactive evolutionary projects, including the Biomorpher Grasshopper plugin for architects [98]. Interactive evolutionary algorithms have also been extensively explored in different fields. Based on Takagi’s [221] survey on interactive evolutionary algorithms, IEC is used in a variety of areas, including “graphic design, music, industrial design, and image processing” [221].

With the prevalence of AI in architectural design, researchers in architecture have created frameworks for using AI algorithms in the generative design process. In 2023 Zhong et al. introduced Building-GNN, a deep learning system that creates 3D-voxel building models in collaboration with architects [248].

Interactive Genetic Algorithms (IGA) in Architecture

An early example of the use of interactive genetic algorithms in architecture was John Frazer and Peter Graham’s work in 1993 [86]. Their project aimed to design an optimized yacht by combining the user's selections (“artificial selection”) in addition to the natural selection process of the algorithm. They assigned the curves’ control points as the variables in the genetic algorithm. User selection was based on the user's preference and expertise [86].

Another early example of these algorithms is discussed in the work by Coates et al. in 1999 [43]. In this project, users are presented with nine solutions, where they can select the parents for breeding, which are then used in the creation of the next generation of designs [43]. For the final evolution at the bottom right, the user was interested in a more regular branching form, and that criteria guided the design evolution accordingly [43]. The initial generation has more variety and randomness in the overall forms achieved, while the last generation has similar growth patterns among individuals.

The use of interactive genetic algorithms was also explored in research by Banerjee et al. [15] in 2008. This project implements IGAs to generate floor plans in a collaborative “peer-peer” methodology called “Interactive Genetic Algorithm Peer-to-Peer (IGAP)” [181]. In their proposed platform, users were given nine design options in each generation, and they implemented both

qualitative and quantitative objectives (such as area minimization). During the process, each peer had the ability to communicate with other peers in the form of requesting their best solutions in each generation [181]. If multiple users were involved in the optimization process, the "requesting user" was provided with a group of best solutions from their peers [181].

Other examples of the use of interactive optimization in architecture include Mor et al.'s research in 2011 [158]. Their work introduces an interactive "biomimetic space installation", where the user engages and collaborates with the "swarm agents" by adding spaces in real-time as "building blocks" in the platform [158]. Danhaive and Mueller explored a recent example of using IGAs in architecture in 2015 [53]. This project aimed to integrate the designer's feedback into the structural optimization process using interactive evolutionary algorithms [53]. The interaction happened in an interface called "Stormcloud" while utilizing Grasshopper under the hood [53]. Based on the visualizations presented in Stormcloud using a mutation rate of 0.2 and population size of thirty, the top eight designs were presented to users.

A more accessible use of IGAs in architecture using a plugin called Biomorpher was introduced by Harding and Brandt-Olsen in 2018 [98]. Biomorpher is a plugin in Grasshopper, and users can use parametric models for exploration using their proposed IGA [98]. All the variable values in the parametric model are turned into a normalized number between 0.0 and 1.0 to create genotypes [98]. Users control settings such as the population size, crossover, and mutation rate [98]. Users can add performance-based objectives to the IGA if needed [98]. Another characteristic of their tool is that if a design option is selected (by checking the box next to the cluster's displayed centroid), the algorithm sets the fitness score to 1.0 for the centroid and all the individuals in that same cluster [98]. Biomorpher's cluster centroids count (number of options displayed to the user in each generation) is 12, and their initial generation population is 100. The mutation rates are between 0.20 and 0.01; they used no crossovers.

Despite the benefits of including humans in the optimization process in IGAs, the optimization process can result in human fatigue [221]. Users might face many design options to choose from and evaluate, which can become tiring over time [221]. If the algorithm does not converge to an optimal solution based on the user's feedback during early generations, the continuation of generations can make IGAs impractical [221]. Another critical issue in this context is how to reach convergence with only a small population size and as few generations as possible while still reaching a desired solution for the users [221]. User's "gradual cognition" during the optimization process also needs to be considered, which can make their evaluations gradually shift [92]. To solve the mentioned problems and improve the overall efficiency of IGAs, there have been different approaches suggested in the literature, including the use of surrogate-assisted IGAs to estimate fitness values as discussed by Sun et al. [216], [218] and Gong et al. [91], user interface design improvements, and selecting the suitable clustering methods as discussed by Sun et al. [218]. A combination of these solutions can be used to solve fatigue issues [216].

IGAs with interval-based fitness

Using a single value as fitness is the most common configuration in optimization methods [92]. However, in the case of interactive optimizations, the user's evaluation of the individual is uncertain. It can shift gradually, which makes it challenging to represent fitness as a single value

[92]. To solve this, Gong et al. [92] discussed assigning intervals to show the user's fitness values range. When working with intervals, one of the challenges is that comparing two intervals is more complex than comparing two individual values while comparing fitness values in the tournament selection process [216]. To solve this challenge, a probability can be implemented to determine "dominance" [39]. Gong et al. [92] also discussed that assigning interval fitness to evaluate individuals is convenient for the user and helps reduce the risk of user fatigue. However, even using interval fitness will not be sufficient if the user needs to evaluate a large population in each generation [92]. It is helpful to use a surrogate model to approximate this fitness value for unevaluated solutions [216].

Surrogate-Assisted IGA

A surrogate model is an approximation used to make a computationally less expensive model [65]. The typical methods to create a surrogate model include "rational functions, radial basis functions, and support vector machines" [65]. An example of surrogate-assisted IGAs was presented in a publication by Gong et al. [91] in 2009. In their proposed method, the surrogate models were made based on the interval fitness of an individual, using two "radial basis function networks (RBFN)" [91]. The two RBFNs were used to model the fitness's upper and lower limits, then trained and used to estimate the fitness value of every individual in the population [91].

As discussed by Uykan et al. [232], RBFNs have three components: an input, a hidden output component, and an output component. The output calculates the final outputs using a weighted sum using the hidden layer outputs [232]. An RBFN can find a close approximation to "any continuous function" using the Gaussian kernel function [232].

The surrogate model approach helps with better performance in the algorithm because more samples are evaluated, and user fatigue is reduced by limiting the required number of individuals evaluated by the users [216]. The training process of surrogate models can be done using supervised learning algorithms, which can be problematic due to the limited proportion of evaluated individuals for larger populations [216]. To solve this issue, we can use a semi-supervised learning (SSL) method [216]. In unsupervised learning, the algorithm trains samples without labeled data, while a supervised learning algorithm uses labeled samples to train the function for predicting future results [249]. Semi-supervised Learning (SSL) allows for training learners with only a few labeled samples [249]. Then, it uses the rest of the unlabeled data, which can be a large amount, to improve the initial learners [218].

To use large populations in IGAs, Kim [126] used groups of large populations clustered using the "k-means clustering" method. For the IGA, the cluster's centroid can be evaluated as a method to reduce the number of user evaluations rather than looking at every individual in the population. In "A New Surrogate-Assisted Interactive Genetic Algorithm", Sun et al. [216] experimented with a population size of 200 with 10 cluster centroids displayed to the user for evaluation in each generation.

The steps used in a semi-supervised IGA with a large population were discussed by Sun et al. [216] as follows:

1. A large population of individuals is randomly generated [216].
2. The population is clustered using the k-means clustering method [216].
3. Each cluster has an individual representative of its centroid, which will be the individual the user will evaluate using the interval fitness method [216].
4. Then, a semi-supervised learning method is used to train two RBFNs as "learners," and this method is "co-training" [216].
5. The surrogate model is generated using the two learners [216].
6. The generated surrogate models are used to estimate the fitness value of the unevaluated individuals [216].
7. Certain individuals are selected to be re-evaluated by the user to understand the accuracy of the estimates and follow the user's "gradual cognition" [216].
8. This process continues until the user finds a desirable solution [216].

Based on Sun et al.'s [216] proposal, when users are sure about their preference of an option, the mean of the interval $c(f(\ x_i(t))$will be maximized (it will be closer to the upper end of the interval). The width of the interval $w\ (f(x_i(t)))$ will be minimized because the uncertainty bounds are minimized [216]. Based on this assumption, to find optimized options in future generations, the algorithm needs to be closer to the maximum $c(f(\ x_i(t)))$ and minimum $w\ (f(x_i(t)))$ [216].

The presented surrogate-assisted methods in this section describe how using semi-supervised learning methods and interactive genetic algorithms can help include human wisdom and intuition in a generative process. These methods generate larger populations in the optimization process while selectively displaying options to users to evaluate and make decisions. The users work with the algorithm to evaluate the solutions and can re-evaluate them throughout the process as their cognition changes during the experience. These methods can be applied to include humans with different backgrounds in the design process and help them guide the optimization process to include their subjective feedback [216]. As discussed in the example of Biomorpher in this section [98], objective functions such as energy or daylight performance can also be added to such algorithms to get a combination of quantitative and qualitative feedback to improve and generate new design solutions.

Data Visualization in Generative Design

Visualization methods used in generative design processes play an important role in the user's experience and cognition during and after the design process. Despite the increasingly effective computational methods available to integrate into the design process, it is challenging to visualize large design solution spaces to users efficiently while avoiding user fatigue [37]. Data visualization

techniques have explored human visual pattern recognition to help design these complex interactions [37]. While it is important to show the user the final best solution or the set of best designs (Pareto front), allowing the user to access only limited solutions does not provide a full exploration chance [37]. Such explorations can help the user to understand deeper relationships between the objectives and solutions rather than only searching for an optimal solution [37]. Maher et al. [146] discuss how the design solution space and the goals can be "ill-defined" in generative design, and it is important to allow the fitness functions and solutions to evolve. To enable exploration, they propose a GA that enables "co-evolution" of fitness function and solutions in a generative design process [146].

To avoid "visualization clutter" when representing large data, methods such as parallel coordinate plots (PCP) can be used in generative design projects. PCP plots all "dependent and independent dimensions" of data on a group of parallel lines, and in the next step, lines representing the solutions will form [37]. Another version of this method can be done by creating a "polar plot" [37]. PCP enables users to see the "patterns and correlations" within the data set and is widely used to inspect solution spaces [2].

In addition to visualization of generated results, it is important to create a fully interactive optimization tool to help engage the user in the optimization exploration process. As an example, Octopus in Grasshopper allows for user interactions within the solution space and for modifications of the objectives and optimization settings while the optimization is still in progress [233]. This method differs from Galapagos and Goat, where the problem definition is fixed and not interactive once it is running [233].

In 2017, Brown and Mueller discussed methods of "simplifying the solution space" automatically by keeping potentially important variables, removing unnecessary variables, and allowing the user to work with the remaining important variables [24]. Another method that can be used to simplify the solution spaces is the clustering method, where the solutions are placed in groups based on different similarity-based methods [2]. They applied a clustering method to "eleven bar truss" optimization results from an EA optimization [24]. The first cluster is the best-performing cluster, while the second cluster is narrow but exhibits similar performance results [24]. Cluster three is also noted because it has the most "geometric diversity" from the others [24].

Other methods of visualizations are explored in projects such as Dreamlens and Cupid, where design options' visualizations and their related data, such as scatter plots or PCP, are linked and can be filtered [2]. In Cupid, geometric shapes are grouped using a clustering method and nested within a parallel coordinate system [16]. The clustering of shapes is based on their shape similarities, and the results are displayed in a radial tree combined with a PCP [16]. Users can also modify the clusters based on their decision by either merging or splitting them. The example below shows the different visualization methods and clustering in Cupid for a cup 3D model form design, which the user can interact with [16]. A vast array of visualization methods in generative design enables users to understand the solution space further while avoiding issues such as visual clutter and user fatigue. Users can explore the solution space further to understand potential associations of objective results and topics such as shape geometry similarities or other attributes of their designs.

2.2 Participatory Design in Architecture

2.2.1 Early-Stage Architectural Design

The importance of massing in the design process was discussed in Ferris's 1922 New York Times article [263], where they discussed new zoning laws in New York City and what that entailed regarding the city's future image. At the early design stage, architects design a volume that becomes a reference for applying zoning requirements and building codes to the design [142]. This representation should respond to the site itself, the site's context, the program (defined by the client), and the overall aesthetic goals of the design team and the client. Finding this representative volume is a complex and essential task as it solidifies the overall building geometry during the early design stages. In this stage, called the massing study, many theoretical possibilities of shapes, sizes, and volumes in the plot are studied [142]. In California, the Building code proposes limitations on height and area [264]. Per these regulations, the height above grade depends on the occupancy type and type of construction [264]. The allowable area can also be determined using the building code, where occupancy classification and construction type impact the allowable area again. In addition to the building codes, architects need to refer to the zoning code based on the site's location, which might limit these numbers further. This stage, as discussed in the educational setting, includes physical prototyping and digital drawings, including diagrams, to help the team with decision-making. At this point, there is a prime opportunity to include stakeholders and peers to receive feedback and collaboratively work on improved design solutions. Generative design at these early stages can help designers find novel solutions to different design objectives.

2.2.2 Design Review Process in Architecture

Providing feedback, a practice that architects call "critiquing," is a method by which architects and architectural design students provide feedback regarding architectural design projects in a design studio [172]. In architectural design education, students are initially given a design problem to solve creatively [137]. The process of learning in architecture is considered "problem-based learning (PBL)", where students guide their learning process through their exploration for solutions [137]. In the professional practice setting, architects are usually provided with a detailed program by the client, and they start exploring different design options in the studio environment at their architectural office. The studio is usually an open space office where designers can informally discuss their progress with their colleagues and frequently share their outcomes with the clients.

In the educational setting, the instructor sets the project goals and their details based on predefined learning criteria to be achieved during the semester [14]. The design problems posed to the students are usually in the form of a detailed program and the cultural or social aspects they need to respond to for a given site [14]. Students go through several exercises to evolve their designs, which might start as hand sketches and develop into drawings, digital prototypes, and finally physical prototypes [14]. Design learning is a challenging process, where design activities mainly happen using representations such as drawings and 3D models, and these representations serve as the main tool for communication between the designer and their reviewers [206].

In architecture education, students' assessments are based on their presentations, a method carried out in the architecture practice as internal design reviews and client meetings [172]. Architecture students work on their design projects and present them to their professors, which mirrors the role of design directors in professional practice. A big part of the learning process for students is improving based on the feedback from their instructors and classmates (peers) [172]. The instructional methods today, including providing feedback, are based on "the Ecole des Beaux-Arts" traditions with the school's concept of atelier [172]. The Ecole des Beaux-Arts was established in 1819, but its education style dates back to 1795 when the Institut de France was founded [88]. Learning in the atelier happened under an instructor's guidance in a student-organized setting, where students get involved in each other's design process informally [88], [172].

A design reviewer (in the case of architecture school, it can be the instructor, an outside reviewer, or peers) uses a "repertoire", which is a combination of examples and design ideas that architects build throughout the years as they learn architecture [172], [193]. When an architect reviews a design project, they go through their repertoire to provide their feedback [172]. Research conducted by Uluoğlu on architecture courses at U.C. Berkeley, MIT, Carnegie-Mellon University, Harvard Graduate School of Design, and Istanbul Technical University resulted in findings about the importance of communication methods used by a studio instructor to convey design concepts and how communication is a more important characteristic than being a good designer themselves [231].

When discussing studio critique, it is critical to introduce the typical studio design stages. Uluoğlu [231], in their research, discussed the stages of studio work, where the studio instructor uses methods such as lectures and readings to convey "abstract knowledge" in the studio. These methods are used throughout the stages of design in the studio process, and they include initial introductions to the design, analysis of site and massing options, design option proposals, and further design developments of a selected design option [231]. Eventually, it progresses to details related to building systems and specific technologies [231].

As discussed by Oh et al., during the early massing phase, the role of the instructor will be to guide the students with massing placement and its overall form [172]. In the later stages of the design, the instructor might reference related precedents for the students to help them with their design process exploration [172]. Regarding the critiquing setting, the categories include desk crits, group crits, interim review, final review, and informal interactions [172]. The discussions and criticism provided to the designers play an important role in the evolution of their design process [14].

Desk crits are an informal discussion between the student and their instructor throughout the studio course [172]. Group crits allow students to learn from their peers and instructors and enable students to learn from each other's methodologies [172]. As described by Sarah Hirschman, the instructor of the integrated studio at UC Berkeley, one of the methods used during the integrated studio early massing exercises is asking students to select a design by one of their peers and provide feedback for their peers and discuss how to improve those designs [106].

Interim and final reviews usually include external reviewers, who are usually experts such as practicing architects or engineers based on the topic of the studio [172]. These reviews are studio-

wide and usually represent a milestone in the project phase, such as the end of Pre-design or Schematic Design (SD) [172].

A common time to have the first interim review is usually at the end of the pre-design phase, where students have analyzed the building site, including its historical background, zoning requirements, neighborhood, and environmental aspects [172]. One of the important aspects of a studio is the informal interaction, which was challenging to achieve during remote instructions due to the COVID-19 pandemic [172].

In a professional setting, architects who work collaboratively on the same project attend frequent internal meetings, where they discuss their progress and challenges with their colleagues and provide feedback to each other. Senior designers and project managers will frequently discuss the junior designers' progress during informal interaction, providing feedback and directions to the designers, similar to desk crits in the educational setting. Junior designers can develop their expertise within the professional practice using these informal interactions with their peers. As the project progresses, the architecture team shares the design with the client (similar to interim reviews in the educational setting). It subsequently receives further feedback on how to improve the design. In addition to these reviews, depending on the firm's size, there might be office-wide design reviews where the company leaders and design directors provide design feedback to design teams.

As discussed by Uluoğlu, the content of critiques depends on the design stage and the issues the designers face in reaching their final designs [231]. Instructors might create a more supervised design studio environment during the first years of architectural education, while towards the end, the students have more design freedom as their knowledge in design has increased [231]. Third-year students typically spend 1-2 weeks on massing studies [106]. During the final years of undergraduate architecture, the studios can be conducted with more focus on the later stages of the design, such as the schematic and design development stages [231].

When it comes to relationships in the literature, Oh et al. discussed the relationships within the studio as: “master-apprentice, user and designer, and peer relationships” [172]. Within the user and designer relationship, the instructor plays the role of the user or client by providing neutral feedback on the design. Peer relationships include peer critiquing as an informal method of engaging in a studio by providing feedback. Peer critiquing allows for collaborative learning in the studio and teaches designers how to integrate peer’s feedback in the design process [172].

The communication methods used in the design studio include: “speech, written comments, drawings and gestures” [172]. When using speech, reviewers combine different methods of communication, such as drawings (sketches) which can be more abstract diagrams to show topics such as adjacency or overall massing [172]. Written comments are less common as a communication method in an educational setting; however they became an important mode of communication during COVID-19 remote instructions. As an example, in 2020 and 2021, design studios at UC Berkeley College of Environmental Design used an online platform called Miro to view and present design projects remotely [271]. The common methods of communication used on Miro are drawings and leaving comments/sticky notes, which enable further asynchronous communication with students.

Written comments and drawings (markups) are common modes of communication in professional practice. They are used internally and externally in city and shop drawing reviews in the CD phase. Another method of communication, which is difficult to replicate during remote instruction and design collaborations, is gestures that are typically used to communicate abstract design concepts [172]. The use of hand gestures in remote conditions to convey design feedback is usually done by drawing over a design, as discussed in the case of Miro. Delivery of feedback is another important factor in a design critique setting, and the abstract explanations and examples provided to the designer play an important role in design development [172].

2.2.3 Participatory Design in Architecture

Participatory Design Background

Participatory design is one of the approaches used in human-centered design [210]. Users play a crucial role in human-centered design (HCD) by helping designers and research teams understand the user's perspective and preferences [210]. HCD processes introduce two main challenges:

1. Experts versus Users Challenge: Tension between experts (designers and research team members) and users is often due to the knowledge gap between the two groups [210].
2. Researchers versus Designers Challenge: Researchers and designers experience a tension because of the differences between research findings during the participatory research process and the nature of design solution explorations [210].

Participatory design initially started in Europe as part of the "Scandinavian workplace democracy movement" and later was introduced in England, where labor unions expressed their concerns about achieving workplace democracy with the introduction of new technologies [161]. Participatory design investigated technologies and how they can impact people's lives and society [161]. In 1971, the Design Research Society conducted a pivotal conference on "design participation" to discuss the effects of user participation in planning and design [49]. In the conference proceedings, Nigel Cross discussed how including users in the design process can avoid unexpected problems from the early stages of design [49]. The conference's outcome, as Cross described, was "a blurry distinction" between the designer and the user [49].

Later, in the 1990s, the PDC conference established participatory design as a research field where researchers such as Namioka and Schuler discussed their stance on using computers and how they are tools to facilitate human workflow rather than automate it [165]. They were discussed to make the best decisions regarding their work more reliably than computer users' [165].

The concept of participatory design (PD) has been defined as a design approach where users play an essential role in making decisions regarding the technologies shaping their environments [213]**.** Halskov and Hansen identified five fundamental facets for participatory design: politics, people, context, methods, and product [97]. In this context, people play an essential player in decision-making and their inclusion is essential for success [97]. The shared value in all these facets is people, and the technologies will revolve around the people [97].

Participatory design moved beyond the workplace concept over time with the introduction of new technologies [97]. Mainsah and Morrison's work that introduces the use of social media and participatory design to engage users is an example of these new possibilities [147].

Participatory Design Applications in Architecture

In 1984, King Charles of England gave a pivotal speech regarding the use of participatory design in architecture and urban planning, in which he discussed the urgency of considering the user's needs during the design process [238]. He called "Community Architecture," or participatory design, a new hope [238]. Participatory design in architecture and its methods have been discussed in publications as early as 1968-1969, describing the importance of including users in design [17]. This approach directly conflicts with architects' expectations for their role in the process and the culture revolving around the importance of architects pervasive in the architecture industry [17], [238]. Such theories started questioning power dynamics and architecture's political essence [17]. In 2018, participatory design researcher Rachel Luck defined the main aspects of participatory design in architecture and urban planning, which can also be extended to other fields [143]. These aspects include: "equalizing power relations, situation-based actions, mutual learning, tools and techniques, alternative visions about technology, and democratic practices" [143].

There are different approaches toward the role of designers compared to the users in the PD approach. Knevitt and Wates described community architecture as a method that enables users' inclusion in the design process as clients who have the power to make changes to the built environment, and the relationship between the users and designers is a partnership [238]. In 2018, Beyrele, in their PhD dissertation, discussed the different roles that architects have played in the PD process, where they can be considered "professional amateurs", who need to allow the users to take control of their environments [154], [17].

Table 2.3. Differences between community architecture and conventional architecture [238].

	Conventional architecture	**Community architecture**
User's role	Users are not the decision-makers, and decisions are left to the experts.	Users are perceived as the clients and active participants in the design process.
User and expert relationships	Users and experts have few or potentially no interactions. Developers might make low-level attempts to consult with the users.	Experts and users are partners.
Role of experts	Experts are exclusive and not accessible by users.	Experts facilitate the inclusion of users during the design process.
Project Scale	Large scale	Small-scale
Motivations	Experts are motivated by reputation and future opportunities. The developers are interested in the financial aspects of the project.	The motivation is to improve communities' living conditions.

Note: Table is created based on Wates and Knevitt [238], with modifications by author.

Kossac et al.'s research in 2010 discussed how the integration of users in the design process can be conflicting despite the resulting benefit of user empowerment [132]. To solve these conflicts, architects need help from the rest of the PD team to solve these issues [132].

User participation in urban planning has been investigated and emphasized extensively. Brody et al. in 2003 compared user engagements in local planning practices within different states in the US, along with their strengths and weaknesses [22]. Although many states have requirements regarding involving users in the planning process and decision-making (such as Oregon and Washington), the nature and the stage of this involvement can be very different [22].

User participation can help mitigate concerns about sustainable housing practices used in prefabrication construction, where design solutions are produced without the future residents' involvement [21]. Understanding the users' future needs to make these solutions more flexible and resilient is crucial, and it demonstrates the importance of user inclusion in the design process [21].

To further emphasize the importance of participatory design in architecture, successful examples will be discussed with different building typologies based on Forsyth et al.'s "Architecture, Participation and Society" [117].

An example of applying participatory design was applied to the Kingsdale School project in South London, which was awarded AIA's Excellence in Design Award in 2005 [117]. The early participatory design activities started one year before the architectural design started, and the participation continued throughout the design phases, including construction through the completion of the project [117]. The architects in this process acted as mediators who brought in their design ideas for collaboration with the community [117]. A conflict was encountered during the participation process while working with the affluent local community, which initially did not support the project. However, their perception of the design evolved during the project's progress [117]. The participatory methods used during early design stages included hands-on activities such as "collages, and photography exercises by students to point out issues in the existing environment, and introducing precedents by architects" [117]. They also used surveys and presentations to communicate with the school community [117].

An example application of participatory design for a residential typology was in a social housing project called "Canmore Place" in a rural area of Kincardine O'Neil, Scotland [117]. This project was completed in three stages: 1. Research phase with the involvement of users, 2. Participatory design exercises with the local community during the design phase before construction; 3. Participatory activities during the occupancy phase [117]. In the early stage of the project, a combination of researchers in different social sciences conducted research with users to understand their needs and help the designers make future design decisions regarding the project [117]. After developing design ideas based on this early stage, they presented the different design options to the users and stakeholders on a community scale. The design kept evolving while keeping users involved [117]. During the later stages of the design, decisions such as materials and floor plans were discussed with the community to get feedback about the project [117]. The community activities included displaying 3D models, photographs, and drawings to communicate the concepts with the community [117]. During the early occupancy, users' motions within the space were

analyzed to understand how the house was being used, and the design team conducted studies to see how users planned to modify the layout of the units [117].

Another example of participatory design is explored in the Lee Valley Millennium Centre in Liverpool [117]. Community involvement started from the initial funding stages over several years [117]. The architects used two surveys to understand user preferences before the design process started on a larger community scale. In the later design stages, the participatory meetings directly included local community representatives [117]. During the community discussions, they used massing prototypes and a physical prototype showing the interior layout to better understand the spatial qualities of the proposed options [117]. The massing models that showed the context were preferred to the digital 3D model of the massing [117].

As seen in the last few examples, participatory design in architecture can happen during different design stages and in different building typologies. It is important to include users from an early stage of design for the highest level of success in user engagement since, as the project evolves, user involvement becomes more and more complex and closer to informational sessions rather than design activities. In a participatory design process, large communities might be interested in participating. However, with traditional methods of participation, it is difficult to include and analyze many users' preferences and needs. One way to tackle this issue was discussed at the Lee Valley Millennium Centre, where they formed community representative groups for design collaborations [117].

In addition to the field of architecture, participatory design has been explored in larger-scale urban planning projects. An urban scale study conducted by Kunze et al. proposed a participatory planning approach that engages stakeholders in the design and development of a development in Porto Alegre, Brazil [136]. This project looked at the design both on an urban scale and the architectural design scale, where users were introduced to the initial planning information to discuss potential ideas and design solutions [136]. Their participatory approach used communication on a website, in addition to conducting workshops, to enable user engagement [136]. In other large-scale architectural projects, designers develop web-based tools to enable users (future residents) to create unit parcels in large housing projects and show their preferences and spatial needs for layout optimization [100].

Integration of users in the design process has been explored in other fields of design, including healthcare design. In 2019, Tiainen and Jouppila explored a co-design process of a patient room design by including 238 users from different healthcare backgrounds to participate [227]. The users were active participants in this process and shared valuable feedback based on their backgrounds by evaluating the evolution of virtual design options with greater details over time [227]. Using participatory design methods has been explored in other public projects, such as designing a museum exhibition space where users could evaluate the exhibitions and provide design concepts for exhibition design [225].

With the advancements in technology and wider access to immersive technologies and web-based platforms, researchers and designers can involve larger groups of users in the design process. On the other hand, users can better understand the proposed designs' spatial and functional qualities and the potential environmental consequences of their suggested design decisions. Users' feedback

can be gathered and analyzed at different stages of design more efficiently, and engaging the users with different design activities can lead to adaptive and responsive architectural designs.

Participatory Computational Design in Architecture

Based on a literature review in participatory design, the following publications were selected that integrated computational design in a participatory approach. The keywords utilized include "participatory design", "participatory generative design", "participatory parametric design" and "participatory computational design" on CumInCAD and Google Scholar from 2013 to 2024. The overall context of most of these projects is urban design and larger-scale decision-making processes. The table below summarizes the findings from the publications. This section will discuss seven of the most relevant publications to this dissertation.

Table 2.4. Reviewed computational participatory design publications.

Publication Title	Authors	Year	Description
Interactive generative system supporting participatory house design [139]	Kwieciński and Słyk	2023	A participatory generative design tool for house layout generation for non-experts.
DeciGenArch A generative design methodology for participatory architectural configuration via multi-criteria decision analysis [204]	Soman et al.	2022	Focused on zoning, DeciGenArch combines Multi-Criteria Decision-Analysis (MCDA) with a Multi-Agent System (MAS).
PlacemakingAI: Participatory Urban Design with Generative Adversarial Networks [125]	Kim et al.	2022	Urban planning tool using Generative Adversarial Networks (GANs) aims to include citizens, designers, and stakeholders.
The Development of Design Support System for Public Participation of Community Public Space Design Using Mixed Reality [52]	Dan et al.	2021	A mixed reality method of participatory design for design collaboration of a public space.
V-Dream: Immersive Exploration of Generative Design Solution Space [121]	Keshavarzi et al.	2020	An interactive spatial exploration in virtual reality that uses a generative analysis framework for exploring large-scale solution space. The evaluations provided to the user in VR were both qualitative and performance based.

Publication Title	Authors	Year	Description
Encouraging Community Participation in Design Decision-Making Through Reactive Scripting [128]	Kimm and Burry	2020	A system that uses “abstracted expert knowledge” in algorithms that enable users to interact with the environment in a gamified environment and see real-time changes based on their design decisions.
Citizen Engagement through Design Space Exploration: Integrating citizen knowledge and expert design in computational urban planning [129]	Knecht et al.	2019	A parametric design method for a designer to share designs with participants using an online platform called Beta.Speckle.
Participation-based Parametric Design in Early Stages - A participative design process for spatial planning in office building [51]	Daher et al.	2018	A proposed participatory method where designers oversee defining parameters, while users have the ability to make spatial changes based on their needs.
Customised Collaborative Urban Design - A Collective User-based Urban Information System through Gaming [11]	Bai et al.	2018	A study in which they gather urban use preferences that result in the collective decision in a gamified environment. Preference was analyzed to find an optimal layout for the community.
WP-BIM: Web-based Parametric BIM Towards Online Collaborative Design and Optimization [244]	W. Yan	2017	This describes a web-based parametric design tool that uses BIM models and visual programming-based user interfaces for parametric design and optimizations. Their goal was to create a method for design teams and clients to work collaboratively.
Participatory Design Supported with Design System and Augmented Reality [138]	Kwiecińskі1 et al.	2017	This paper investigates use of augmented reality and parametric design combined with participatory design. Their experiments were conducted during workshops to test the developed grasshopper tools.
Tangible Mixed Reality On-Site: Interactive Augmented Visualisations from Architectural Working Models in Urban Design [194]	Schubert et al.	2015	An interactive and collaborative tangible mixed reality system that allows for real time visualization of simulations such as computational fluid dynamics and visualizations of architectural design in an AR environment.
Parametric Design Strategies for Collaborative and Participatory Urban Design [211]	Steinø et al.	2013	A proposed system using parametric design to create different design options on an urban design scale to increase stakeholder engagement. They

Publication Title	Authors	Year	Description
			conducted workshops to create the parametric approaches.
A Discursive Grammar for Customizing Mass Housing [72]	J. Duarte	2003	A web-based platform that uses discursive grammar to generate house designs based on a user's defined goals.

An example of participation in the parametric design process is presented in Steinø et al.'s 2013 publication [211]. This research discussed parametric design in collaborative urban design among non-designers and designers. They conducted a workshop to create different design options using parametric design to create urban renewal projects. The main focus of workshops was for the student designers to parametrize different aspects of the urban model using the City Engine platform [211]. As for interactions, they proposed a tangible interface as model blocks or game objects to allow non-designers to interact with the design (instead of a GUI from City Engine) [211].

In 2018, Daher et al. proposed a parametric approach to design while including participants in the early stage of the design process [51]. Their proposed approach was tested in a spatial planning exercise of an office building. The users, combined with the planning experts, were able to modify objectives, manipulate the model, modify variables, and visualize the interactions as well as evaluate them in an interactive environment [51]. The user's participation was made possible by modifying proximities between spaces and doing operations such as adding or removing walls [51]. Regarding visualization, they proposed using Hololens, combined with a tangible interface tabletop, to interact with the models in real-time [51]. The participation levels vary; in some cases, the participation was more for users to get informed during the process, and in others, there was direct participation where users were involved in setting objectives and weighing the importance level of objectives [51].

In 2018, an exploration of using immersive technologies in participatory design was conducted by Bai et al. [11]. This study focused on immersive technologies in participatory design, where urban use preferences were gathered from participants, and the results were displayed in AR. The collective decisions and preferences were analyzed to find an optimal layout for the community [11].

In 2019, Knecht et al. proposed a parametric design method for the designer to share a design with participants using an online platform called Beta.Speckle [129]. During the usability studies of their approach, they discovered that user preferences were based on "visual order" and topics such as "economic, social, and environmental values" [129]. Speckle could read data input and output from Grasshopper and Dynamo, and using the online platform, designers and stakeholders could communicate their ideas [129].

In 2020, Kimm and Burry [128] proposed a system that enabled users to interact with the environment in a gamified environment and see real-time changes based on their design decisions

[128]. With their proposed approach, the users could visualize and interact with the virtual urban environment, directly impacting it without needing an expert [128]. Their prototype used Unity3D as the game engine and OpenStreetMap to reconstruct the urban model in the virtual space focused on a rural area in Northern India [128]. Their case study looked at visualizing paths created based on users' decisions regarding the placement of resources and infrastructures within the neighborhood over time within the experience [128].

Dan et al. in 2021 introduced an MR-based participatory design tool to encourage user participation in community planning [52]. They created an interactive interface in their proposed system, allowing interactive visualizations and design modifications [52]. They tested their concept of designing a community public space in Japan and invited 12 participants to visualize the design proposals in MR on the actual site [52]. The users could modify design elements (such as furniture) or change the parameters such as dimensions and materials [52]. Based on their results, most participants found the proposed MR system helpful in allowing them to express their design preferences and intentions [52].

In 2023, Kwieciński and Słyk introduced HOPLA-Home Planner tool, a generative design tool for house layout design to non-experts [139]. The flexibility in the proposed tool posed a challenge where the final layouts did not result in high quality designs [139].

2.2.4 User Feedback in Crowdsourcing

The use of peer feedback and combining experts' and non-expert participants' inputs has been explored in crowdsourced systems. As discussed by Hossain and Kauranen, the concept of crowdsourcing can be traced back to 1714 with the longitude calculation competition in England, with its more contemporary emergence in Toyota's logo design competition in 1936 [109].

In 2006, Jeff Howe, in an article in Wired Magazine, discussed crowdsourcing as a methodology that enables regular people to participate in problem-solving activities [110]. Crowdsourcing systems' main components are humans who can solve different problems efficiently [66]. Examples of these systems include Wikipedia and Mechanical Turk-based systems [66].

As discussed in this dissertation, design reviews are essential in providing valuable feedback and are necessary for the progression of a design project. Without access to users with a design background in the studio environment, replicating the design feedback process can be difficult, forcing the designers to find other resources for feedback [245].

There are platforms where designers can get feedback for their design proposals, including Dribbble [245]. The quality of comments on such platforms has been discussed in Yuan et al.'s [245] and Xu and Bailey's [241] research. Xu and Bailey discussed how only a small portion of the feedback is useful [241]. Part of this issue can be due to the non-expert not being provided with any objectives to provide specific feedback to other users [245]. The comments provided might not be "facilitative" or "directive", and they can be mainly positive and brief comments [172].

Yuan et al. [245] discussed the benefits of providing rubrics to non-design experts participating in the design evaluation process to tackle the issues of unproductive feedback, as discussed earlier

[245]. Crowd workers who provide feedback on platforms like Amazon Mechanical Turk are likely non-experts in our desired field [272]. In order to receive more useful feedback from non-expert users, strategies can be employed, such as creating smaller tasks for user evaluations and introducing rubrics [245].

Voyant is an example of a system that enables designers to receive crowd-sourced feedback online, and it breaks down the feedback process into smaller tasks [242]. The designer will select the feedback type most suitable for the designs that they uploaded to Voyant [242]. Although this study is more related to 2D content design, it can be helpful to see approaches that could also be used for guiding and separating user feedback in architectural design processes, such as for massing design options. This can be especially useful if the massing has multiple components, which can be commented on and modified individually.

Another related example to the topic of non-expert feedback is CrowdCrit [145]. CrowdCrit follows a similar approach to Voyant by guiding the non-experts in their feedback process [145]. Crowdcrit's platform provides feedback to the crowd worker, which they can choose from based on known design principles [145]. The users can also annotate using the text-based feedback provided in CrowdCrit [145]. This system helps non-expert workers by enabling them to provide feedback that will be useful to the designer, and the designers will have access to the accumulated feedback [145]. The experiment results using CrowdCrit showed little difference between expert and non-expert feedback [145].

In another research project by Yuan et al. [245], they followed the strategies used in CrowdCrit by providing the design principles in a rubric to help increase the usefulness of the produced feedback by the crowd workers. They compared the value of feedback provided by non-experts with access to rubrics versus experts without rubrics [245]. They also used the Crowdcrit platform for this research. Based on their final results with the introduction of rubrics, non-expert feedback was not substantially different from the experts' feedback, and providing rubrics most improved the responses from the non-experts [245]. Other methods of improving non-expert feedback quality are discussed by Dow et al. [70], who enabled self-evaluation by the crowd workers and provided expert feedback to the crowd workers to improve their final submitted feedback.

In the architectural design context, participatory approaches require the inclusion of users from different backgrounds, with different levels of expertise and different expectations from the participation and the final building design. As an example, if the users are clients with a financial interest in the final result, they might be interested in learning about the number of available units, the potential complexities of the construction, the amount of floor area, and the associated costs with the construction for the various designs. On the other hand, if the participants in the process are individuals from the local community, they might be more interested in how the proposed building will change the overall aesthetics of the neighborhood, what the financial benefits of the proposed project are to the local community, the environmental impacts of the building, the building's height, its shading impacts, or issues such as access and traffic to the building. The architectural designers who are the experts (including intermediate-level experts) in this setting have different motivations and visions about the building when compared to the client and the local community during the participation. Architects have comprehensive training in architectural

design theories, technologies, and history that can create compelling designs for the community when combined with the user's perspective of the design.

Based on each user's different motivations and needs, it can be helpful to provide further information to non-experts, such as specific definitions or requests that can help those users with their critique process and increase the usefulness of their provided feedback. As mentioned in the studies above, users can be given different rubrics and objectives to guide the design during the design process. If participants are local community members, a useful rubric for them could include the definitions and limits of FAR (floor area ratio, which reflects the amount of built space on the lot), building height contextualized by the maximum allowed height, potential shading impact on the area, the amount of provided open space to the community, and data regarding the embodied carbon of the building and other environmental impact information. The provided definitions and numerical data (including a baseline for comparison) associated with each can help the users make decisions and provide useful comments during an architectural design process. In addition to quantitative data and definitions, users can also be exposed to architectural design definitions with topics such as scale and proportions. However, since understanding architectural design concepts requires long exposure to a wide range of topics and precedents, it is more difficult to describe architectural concepts by brief written statements [193]. Therefore, creating relevant rubrics based on quantitative design aspects can be helpful in this dissertation.

2.2.5 Participatory and Collaborative Interactive Evolutionary Generative Design

As discussed in the previous section, users in the participatory design come from different backgrounds, with different levels of expertise and different expectations from the participation and for the final building design, such as financial, community, or environmental impacts. These users with different motivations can provide different insights based on their backgrounds and help guide the design process. In the examples discussed in this section, we will see how interactive evolutionary design has been applied in different fields and the impact of participants and their motivations during the design process.

An example of an interactive evolutionary algorithm is introduced in WRESTORE, an interactive watershed decision-making platform initially published in 2013 [178]. This tool created design options based on user preference (user's assigned ratings), displaying options for different conservation watershed landscapes [178]. To analyze WRESTORE's success, usability testing was conducted on a combination of stakeholders and non-stakeholders participating in the optimization process [178]. The results showed that both user groups' time to optimize design options through different generations reduced in each generation once the users got accustomed to the user interface of the platform [178]. Furthermore, stakeholders were more interested in obtaining further information about the options in the platform to help them with their decision-making process, which shows they had high engagement rates [178].

WRESTORE allows the user to add their subjective feedback to optimize the locations and typology of the water conservation [9]. Using maps, users are presented with multiple design options in each experiment and are asked to rate each option. A rating on a Likert scale is then

used, combined with physical attributes relevant to watershed planning, to create a new generation of design alternatives [9]. WRESTORE intended to include a large population of stakeholders to participate in the design [9]. Based on their reported analysis, including other participants (non-stakeholders) was useful since their understanding of the platform and goals increased during the experiment [9].

The use of Interactive Genetic Algorithms (IGAs) in the context of participatory design was explored in research by Banerjee et al. in 2008 [15]. This project implemented IGAs to generate floor plans in a collaborative "peer-peer" methodology, which they called "Interactive Genetic Algorithm Peer-to-Peer (IGAP)" [181]. In their proposed platform, users were given nine design options in each generation, and they implemented both qualitative and quantitative objectives in addition to the user's preference. Their proposed system enabled users to selectively "inject" individuals from the presented peers' design, and this added individual could be seen in real-time by other participants [181]. Based on their user studies, they did not clearly show a differentiation with regard to the creativity of the outputs between the floor plans generated by each user and the collaborative results [181]. An issue with this approach was asking for a creativity rating regarding floor plan layout and design from participants who were all non-expert users (they indicated that the users were engineering, math, social science, and undecided majors) [181]. However, a benefit of their proposed method was that it could help optimize floor plan options by engaging future users in the design process [181]. Including a combination of non-experts and experts in the participation process of IGAP could have been more successful as a participatory method, where the users are interested in the final solutions rather than a purely experimental context about design creativity. A similar approach has been discussed in Project Discover, where they considered topics such as adjacency, work style, and performance-based measures such as daylighting for their optimization process [164].

In 2016, Dou et al. proposed a "multi-stage interactive genetic algorithm (MS-IGA)" that enabled users' and designers' participation in a collaborative optimization process [69]. In their proposed method, the system was in charge of deciding whether or not there needed to be a new generation. They discussed this feature as a proposed method to reduce user fatigue compared to other IGA methods [69]. They tested their proposed system on 30 graduate students, and their results showed that the MS-IGA could converge to ideal solutions more efficiently than the more traditional IGAs they compared against in the experiments [69]. Additionally, based on their eye tracking studies during these experiments, their final findings did show that their proposed application helped with fatigue reduction as desired [69]. Although this system is discussed as a collaborative system, the process is in only one direction, where the designer benefits from understanding user preferences. However, there does not seem to be a real participatory method of user integration in the design process. Their algorithm's method of combining the other user's preferences is similar to Banerjee et al.'s method, except it has less engagement among users and allows for a less democratic and less empowering experience than an automated process [15].

2.3 User Assessment in Virtual and Augmented Reality

2.3.1 Augmented and Virtual Reality in Architectural Design

Virtual reality (VR) is a computer-based simulation of an environment that can be visualized and interacted with using head-mounted devices and other hardware [273]. Freina and Ott [87], in their 2015 literature review, discussed the origins of VR from the 1960s, which was itself rooted in the 19th-century 360 art. Sensorama was introduced in the 1960s and created an immersive experience with features such as sounds and smell alongside the displayed film [87].

There are different types of VR with various levels of immersion, including "non-immersive desktop virtual environments (DVE), semi-immersive systems, and fully immersive virtual environments (IVE)" [175]. Regarding devices, VR can be categorized as either head-based VR or stationary displays [247]. In non-immersive VR, the content is displayed on a screen, and no wearables are required during the experience [214]. Immersive VR requires using head-mounted displays (HMD), providing a controlled interface to observe and gather user's responses [214]. An example of immersive VR is the Cave Automatic Virtual Environment (CAVE) project that creates the feeling of immersion by combining projection-based displays, but due to its physical space requirements, it can be challenging to scale for larger audiences [214], [87].

In the past, the difficulties associated with HMDs were the costs, availability, and discomfort issues [87]. A pivotal moment in the widespread use of HMD VR was the introduction of Oculus Rift in 2016, which had accessible pricing for consumers [87]. Meta Quest 2 is a wireless HMD with an LCD resolution of 1832×1920px per eye that acts as a standalone device [274]. HTC Vive is another example that offers VR HMDs. One of their products, the HTC Vive Flow, offers VR glasses using a personal phone as a controller [206]. Additionally, the HTC Vive Pro series, including the HTC Vive Pro Eye with Dual-OLED displays [207], has integrated eye tracking technology. Meta Quest 3 is a mixed reality HMD where sensors follow the user's gestures (e.g., hand movements) to navigate experiences [275:3].

Augmented reality (AR) uses computer graphics to overlay virtual content on the real (physical) world [94]. Hololens 2 is an AR HMD with four visible light cameras and two Infrared (IR) cameras for eye tracking [195]. In addition to HMDs, personal handheld devices such as iPhones and iPads can be used to access AR. According to a literature review by Chi et al. [39], in 1998 and 1999, two pivotal AR conferences titled the International Workshop and Symposium on Augmented Reality (ISMR) and the International Symposium on Mixed Reality (IWAR) kicked off a focus on AR technical development.

In a literature review by Suh and Prophet in 2018, they provided a summary of the main "theoretical frameworks" that were employed within immersive technologies discussed in the publications between 2010-2017 [214]. These theoretical foundations include Cognitive load theory, Presence theory, Flow theory, and Stimuli-organism-response (S-O-R) framework [214]. Regarding the research methodologies applied in the literature, they found four different methods were primarily employed using AR: experiment, case study, survey, and interview. Most projects included experiments, and 22% of the total studies used combined methods [214].

The initial introduction of VR technology to the built environment was in the 1990s. Now, it is considered a common technology in the Architecture, Engineering, and Construction (AEC) industry during design reviews and site walkthroughs [247]. Based on a literature review by Zhang et al., the research in VR related to the built environment can be divided into "architectural and engineering design, construction project management, human behavior and perception, construction safety, engineering education, and construction equipment" [247]. The publications containing discussions regarding using VR technologies in the design review process are most prominent in the categories of architectural and engineering design, construction project management, and human behavior [247].

In architectural and engineering design, design interpretation and visualization are explored by introducing VR-based design review as a solution for stakeholders to understand spatial design [247]. These systems can also include real-time interaction between users and buildings for further design modifications, as introduced in 2018 by Du et al. [71]. VR methods are especially helpful in fostering engagement in the early design stage, building an understanding of the space (scale), and increasing user satisfaction [247].

Design model development in VR is another important application in architecture [247]. An early example of the use of VR in the initial stages of the design was introduced in 1995 by Donath and Regenbrecht's VR platform, where users can make voxel-based 3D models in an immersive environment [67]. Tiltbrush, a 3D painting application in virtual reality released in 2016, is another example of this application [276]. Liftoff is a 3D modeling tool in VR that uses image processing to create curves and generate 3D geometries, and users can interact with the generated models [115]. Visualization of conceptual design using AR has been explored in architecture in the case of the Fologram application [277]. Fologram enables designers to visualize their 3D designs through Grasshopper in their AR application [277]. While the handheld devices for augmented reality such as iPads and iPhones are most widely used, using HMD AR technology such as Microsoft Hololens enables a flexible hands-free AR experience [81]. Using portable AR devices, users can visualize site-specific 3D models within the physical space. This use case is currently the most popular application outside a research setting for these technologies in architecture [39]. In 2020, Keshavarzi et al. created a framework for optimizing mutual virtual space in AR and VR environments, which is crucial in remote settings [123]. Kesharvarzi in 2022 introduced context-aware generative systems that addressed adaptable experiences in AR, VR and MR [120].

Regarding trends and visions for integrating VR and AR technologies in the AEC industries, Noghabaei et al. conducted two surveys to assess the state of AR and VR in architectural design and engineering [170]. Their results showed that residential and commercial sectors in the industry have the highest integration of these tools within their work, with the most anticipated growth for the healthcare sector [170].

Delgado et al. [54] in 2020 conducted research among fifty-four participants within the AEC industry to investigate the applications of AR and VR in the industry. Based on their findings, there are currently six main applications for AR and VR in the field, including: "Stakeholder engagement, design support, design review, construction support, operations and management, and using [54] Table 2.5 is based on Delgado et al.'s research on the use cases of AR/VR in the

field of AEC, and it includes additional examples by the author [54]. Some of the most relevant literature on using augmented and virtual reality in architecture will be discussed further.

Table 2.5. Use of AR and VR in architecture [54].

Use-Case	Literature	Author	Year
Stakeholder engagement	The Development of Design Support System for Public Participation in Community Public Space Design Using Mixed Reality [52]	Dan et al.	2021
	Challenges and enablers of augmented reality technology for in situ walkthrough applications [163]	Mutis, A. Ambekar	2020
	Examining Effectiveness and User Experiences in 3D Mobile based Augmented Reality for MSU Virtual Tour [7]	Andri et al.	2019
	Towards an Architecture of Collaborative Objects [93]	A. Grasser	2019
	CityScopeAR: Urban Design and: Crowdsourced Engagement Platform [171]	Noyman et al.	2018
	Shared Realities: A Method for Adaptive Design Incorporating Real-Time user Feedback using Virtual Reality and 3D Depth-Sensing Systems [135]	Krietemeyer, et al.	2017
	Augmented reality tour system for immersive experience of cultural heritage [127]	K. Kim et al.	2009
Design Support	RadVR: A 6DOF Virtual Reality Daylighting Analysis Tool [122]	Keshavarzi et al.	2019
	Contextual Spatial Computing: A Generative Approach	Keshavarzi, Mohammad	2022
	Shared Immersive Environments For Parametric Model Manipulation [99]	Hawton et al.	2018
	Signage visibility analysis and optimization system using BIM-enabled virtual reality (VR) environments [159]	Motamedi et al.	2017
	Integrating building information modeling and virtual reality development engines for building indoor lighting design [166]	Natephra et al.	2017
	Augmented reality applications in design and manufacturing [167]	Nee et al.	2012
	A rapid prototyping software infrastructure for user interfaces in ubiquitous augmented reality [189]	Sandor and Klinker	2005
	The First Virtual Environment Design Studio [192]	Schnabel et al.	2001

Use-Case	Literature	Author	Year
Design review	Effect of Immersive VR on Communication Patterns in Architectural Design Critiques [206]	Sopher and Gero	2021
	Design and Initial Evaluation of a VR based Immersive and Interactive Architectural Design Discussion System [111]	Hsu et al.	2020
	Research on visual comfort of underground commercial streets' pavement in China on the basis of virtual simulation [215]	Sun et al.	2019
	Tangible Mixed Reality On-Site: Interactive Augmented Visualisations from Architectural Working Models in Urban Design [194]	Schubert et al.	2015
	Collaborative visualization of engineering processes using tabletop augmented reality [68]	Dong et al.	2013
	Virtual reality mock-up for design review of hospital patient rooms [75]	Dunston et al.	2011
Construction support	Integrating mobile Building Information Modelling and Augmented Reality systems: An experimental study [41]	Chu et al.	2018
	Discrete Event Simulation and Virtual Reality Use in Industry: New Opportunities and Future Trend [230]	Turner et al.	2016
	Virtual reality-based teleoperation construction robot control system with 3D visor device [222]	Tang et al.	2010
	The use of virtual reality and intelligent database systems for procedure planning, visualization, and real-time component tracking in remote handling operations [185]	Robbins et al.	2009

Note: Table is created based on Delgado et al. [54], with modifications by author.

Using AR and VR technology can help provide a realistic representation of the design and better engagement of stakeholders with the design, which can be more beneficial than traditional 2D drawings [54]. In 2019, Mutis and Ambekar developed an AR technology called i-Tracker that superimposed 3D content on the physical environment for construction site visualizations [163]. Some challenges in using AR in this context include registration inaccuracies, drifting, and difficulties visualizing larger objects at further distances [54].

The interaction of users and designers during the design and review process can be mediated using AR and VR technologies. As early as 2001, Schnabel et al. discussed a virtual studio environment where a joint design studio program in architecture used virtual environments and HMDs as a remote communication method [192]. Students worked in teams, using the virtual platform as a collaborative medium [192]. Their virtual environment facilitated direct interaction between the concept and design, allowing the students to work with and understand their model more directly [192].

There is a body of research on the use of immersive virtual reality in an architecture studio environment to review student projects, such as the research by Sopher and Gero [206], where a case study was defined based on a third-year architecture studio course and immersive VR (IVR) was used as well as non-immersive desk-crit methods. Their results showed that the IVR encourages learner activity and higher student-centered communication [206].

In 2017, Krietemeyer et al. investigated using immersive technologies to utilize user feedback in the design process [135]. In this research, they developed a computational workflow that combined VR and depth sensing for an HCD design. This workflow was tested in an interactive installation. The users' point cloud data was gathered and accessed in real-time to enable the designers to observe the users' behavior remotely [135]. The content shared with the users was parametric so that the users could modify them, and the designer could see the results in real-time in VR [135].

In computational design, different research projects explored real-time object manipulations and visualizations, remote collaborations, and participatory design. In 2018, Hawton et al. explored parametric modeling and the generation of meshes from Grasshopper into Unity to streamline the use of computational design and object manipulations within immersive environments for designers [99]. In their workflow, data input defined within Grasshopper can be changed within the VR/AR environment, and the changes will be communicated to Grasshopper using a web server [99].

A more recent example of such a process was explored in 2020 by Hsu et al. [111]. In this project, the designers created a "VR design discussion system" for remote design meetings. Users can visualize full-scale architectural models in VR and communicate using voice while sketching in space and on surfaces. Objects can also be modified live during the discussion [111]. Their system integrated Grasshopper and Rhino with the immersive virtual environment [111]. This system attempted to address some issues with traditional design discussions, including the inability to manipulate objects in real-time or reverse changes during the discussion [111].

Dan et al. in 2021 explored participatory design workflows in mixed-reality (MR) [52]. They used MR to let the public view 3D design proposals interactively and adjust designs based on their preferences [52]. In 2018, an exploration of using immersive technologies in participatory design was conducted by Bai et al., where land use preferences were gathered from users with the results displayed in AR, and the collective decision and preference were analyzed to find an optimal layout for the community [11]. They discuss the idea of gamification of the urban design process, where the participants could arrange the site with different functions with QR codes [11].

A more conceptual exploration of the use of augmented reality in architectural design was explored in the design of VoxelCO [93]. VoxelCO allows for the real-time placement and manipulation of virtual voxels, which can be done collaboratively in a multiplayer environment [93]. VoxelCO was used to design and guide the construction of an exhibition space in a case study using AR to assemble wire cubes in the visualized positions of the virtual voxels [93].

2.3.2 The Role of User Feedback Collection in AR and VR

Critiquing in the architectural design field is the method by which feedback is most commonly communicated within the studio environment [172]. Designers learn through the Problem-Based Learning (PBL) method while going through an exploratory journey to find design solutions [137]. Designers in a professional setting discuss their work informally, in addition to the formal reviews where they receive feedback regarding their design and guidance on moving forward. Knowledge of architectural design is conveyed through the feedback designers receive from their instructors in formal review settings and what they learn from their peers' informal feedback [193], [172].

Delivery is an important factor in a design critique setting, where the explanations and examples provided to the designer play an important role in future design outcomes [172]. The reviewer's body language, such as facial expressions and hand gestures, can communicate subtle information about the reviewer's impressions of the design and their decisions [172]. In the literature, nonverbal communication has been associated with deeper subconscious communication between the designer and the reviewer, which can convey messages such as: "confirming" and "refusing" [172].

Communication modes and delivery methods are approached differently in AR/VR [172]. For example, in a remote synchronous studio environment where users join virtually using AR or VR, they might have methods such as drawings for markup and speech. However, the venue can lose some of the user's nonverbal behaviors and communications, such as hand gestures and facial expressions, which cannot be seen by the other participants [172]. This can result in miscommunications during design feedback, and designers making future decisions for their designs based on that feedback may not follow what their reviewers and peers meant.

In addition to the interactions above, during a virtual experience, users still need to continue with their informal interactions where the combination of verbal and nonverbal communication can be important in conveying ideas to one another. According to Tullis and Albert, participants do many more activities than they are asked for during a task including laughing, or looking around [229]. The users might go through a variety of emotions during the interactions, including stress and excitement [229]. Capturing the details of users' behaviors can reveal how peers view designs, which can provide additional critical inputs from an immersive studio environment experience. Several methods have been studied for capturing and visualizing these behaviors, which can give us a better understanding of a peer's experience interacting with a design project [229]. In order to analyze a reviewer's verbal expressions in further detail, comment types can be recorded and categorized based on whether they contain "positive, negative, and neutral" content [229].

In order to explore the relationship between "stimulus, user experience, and the outcomes of the immersive technology", researchers in AR and VR have developed a framework called Stimulus-Organism-Response (S-O-R) [214]. Based on the S-O-R model, "stimuli" result in changes in the user's emotions, which then turn into behavioral responses in the user [133]. In this framework, understanding the users' "cognitive and affective reactions" can help designers understand if their

current design strategies positively impact the users [214]. In addition, it can help identify the areas of improvement and inform decisions towards further evolving the design [214].

When users interact with stimuli (in this case, elements of the virtual environment), they will display cognitive reactions accordingly [36]. Based on Suh et al.'s studies, the primary cognitive reactions concerning immersive technology include enjoyment, fun, pleasure, arousal, and dominance [214].

Topics such as the degree of immersion that the user experiences while using immersive technologies have been extensively explored. There are three different established stages of immersion: the first stage can be described as "engagement" within the immersive experience, the second stage is "engrossment" [23], and the final stage is called total immersion [23]. In the VR context, immersion usually refers to spatial immersion, where the environment and the stimuli are realistic [87].

When studying user's experiences within these AR and VR environments, questionnaires are typically used [214]. According to Martinez et al., in an immersive environment, self-reported assessments are the most common assessment tools to understand the user's "perception" [151]. Despite the usefulness of such data, there are limitations associated with self-reported data. According to Tullis and Albert, users are inclined to provide more positive feedback when interacting directly with the designer/researcher [229]. This can be mitigated by using anonymous web-based questionnaires to help resolve the issues of "social desirability bias" [229]. In order to remove the limitations and biases inherent with the subjective feedback, objective measures such as the user's physiological responses can be used to "corroborate" these subjective measures [113]. Martinez et al. discussed using biometric and physiological measurements as a useful option to study user behavior and performance within immersive environments [151].

Jordan Smith, in 2015, introduced the benefits of immersive virtual environments in observing and evaluating a user's behavioral responses, and the use of eye tracking to assess physical environment changes [202]. Eye tracking systems, with their ability to measure users' responses to the environment, primarily seek to quantify attention, gaze patterns, or user's pupil size [202]. Other measurable physiological responses that can be measured within immersive environments using the available tools are heart rate, skin conductance, skin temperature, respiration patterns, and head movements using head tracking systems [202]. These measurements can help designers/researchers understand the impact of different environmental stimuli on a user's behavior responses [202].

As discussed by Ergan et al., human experiences can be reflected in the body's physical status and cognitive conditions [80]. In the architectural design process, using physiological measurements combined with subjective questionnaires, designers can compare users' responses to different design options and their impressions of different design components. By allowing users to experience architectural design within an immersive augmented or virtual environment, their physiological responses can reveal many aspects of their mental processes, including attention and comprehension. Studying users' responses to the stimulus, in this case, architectural design can help the designers learn about the impacts of their designs. Ergan discussed a summary of physiological measurement techniques that can be applied in the field of architecture [80]. The

table below from Ergan et al.'s research demonstrates the standard biometric sensors used in architecture and the data captured using these methods.

Table 2.6. Overview of standard biometric sensors used in architecture [80].

Sensor	Primary data captured	Valence or arousal	Focus
EEG	Brain activity	Valence and arousal	Electrical brain activity
GSR	Skin conductance	Arousal	conductance response
PPG	Heart activity	Arousal	Change in HRV
f-EMG	Facial muscle reactions	Valence and arousal	Facial reactions
Eye tracking	Eye movement	Valence and arousal	Fixation zones and points of interest
fMRI	Brain activity	Valence and arousal	Blood flow of neural activity
Electrocardiogram (ECG or EKG)	Heart activity	Arousal	HRV and HR
MEG	Brain activity	Valence and arousal	Magnetic neural activity
PET	Brain activity	Valence and arousal	Nuclear imaging of the brain

Note: Table is created based on Ergan et al. [80], with modifications by author.

The literature provides examples of how to understand a user's response to different spatial typologies, which can help designers make future design decisions based on the experience. Spatial experiences in architecture impact a human's cognitive, emotional, and physiological conditions [80].

Subjective methods combined with objective measurements and methods can help understand the impacts of architectural form on human experiences, and the use of available wearable technologies can help collect unbiased, physiological, emotional, and cognitive data [80]. Franz et al., in a virtual-reality-based experiment, tested different design options of a room on 16 participants, and their results showed that there was a strong correlation between scene features ("proportions, area, openness") and the rated experience by users [85]. Other areas that can benefit from collecting users' physiological responses and movements within VR and AR include wayfinding and defining the architectural scale and adjacencies [103]. Understanding physiological responses can help guide the designer in making decisions on users' preferences and which direction the design can take to impact the user's experience positively.

2.3.3 Physiological Assessment Methods

Using physiological measurements can help understand the user's behavior and performance more deeply within an AR and VR environment [151]. User's emotions experienced while exposed to stimuli results in responses in the human body, and it is these functions that can be used in devices to measure human physiological responses [151]. These devices can assess different cognitive and emotional conditions [151]. The table below shows the standard methods of measurement used in human behavior research and their associated biometric signals and sensors.

Table 2.7. Standard measurement techniques in human behavior studies [149].

Technique	**Biometric Signal Measured**	**Sensor**	**Features**	**Psychological or Behavioral Construct Inferred**
electrodermal activity	Changes in skin conductance	Electrodes attached to fingers, palms or soles	Skin conductance response, tonic activity, and phasic activity	Attention and arousal
heart rate variability	Variability in heart contraction intervals	Electrodes attached to chest or limbs or optical sensor attached to finger, toe, or earlobe	Time domain, frequency domain, non-linear domain	Stress, anxiety, arousal and valence
electroenceph alogram	Changes in the electrical activity of the brain	Electrodes placed on the scalp	Frequency band power, functional connectivity, event-related potentials	Attention, mental workload, drowsiness, fatigue, arousal and valence
facial expression analysis	The activity of facial muscles	Camera points towards face	Position and orientation of head. Activation of action units	Basic emotions, engagement, arousal and valence
eye tracking	Corneal reflection and pupil dilation	Infrared cameras	Eye movements (gaze, fixation, saccades), blinks, pupil dilation	Visual attention, engagement, drowsiness and fatigue

Technique	Biometric Signal Measured	Sensor	Features	Psychological or Behavioral Construct Inferred
functional magnetic resonance imaging	Concentrations of oxygenated vs. deoxygenated hemoglobin in the blood vessels of the brain	Magnetic resonance signal	blood-oxygen-level dependent	Motor execution, attention, memory, pain, anxiety, hunger, fear, arousal, and valence

Note: Table is created based on Marín-Morales et al. [149], with modifications by author.

According to Martinez in their literature review on the technologies used to measure a user's physiological responses in immersive environments, the most common methods used are heart rate measurements and skin conductance respectively [151]. Other standard methodologies include respiration rates, EEG, EMG, Blood Flow Velocity, Salivary Alpha-Amylase, and eye tracking [151]. They categorize these methods into standard technologies, original, and non-typical.

2.3.4 User Feedback and Participatory Design

Participation in the design process can help a community reach democratic decisions regarding designs in their community, improve the quality of proposed designs, and avoid early design mistakes by including people impacted by those decisions [213]. As discussed by Yan and Tamke, in the architectural design field, designers usually do not seek feedback until they have completed an initial design and associated visualization [243]. Community participants in the design process would be able to share their unique wisdom and experiences about the built environment and help designers reach the best decisions throughout the different stages of design. Users in the community can feel empowered by this participation and feel ownership and control of their built environment. In addition to including non-experts, receiving feedback from peers and experts in the design field can help improve design solutions. There are contemporary examples of architectural design projects with negative environmental impacts that the inclusion of other experts and community members could have resolved. The 20 Fenchurch Street office building's highly reflective facade in London, the Museum Tower's glare issues affecting Nasher Sculpture Center in Dallas, and Foster+Partners's recent Tulip project are examples of such missed opportunities that could have been resolved by a wider range of experts and non-experts in earlier stages of design [278].

While post and pre-study qualitative data collections can be useful for understanding participants' concerns and preferences, involuntary responses can be recorded and used to analyze and understand the design's impact and drive design decision-making. Users feel various emotions when interacting with content, including stress, excitement, or surprise, and their attention to different aspects of our designs can provide insights into their thoughts beyond only subjective words [229]. Using anonymous self-assessment methods, combined with gathering users' physiological responses, can give designers a deeper understanding of the users' reactions to their

different proposed concepts. Measuring users' responses, including eye tracking, can help researchers or designers understand a user's interests in the provided content [229]. Depending on the accessibility of the technologies, researchers can reach participants remotely or in person, synchronously or asynchronously. This will enable users' voices and preferences to be heard and allow people otherwise disenfranchised from their community's design to be engaged with decisions made for their environment. Asynchronous methods can also enable higher user engagement from different locations, increasing the number of participants.

With immersive environment technologies and data collection methods, the feedback collection can go to the user rather than needing to be in a controlled environment. Immersive technologies can create different levels of realism and immersion, which helps even untrained users better understand the presented 3D concepts (e.g., spatial understanding) than traditional 2D media. Virtual reality, including affordable options such as Oculus Quest, can be integrated into the design process, where researchers and designers can bring the required equipment during community meetings or peer-to-peer design sessions. For even wider audiences, augmented reality can be accessed using smartphones and HMDs. The availability of augmented reality in many people's pockets makes it easier to include users' interactions with the design content, visualize the design content, and provide useful feedback. Using the remote capabilities of AR and VR technologies, users can participate in the design process regardless of their location, background, or expertise.

Collecting a large data set from a wide array of participants including experts and non-experts can help reduce the bias of the resulting datasets that can arise when methods such as convenience sampling are used in an experimental context. In addition to giving users wider access, gathering user feedback passively based on quantitative rather than qualitative methods, such as interviews, can help remove biases during the data collection process. In the example of the Shared-Realities project, the user's data is collected and visualized as 3D point clouds without the bias or discussion of race, age, gender, or expertise within the virtual environment [135]. A wide array of test subjects can ensure that the final resulting design is as widely accepted as possible without undue influence by the biases of a smaller set of participants selected based on convenience sampling.

A successful framework in participatory design should gather and provide overall guidance to the designers and researchers throughout the process while providing a platform for participants with different backgrounds to engage with each other and the design team. This framework can be integrated into the design process starting with the pre-design stage, with periodic inputs collected through the later design stages. Users in the early stages of pre-design can participate in collaborative activities to visualize early 3D design concepts and, using computational design methods, guide the conceptual designs in a direction they find more desirable. Participants' feedback can also get aggregated and help evolve a design decision over time. While it is important to collect feedback regardless of background to capture the correct community impact and help provide the correct questions, rubrics, and information to the users, it is necessary to seek a mix of both expert and non-expert feedback explicitly to ensure the right expertise is included in the study.

In summary, technologies and methods allowing broader user feedback collection are critical to opening up participatory design. By allowing users with different experiences and backgrounds to engage and participate in the guidance of initial massing in the early stages, we allow for a larger impact on our design decisions while empowering the end users of our spaces. This will ultimately result in a better design for the community in ways that traditional design review methods cannot.

2.3.5 Eye Tracking

Eye tracking data can indicate emotions and reactions to stimuli in user assessment studies and specify the points of interest in the subject that they are viewing [151]. The use of eye tracking is a common method in user studies due to wide accessibility to eye tracking systems and analysis tools, accuracy of the implemented systems, and availability in hand-held devices and HMDs [229]. Understanding the user's attention when interacting with the stimuli can reveal areas that require further modifications in the design or areas that they found particularly engaging [229]. An example of this application can be in web design, where eye tracking studies show unnoticed areas of the screen [229].

Earliest methods of eye tracking date back to the 1800s, when researchers used intrusive methods to track a user's eye movements which involved using anesthesia on the eyes [108]. Then, first introduced in 1900 by Dodge and Cline, methods involving photography for analyzing eye movements based on the light reflection from the fovea started being adopted as a nonintrusive method of collecting eye tracking data [108]. In the 1950s to 1970s, methods such as Electrooculography (EOG) were introduced that measured eye movements based on the electrical potential differences between two electrodes near the eye [108]. This works because the eye is positively charged on the cornea and negatively charged on the retina area so a motion would be detected as a change in measured voltage [160].

A less intrusive method of eye tracking is the video-based method, which uses a source of light, usually infrared light by measuring the "corneal reflection" of the light relative to the central point of the pupil [74]. As the user moves their eye, the position of the corneal reflection (relative to the pupil) changes [229]. This position stays consistent with small head movements [74]. Calibration is the first step in user assessment in eye tracking, and deviations less than one degree are considered acceptable in the eye tracking calibration process [229].

Using eye tracking, the researchers can study a user's thought process throughout experiments when exposed to a set of different stimuli [77]. Yarbus used an example of this method in 1967, where they asked different questions from subjects while viewing the same stimulus (painting), and the results indicated that the user's attention on the stimulus "corroborated" with the question's content [77]. Research in Functional Magnetic Resonance Imaging (fMRI) and cortical simulation shows a relationship between attention and gaze [77]. Companies such as Tobii have developed both eye trackers and screen-based systems such as Tobii Pro Glass 3, where 16 IR illuminators, four eye cameras, and scene cameras are used to create an unobstructed experience [180]. Examples of screen-based eye trackers include Tobii Pro Fusion, which can

capture gaze data up to 250 Hz and is a video-based pupil and corneal reflection tracking system that can work in dark and bright illumination mode [251].

2.3.6 Limitations

There are strengths and weaknesses associated with different physiological measurement techniques [114]. In order to overcome the disadvantages associated with each, researchers can combine different methods to gather more reliable data [114]. A common limitation of some physiological measurement devices is that they are obtrusive with limited user physical freedom. These devices often cause discomfort, limiting the duration of a session using these devices. In addition to discomfort, body-mounted devices can experience interference from unmeasured physiological responses such as body sweat [228]. Eye tracking methods using infrared lighting are unobtrusive, making them desirable [229]. However, when they are not mounted to the head, their calibration can be affected by undesirable motion by the user. To increase the precision of physiological measurements, methods such as combining physiological measurements and machine learning algorithms (e.g., Neural networks) have been used [149]. Other methods, such as questionnaires, can also help increase the precision of these results and help the researcher understand how results vary relative to the specific user [228]. The other challenge can be data analysis of the large amounts of data generated by these devices, especially when multiple devices are combined [228]. Additionally, some of these systems are limited to lab environments due to the need for access to technology. In contrast, methods such as phone applications, web-based systems, and systems that depend on smartphone cameras or webcams will be able to reach remote users. The table below shows the positive and negative aspects of different measurement methods.

Table 2.8. Positive and Negative Factors of Physiological Sensors [114].

Physiological Measures	**Advantage**	**Limitation**
Eye Position Tracking	Able to identify several cognitive states.	Does not provide information when there are no visual targets. User's extreme movements can cause issues with calibration.
Pupil Size	Able to identify several cognitive states.	Impacted by sudden changes of image intensity.
Heart rate	Provides continuous information.	Slow rise and decline in heart rate based on stimulus.
Skin conductivity	Provides continuous information.	User movements and slow rise and decline in rate can impact the results.
Temperature	Provides continuous information.	Prolonged rise and decline in temperature change.

Note: Table is created based on Ikehara and Crosby [114], with modifications by author.

2.3.7 Eye tracking in AR and VR

Eye tracking as a physiological assessment tool has often been applied to immersive environments to collect in-depth feedback from the user. Early work concerning combining eye tracking with immersive environments was discussed by Duchoswki in 2003 [74]. They implemented a binocular eye tracker fitted inside an HMD at Clemson University's Virtual Reality Laboratory [74]. A significant criticism of this system's algorithm is that it tended to underestimate fixation durations and numbers, which subsequently inaccurately reflected the attention of the user [140]. Eye tracking metrics will be discussed in the next section.

In immersive environments, FOVE and Tobii are two companies that have developed the main industry-leading eye tracking HMDs. The HTC Vive Pro series, specifically the HTC Vive Pro Eye, has Dual-OLED displays [279] along with integrated eye tracking technology using near-infrared from Tobii Eye Tracking technology to scan the wearer's eye movements [280], [281]. The HTC Vive Pro Eye has nine infrared (IR) LEDs integrated with one IR camera per eye to track the user's eye movements which can gather data regarding the user's pupil position, pupil size, eye openness, and gaze direction [280]. As described in the previous section, Tobii Pro uses infrared lights to create reflections on the user's eyes, and by analyzing the created corneal reflection patterns using images captured from the camera, the system calculates the 3-dimensional data about the eyes [281].

Apple Vision Pro is an HMD for spatial computing with glass as an "optical surface" [282]. Using VisionOS as a framework, apps can be visualized and interacted with in the user's space [282]. VisionPro is equipped with eye tracking using LEDs and Infrared cameras, enabling users to interact with objects without needing controllers [282].

FOVE 0, NeU-VR, and FOVE 480 are other HMDs with eye tracking [257]. NeU-VR has a simultaneous IR-based brain activity monitor and eye tracking system [257]. Hololens 2 is an AR HMD with four visible light cameras and two IR cameras for eye tracking [195:2]. Other devices with eye tracking capabilities include Varjo XR-3, VR-3, and Aero, with a gaze-tracking frequency of 100 Hz (default) or 200 Hz [283].

In the Hololens 2, the researchers can access gaze origin and directions at around 30 Hz, with an accuracy of 1.5 degrees around the focus target [207]. For eye tracking, Hololens 2 has two specific infrared cameras and LEDs that also allow for iris recognition [255]. The tracked eye data results can further be analyzed and visualized back within the immersive environment [207]. In AR, frameworks such as ARCore or ARKit enable the creation of AR experiences. ARKit allows for integrating iOS cameras and motions to create augmented reality experiences and combines camera scene capture, scene processing, and motion tracking [284]. ARkit can access the smartphone's front-facing camera (True depth camera) and identify scene objects, such as faces, which are considered anchors in this setting [285]. The position and rotation of the face, as well as left-eye and right-eye poses, can be detected for the scene and used for further analysis [284]. In 2020, Spang and Voigt-Antons conducted studies to assess the accuracy of the gaze tracker in ARkit 2 [208]. In user studies with 17 participants, their gaze estimation accuracy was 3.18°, or 1.44cm on screen [208].

In a literature review conducted by Li et al. [140] in publications from 2000-2019 about research on eye tracking in immersive virtual reality, areas of study for eye tracking can be divided into the following categories: “Feasibility of eye tracking in immersive environments", “eye trackers in immersive virtual environments”, and “eye movement events identification”. In 2008, Jennette conducted two experiments and questionnaires to analyze the user’s immersion level within an immersive environment compared to non-immersive experiences, with the hypothesis that the user’s eye movements will vary substantially between the immersive and non-immersive environments [118]. They used the subjective measurements to corroborate their quantitative results. Based on their results, generated from the user’s fixation data, the user’s eye movements decreased within an immersive environment, while the opposite happened within the non-immersive environment [118].

In 2009, Murray et al. conducted experiments by adding gaze directions to avatars within virtual environments and combined that with the head movements of avatars [162]. The users could better understand what direction and objects the avatar was looking at when the gaze and head movements of the avatars were combined, which helped with better remote collaboration [162].

Later, in 2019, Cebeci et al. conducted a study on emotional states in virtual environments [35]. Their method combined physiological measurements (pupil dilation, blinks, fixations, saccades, heart rate) and self-assessment questionnaires [35]. Based on their results, the experience with cybersickness (roller coaster experience) had higher saccade speeds, and the unpleasant experience had the most fixation counts, pupil dilations, and saccade rates. The experience with neutral content (campfire) was observed to have reduced the user’s anxiety based on the physiological indicators used. They used the HTC VIVE HMD with an integrated Tobii Pro eye tracker [35].

Calibration for eye tracking can be one of the greatest concerns. For the calibration process in Immersive environments, Zhang et al. in 2010 developed a calibration procedure that correlated “eye position with electrooculogram (EOG) within the 3-D” point of view and linked it to the viewer’s head position in the environment [246]. They used a dynamic 3D object moving through the virtual environment for the calibration [246].

2.3.8 Eye Tracking Metrics

Along with being able to collect the data, it is also important to consider the aspect of interpreting the results and understanding the reasons that certain eye activity is detected. Several studies are exploring the measurements of eye movements during “mental activities, visual search, and perception” [6]. When discussing eye tracking analysis, it is important to look at the terminology and metrics used in the field. Below are some important terms in eye tracking that are mainly based on Blascheck et al.’s [20] literature review on eye tracking data visualization and Holmqvist et al.’s [108] methods.

Stimulus: Stimulus (either 2D or 3D) is any visual content presented to users, which can be dynamic or static and have active or passive content [20].

Fixation: Holmqvist et al. described fixation as eyes remaining still, lasting from milliseconds to a few seconds [108]. Average fixations per the literature are around 200-300 ms. Fixation combines gaze points concentrated on a specified time, duration, and area. Common metrics related to fixation include fixation count, duration, and position [20]. Different methods to calculate fixations, as discussed by Salvucci and Goldberg [188], can be categorized as velocity-based, dispersion-based, and area-based algorithms. An example of a dispersion-based method is the Dispersion-Threshold Identification (I-DT). The algorithm leverages the tendency for fixation points to cluster together [188]. To define the algorithm for I-DT, a "dispersion threshold" and time (duration) are needed [188]. To calculate the dispersion threshold, it is recommended to use ½ - 1 ° of visual angle, which requires the distance from the eyes to the device to be known. The other method they suggest is an exploratory analysis of data and modifying the threshold to suit best the required accuracy of the use case [188]. Based on Salvucci and Goldberg, dispersion-based identification methods usually integrate 100ms as the minimum duration, which is the threshold [188]. An example application of this is as an indirect method to measure task difficulty [114].

Saccade: Saccades are rapid eye motions from one fixation to another [74]. They typically take between 30-80 ms. During this time, we are blind [108]. Saccades can happen without stimuli but indicate a lack of fixation [108].

Smooth pursuit: Smooth pursuit happens when the eyes slowly follow the stimuli [108]. They are relevant when the stimulus is moving, and eyes can match the target's velocity depending on its velocity [74]. Smooth pursuit can indicate a moving fixation on a non-moving target, which otherwise would be hard to track with fixation alone.

Scan path: A sequence of fixations and saccades is called a scan path, showing user behavior and attention information [20].

Area of Interest (AOI): Another topic in eye tracking is area of interest (AOI). One of the most common ways of measuring visual attention in eye tracking is on "specific areas of regions of a scene" [229]. In many research projects, such as studying the design of a web page, they are mainly interested in certain things on the website, not the holistic attention on the page [229]. Heatmaps are a usual method of visualizing attention in AOI, which we will discuss next.

As discussed by Li et al., since classic eye tracking studies focus on 2D content, gaze points' representations were mainly in a two-dimensional space, which is hard to extrapolate to a 3D environment directly [140]. Therefore, there has been research on identifying specific methods and gaze validation strategies for immersive environments [140]. In the literature, states including mental states, mental load, immersion, and perception of virtual environment in immersive VR environments are studied using eye tracking [184], [102], [130], [112], [38], [177], [119], [239], [118]. In 2006, Desurmont et al. [63] proposed a system that identifies the user's point of view in real-time and renders it correctly in their display. It can also detect faces and the 3D position of the eyes in the faces and track them. In another study, Soccini [203] suggested a gaze estimation system for virtual reality environments by combining image features and head movements as input in a Deep Convolutional Neural Network to estimate the 2D gaze points.

Duchowski also discussed a velocity-based algorithm for fixations in 3D, which was criticized for "underestimating fixation durations and number" [140]. In 2013, Diaz et al. discussed algorithms for calculating eye gaze angles in the virtual environment using a monocular eye tracker in a 3D environment [64]. Their angular distance calculation between the gaze and virtual objects helps identify fixations, saccades, and pursuits [64]. In 2019, Al Ghamdi and Alhalabi applied an eye fixation algorithm on eye gaze movements recorded within immersive VR and compared the results with the standard frame-by-frame version [3]. They found a strong correlation between the two analysis methods, where the 2D methods combined with raycasting and event times can be applied in the 3D environment [3].

2.3.9 AR and VR Feedback Collection and Visualization in Architecture

Eye tracking has the potential to provide several cognitive state measures, including attention to the stimuli under study [114]. Researchers can evaluate a user's cognitive load by assessing attention maps to find areas with the highest fixation in the stimuli [114]. However, these solutions often break down when the stimuli are dynamic, where fixations may not be sufficient for analysis purposes [114]. As seen in the table below, despite many possibilities of eye tracking, eye tracking can have certain limitations.

One of the difficulties when conducting user studies using eye tracking is the user's potential movements during the process. Outside lab environments, without proper observation of users, the calibration can easily become invalid. One of the most feasible assessment methods that can be combined with eye tracking to corroborate the measurements is using pre and post-study self-assessment questionnaires to understand the user's experience in more depth. Self-assessment methods have some drawbacks, such as user's issues articulating their opinion or a tendency to provide overly positive feedback to the researchers [229].

Adding objective physiological measurement into the process can help researchers reach better conclusions in the research studies [118]. When studying a user's emotions towards a stimulus, it is difficult only to use one form of physiological measurement to make conclusions. Therefore, it is often necessary to combine different measures, such as skin response and facial expressions, in user assessment studies [151]. In addition to the benefits of combining different measurement devices, some of the physiological measures such as skin conductivity, relative body temperature, and heart rate cost lower compared to other methods of measurement, which makes them more accessible for researchers to combine in the user assessment process [114]. These low-cost methods can be easily combined with eye tracking and provide further information about the user's state while looking at the stimuli. One of the disadvantages of combining many methods can be the data processing and ability of the researchers to correlate the data from the different sensors by the researchers [228].

Table 2.9. Eye tracking related measurements and its strengths and weaknesses [114].

Physiological Measures	Secondary Measures	Potential Cognitive Measures	Strengths of the measurements	Weakness of the measurement
Eye Position Tracking	Gaze Position, Fixation Number, Fixation Duration, Repeat Fixations, Search Patterns	Difficulty, Attention, Stress, Relaxation, Problem Solving, Successful Learner, Higher Level of Reading Skill	Able to determine several cognitive states during a task	Does not provide information when there are no visual targets. Rest periods go unmonitored. User's movements can cause issues with calibration
Pupil Size	Blink Rate, Blink Duration	Fatigue, Difficulty, Strong Emotion, Interest, Mental Activity - Effort, Familiar Recall, Positive / Negative Attitudes, Information Processing Speed	Able to determine several cognitive states during a task	Affected by sudden changes of image intensity

Note: Table is created based on Ikehara and Crosby [114], with modifications by author.

While pupil size is one of the metrics used in the literature to analyze user's subjective preferences, there are some inconsistencies in the results by different researchers, which makes metrics such as pupil size less reliable in measuring user preference if used alone [95]. Furthermore, eye tracking measurements such as fixation durations can be higher based on the positivity or negativity of the emotional stimuli (unpleasant content will often attract higher fixation duration). In contrast, fixation counts can be higher on positive content as the eye may return to the positive content [95]. Although single physiological measurements can help reveal a user's state, such as preference, combining methods such as eye tracking and EEG can help with a comprehensive analytical method in understanding the user's preference and adding more reliability and higher coverage of the data [95]. A combination of EEG and eye tracking measurements are common in user physiological assessments. An example of combining EEG and eye tracking is investigated by Seren's [286] SMI eye trackers combined with EEG devices (Emotiv) research [287]. The combination of EEG and eye tracking allowed the researchers to study the changes in eye tracking data and brain waves simultaneously, and it included consistent data with timestamps [287].

In spatial design research, EEGs can help provide insight into human brain activities [80]. EEGs like Emotiv are easier to integrate and install into a project than functional magnetic resonance

imaging (fMRI) methods [80]. However, EEG data analysis can be complicated based on the unsettled scientific interpretations associated with activity in different parts of the brain [80].

Despite the advantages of using eye tracking combined with EEG, researchers such as Eckstein et al. find eye tracking is a better choice than incorporating EEG and fMRI for measuring brain function indirectly [77]. Testing in eye tracking settings can happen more naturally since the user can be wearing a head-mounted display with integrated eye tracking, so they cause less discomfort compared to settings for fMRI [77]. With recent technologies that include both eye tracking and EEG, this issue can be resolved. In addition, eye trackers are easily portable and can even be done with smartphones, which makes it possible to reach a wide range of participants and allows for a diverse and large pool of participants [77]. Eye trackers have a fast calibration procedure for user testing [77]. Despite the portability of these devices, the camera-based methods can limit research to lab settings or areas where the researchers can reach participants, ruling out remote user testing. The availability of applications that work with the front depth camera, such as Hawkeye, has made user experiments and eye tracking easier and more accessible [77]. Hawkeye can be useful in mitigating difficulties associated with remote eye tracking. However, they tend to have issues such as having poor calibration or losing calibration during the test due to a large head or phone movement. This can result in captured data that is hard to analyze or use by the researcher, which they will not be aware of until after the test is executed.

As described in this section, there are many advantages to using a combination of biosensors and methods along with eye tracking. The results can be much more conclusive as several sources can corroborate the same result. There is a higher reliability for collecting data from the test if one sensor's calibration is off, and the sensors can collect full temporal coverage of the experiment where one sensor may not be able to alone. On the other hand, using certain methods and combinations can become pricier and require more complex test setups, which will require the test to be executed in controlled environments. With the increased need and community outreach benefits afforded by remote testing and collaboration, not all measurement methods can be combined while achieving large groups of participants. This means that for many use cases, running with only eye tracking as the only biometric sensor can be most beneficial when reaching a wide audience outweighs the benefits of higher-quality data. The methods for capturing and then visualizing the feedback are important for integrating AR and VR into a design setting. Since architectural design in AR and VR is mainly related to 3D content, first important methods of visualization of eye tracking will be discussed, including visualizations in 3D environments based on literature, and then examples of collecting and visualizing user feedback using AR and VR technologies specifically in the architectural design field will be presented.

Gaze data collected during eye tracking can be point-based or Area of Interest (AOI) based. In the point-based category, the evaluation is based on the "overall eye movements" and their "spatial or temporal" aspects [20]. On the other hand, for AOI, the evaluation focuses on specific areas of the stimuli based on the research interests. Based on a literature review conducted by Blascheck et al. on visualization of eye tracking data, visualization methods can be divided into different categories: statistical graphics, static vs. animated visualizations, 2D versus 3D, in-context versus not in-context visualizations, and Interactive versus non-interactive visualizations [20]. Visualizing eye tracking data can be based on the recorded point coordinates (x,y) that have

both a spatial and temporal aspect [20]. A common visualization method of eye tracking is overlaying the fixation position on the stimulus, which dates back to 1958 [20]. An attention map (or heat map) is the accumulation of fixation positions [20]. Some conventional methods of visualizing attention maps include luminance maps, 3D landscapes, and 2D topographic maps. Scanpaths use "consecutive fixations" and connect them through lines that represent saccades, which are considered a spatiotemporal visualization method [20]. Usually, fixations are represented as circles, and the radii depend on the duration of each fixation [20].

In order to represent eye tracking results in 3D virtual environments, Stellmach et al. [212] in 2010 explored visual attention in 3D virtual environments and proposed three methods of gaze visualization: superimposed 3D scan path, 3D attentional maps, and models of interest timelines. To visualize scanpaths, they proposed a method contrasting to Duchowski [212]. They had a monocular scan path that showed the intersections of gaze rays and a 3D model. In their method, they used both a sphere-based method, as well as a conical fixation method where the orientation of the cones represented the gaze direction; their positions represented gaze positions, cone base sizes indicated fixation duration and cone height represented the distance of the gaze [212]. Another visualization method is a Three-dimensional Attentional Map. In the method discussed by Stellmach et al. [212], they proposed three methods of visualizing attentional maps in 3D that represented a "superimposed aggregation of attention" on 3D objects: projected, object-based, and surface-based methods. They also introduced a prototype gaze analysis tool called SVEETER that allows for loading 3D models and eye tracking data for visualization [212].

An example of eye tracking data representation in immersive VR was by Herbelin et al. [101], who used "gaze-map chromatic gradient coding" to study a user's behavior in a therapeutic context. Although this method is not applied in an architectural design project, the 3D environment, and the methods to visualize the eye tracking data can apply to architectural design.

Yan and Tamke explored developing a spatial assessment and design methodology utilizing AR. In their 2021 publication titled *"Augmented Reality for Experience-centered Spatial Design"* [243], they discuss the differences between spatial design and product design, where user interaction data with the prototypes are collected and processed to improve and optimize the design [243]. They proposed a design methodology that allowed users to experience design in the early stages [243]. While in the experience, their interactions, movement, and attention within the scene will be collected [243]. Next, the gathered data will be visualized so the designers can understand problems such as navigation and efficiency of their design [243]. The user's HMD's head tracking is collected and visualized as representative of the gaze direction in their project [243]. Based on their results, the areas where intersections happen within the geometry attracted the most attention [243]. They used 3D spheres to visualize the user's navigation within the scene. By visualizing different participants' paths, they found differences in interest for each displayed artwork, including their fixation duration [243].

In other research, Hermund et al. conducted studies where participants navigated through a virtual space in VR, and they used self-assessment in the form of questionnaires, as well as tracking user's positions and their head directions [103]. The difference between the two experiments was that one included markers to help guide the users through the space, and the other was an

unguided experience [103]. Based on their results, users took similar routes regardless of markers within the scene and both groups had similar attention patterns [103]. Their questionnaires asked subjective questions regarding users' experiences of individual and overall building experiences, and their results showed similarities between both user groups [103].

Shemesh et al. in 2016 explored the relationship between the geometry of space and human emotions using EEG technology in Virtual environments [199]. In their experiments, they combined quantitative and qualitative methods of analysis to study the user's experience in four different architectural spaces [199]. Their participants included experts and non-experts, and the two groups showed a significant difference in their preference for curvy spaces [199]. Analyzing the data proved that the brain activity in each space type was different and most brain activity occurred within the first 2 seconds of entering each space [199]. In 2021, Shemesh et al. continued this study by assessing user's emotions within different geometric spaces in virtual reality [198]. This study combined different physiological measurement devices, including EEG, GSR, and eye tracking. Based on their results, a user's positive and negative emotions could be measured with geometric changes to the space [198]. Larger symmetrical spaces positively affected users, with results showing higher emotional changes within the non-designer group [198].

In 2017, Krietemeyer et al. explored gathering and analyzing user feedback in virtual reality [135]. In this research, they developed a computational workflow that combined VR and depth sensing for a user-centered design [135]. Users' point cloud data was gathered and could be accessed in real-time by the designers to observe the users' behavior remotely [135].

As discussed in a previous section, Ergan et al. proposed combining VR experiences with biometric sensors to measure human responses to the architectural features (high and low illuminance levels, symmetry, and asymmetry) [80]. Their experiments included EEG, GSR, EMG, eye tracking, and photoplethysmogram (PPG) [80]. PPG uses infrared light to measure skin blood flow [79]. In their specific case, since their focus was on stress and anxiety in architectural spaces, they focused on using EEG, GSR, and PPG [80].

Other topics investigated using immersive environments include wayfinding in architectural and urban design. In 2008, Edelstein et al. investigated the brain activities of architectural design elements and the concept of wayfinding to understand the human responses associated with wayfinding and re-orienting within the spaces [78]. They used a 360 immersive virtual environment (CAVE) as the environment for the experiment and used EEG to record users' brain activity. Based on their results, visual cues were more subtle over time as the users experienced the unknown environment. Finally, a user's visual engagement with the environment was studied by Simpson et al. using eye tracking within an urban environment [201]. Using eye tracking, Sussman and Ward studied the user's attention on Boston City Hall [254]. In Sussman and Ward's studies on City Hall, users' fixations were located on the building's edges, people, and vehicles within the space, and only 75% of the users looked at the building directly after the first six seconds of the exposure to the building [254]. In 2024, Guo et al. used EEG signal to optimize facade design using EEG metrics, including alpha, beta, gamma, delta, and theta waves as objectives of the optimization process [96]. Their system was designed in VR, and used an NSGA-II algorithm for the optimization process [96].

Chapter 3: Methods

InsightXR is an AR platform that aims for the inclusive participation of users in the early-stage design process regardless of users' physical location while exploring generative design as a method. The increased need for remote design and collaboration was the motivation behind the design of InsightXR. Integrating immersive technology, such as augmented reality, into a user's participation in the early-stage generative design process can help with immersion into the design and foster a better understanding of the project's scale and overall geometry.

InsightXR uses a Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to generate design alternatives inspired by how architects start early massing studies using Operative Design [13]. The Operative Design process in this research is based on the volumetric operations discussed by Mari and Yoo [148] and uses the operative NSGA-II algorithm which will be discussed in more detail in Experiment 1 [13]. InsightXR focuses on applying Operative Generative NSGA-II, using an immersive experience by integrating the algorithm in InsightXR [13]. The design operations explored in this research focus on displacement and subtraction by using expansion and carving operations as the main verbs applied to their proposed operative generative design [13], [148].

InsightXR enables users to visualize 3D designs and provide feedback on those 3D designs using augmented reality (AR) technology asynchronously. The early stages of this research explored VR technology in addition to AR. The VR component was discontinued due to its dependency on HMDs and the importance of accessibility in this research. The proposed research focuses on a user's subjective feedback and preferences, where qualitative feedback impacts the generative process through merging with quantitative data.

In InsightXR, users are divided into two categories: experts and non-experts. Within the application, users are prompted to choose between designer (expert) and reviewer (non-expert or expert) Roles. As a designer, a user can define new generative design projects, participate in a generative design project as a reviewer, or provide feedback on single massing projects.

When providing feedback on a single massing project, users' feedback is collected using two methods: direct feedback as markups (areas with positive, negative, and neutral feedback) on the 3D model and the user's eye tracking data (attention maps and 3D fixation points) [212]. A single design is presented to users, and their opinions regarding that model are collected within AR. User's qualitative feedback and eye tracking data can then be visualized back into Grasshopper for further investigation by designers.

When participating in a generative design project as a reviewer, users' feedback is collected using two methods: direct feedback as a numerical rating value (integers from -3 to +3, with -3 being the least desirable and +3 the most desirable massing), and the user's eye tracking data (attention maps and 3D fixation points) [212]. This process will be further discussed in Section 4.3. In this chapter we will discuss the proposed methods used in the augmented and virtual reality platforms.



Figure 3.1. InsightXR's AR application concept.

3.1 InsightXR Early-Stage AR Design

The early stages of InsightXR, the AR platform focused on creating a platform to allow for providing feedback on a single 3D design. Users' feedback in this process is collected using two methods: direct feedback as markups (areas with positive, negative, and neutral feedback) on the 3D model and the user's eye tracking data (attention maps and 3D fixation points) [212]. In this process, a single 3D design is presented to users, and their preference is collected within AR. User's qualitative feedback and eye tracking data can then be visualized back into Grasshopper for further investigation by designers.

In this process, within the Grasshopper environment, users will design their 3D models parametrically and will upload their final models to the Firebase Storage using the designed components in this research using pyrebase in Python [288], [40]. These components also allow users to upload related images and text-based project descriptions from Grasshopper to Firebase Storage. Once the designer uploads their 3D design, the AR platform can visualize the 3D models. Early stages of InsightXR use Google Firebase to allow users to log in to the experience, as well as store the data (eye tracking, device movements, time steps of each, as well as user's marked-up areas) [289]. The users can then choose from the uploaded models available to view and start their review session. Figure 3.2 shows the login process, and Figure 3.3 shows the overall AR application.

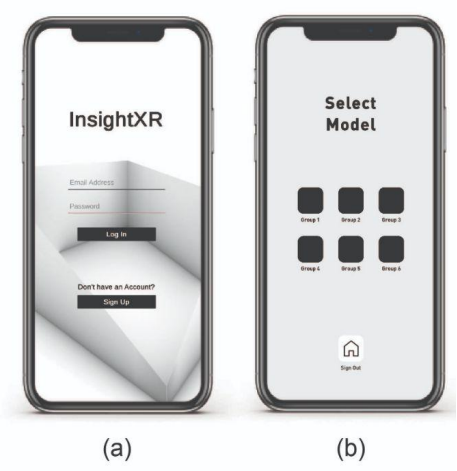


Figure 3.2. a. Login page in the AR Application; b. The model selection page.

The AR environment is built in Unity and uses the AR Foundation framework [290]. We integrated the ARKit XR Plug-in to enable ARKit to use the iOS device camera and motion features to create the AR scene [291]. Users can view the model based on the layers defined in Grasshopper, and they can provide feedback on the design using our feedback tool. In the proposed workflow, users can markup areas they find successful (using the positive button) and unsuccessful (using the negative button) based on the provided rubric by touching the areas visualized on their screen. The user can also markup areas that attract attention but evoke neutral feelings.

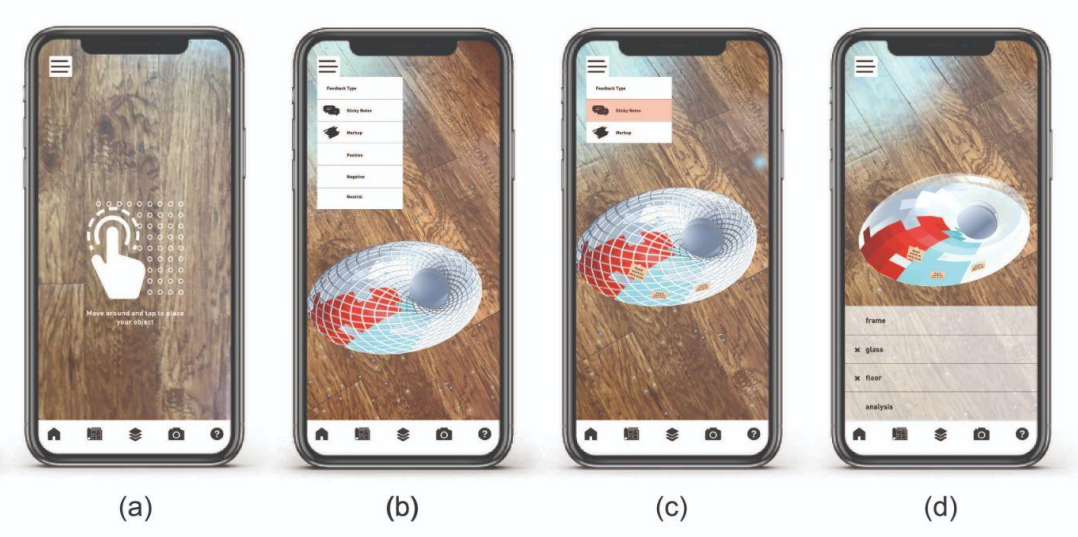


Figure 3.3. a. The user places the model using the point cloud method; b. The user can markup areas of the model with positive, negative and neutral feedback; d. The user can turn the layers of the grasshopper file on and off.

As mentioned before, during the experience, the application collects the user's eye transform (position and rotation) data. In ARKit, the face is considered an anchor in the physical environment. Using the iPhone's front depth camera, the position and rotation of the face detected within the scene can be collected, and using the face, the eye transform data can be detected [291]. Eye transform is collected every 100 ms, based on fixation eye durations discussed in Holmqvist et al., with fixation lengths to be on average, between 200-300 ms. [108]. Since the tasks provided to users impact the fixation data [224], a brief is provided at the beginning of the experience and asks the users to inspect the geometry thoroughly. The viewer is asked to look at the overall geometry to reduce the bias towards only certain sides of the visualized geometry.

In addition, the anchor point transform (position and rotation) of the visualized object is also collected and saved on Google Firebase Storage. Once the user is done reviewing the project, they will click upload, and the data will be saved on Google Fire Storage. Figure 3.4 shows the transforms collected in AR. Users' interactions with the design, eye tracking data, and the design options can also be accessed by the designer after the experience is done in Grasshopper. Figure 3.5 shows an example of eye tracking visualization from our AR platform in Grasshopper.

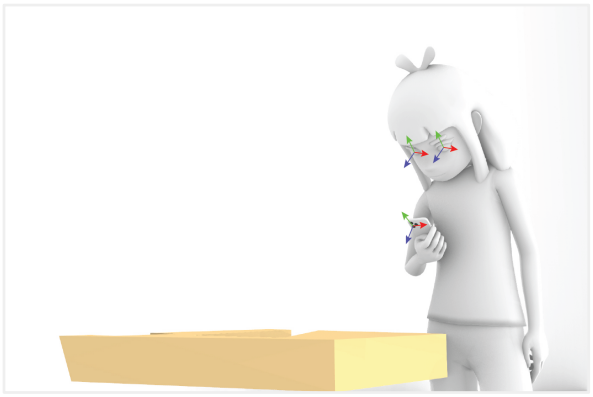


Figure 3.4. Eye tracking and transform data of the phone in AR.

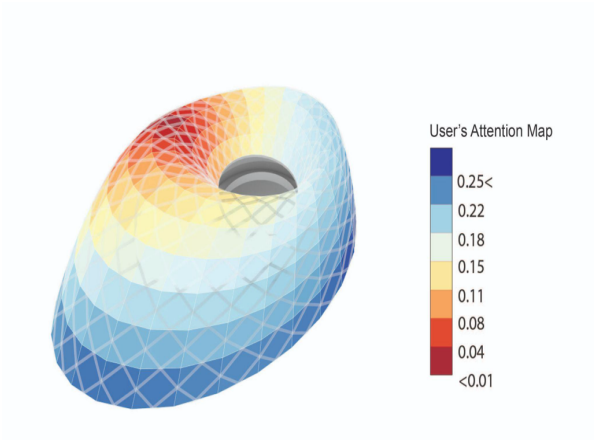


Figure 3.5. User's attention map in Grasshopper from AR application.

The final goal of this early-stage design was to create an AR-based generative process based on user feedback. In this early stage, the algorithm aimed to use the initial 3D models designed by architects as the first generation of the GA (Initial Population). In addition to gathering users' eye tracking data, the application would ask users to provide a numerical value as their rating for the design. Designers could upload up to 3 design options to their Firebase Storage from Grasshopper to start this generative design process. In the next stage, the user will be presented with new massing options based on the provided feedback and their eye tracking data using the proposed IGA [92]. The user will view and rate each provided option again. This process will continue until a desired option is found.

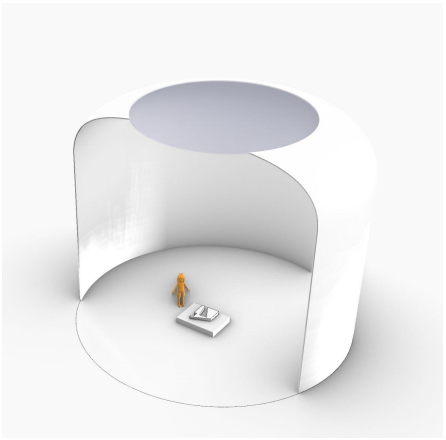
The early-stage AR prototype was informally introduced to 17 architectural designers during the development process, which helped in understanding the pain points in the overall structure and user interface and the impacts of this platform on designers. Most designers stated that they are likely to use InsightXR in a remote setting. One designer stated that their favorite feature of the application is the ability to markup the design. Others were interested in the ability to visualize 3D designs in AR remotely. Another designer found accessing user's feedback later in Grasshopper very useful. Figure 3.6 shows the initial AR application.



Figure 3.6. Initial AR application.

3.2 InsightXR Early-Stage VR Prototype

InightXR's VR experience is designed with a circular form to allow for the highest focus in the central area of the space, where the design options will be displayed to the users. This design is inspired by the C.A.V.U. by James Turrell [256]. Figure 3.7 shows the design of the VR environment. The VR experience uses the HTC Vive Pro Eye, which produces outputs including gaze origin, gaze direction, pupil position, pupil size, and eye openness [34]. The HTC SRanipal SDK in Unity is used to generate these outputs [292].



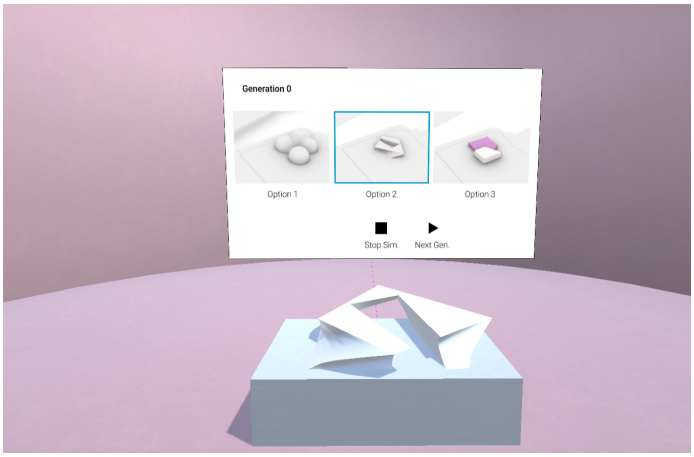


Figure 3.7. (left) 3D Section of the VR Environment. Figure 3.8. (right) User's perspective.

Similar to the AR experience, users can enter the projects remotely and asynchronously. Users' interactions with the design, eye tracking data, and the design options can also be accessed by the designer after the experience is done in Grasshopper. During the experience, similar to the AR experience, the user's eye gaze and markups are collected and uploaded to Firebase.

The goal of the VR platform was to use the initial models to act as the first generation of design, similar to the AR prototype. Users can upload up to 3 design options to their Firebase Storage from Grasshopper. Users can access other users' optimized options and add them to their population during the optimization process. Figure 3.9 shows the different steps of the proposed design participation and optimization in VR. The VR platform idea was not further explored due to the importance of accessibility in this research, which was further explored in AR.

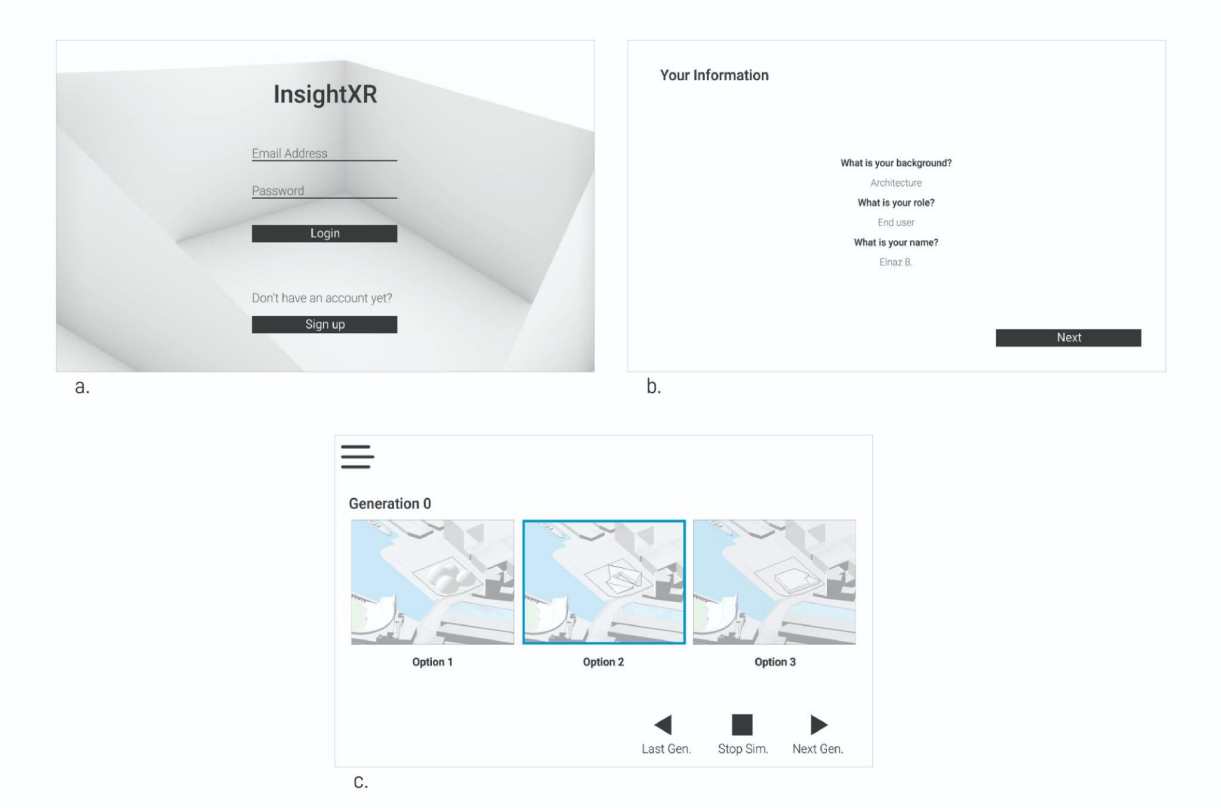


Figure 3.9. VR Environment. A. Login page; b. User information; c. The user selects a design to view and rate.

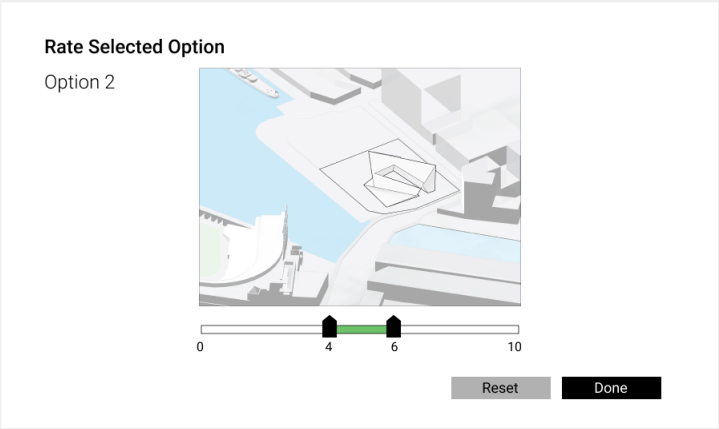


Figure 3.10. The user assigns a rating (interval fitness).

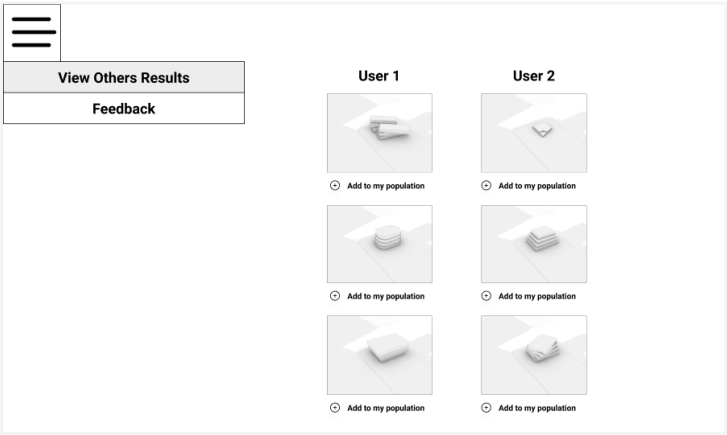


Figure 3.11. Users can access others' highest fitness solutions and add to their solution space.

3.3 InsightXR Final Design

In InsightXR, users are divided into two categories: experts and non-experts. The application makes this distinction by prompting users to choose between designer (expert) and reviewer (non-expert) roles. As a designer, a user can define new generative design projects, participate in a generative design project as an expert, provide feedback on single massing projects and view the results of their generative design project. As a reviewer, users can participate in the generative design process as non-experts. Designers can upload a site model using an .obj format of their site model into the project. Next, they define the variables in the generative design process. Finally, they set the objective functions and the operative design strategies. Once they press generate, a generative design project is created. They can then share the project to allow others to participate in a generative design process for the specific site. The figure shows the process of uploading a model to be reviewed. In order for InsightXR to work in real-time, all massing options are generated using an Operative NSGA-II algorithm and saved by a server both locally and on an Amazon Web Services (AWS) S3 Bucket [293]. InsightXR shows the massing options in real-time by retrieving them from the S3 Bucket and sends the user's rating for each massing

option back to the S3 Bucket. Figure 3.12 shows the designer's process, and figure 3.13 shows the cloud strategy used in InsightXR.



Figure 3.12. Designer's process for uploading a model.



Figure 3.13. Cloud Strategy Used in InsightXR.

In addition to creating a generative design process through uploading a site model, designers can upload a single massing model for review by other experts or non-experts. Users can mark up the massing and leave sticky notes for the designers. Figure 3.14 shows this process.

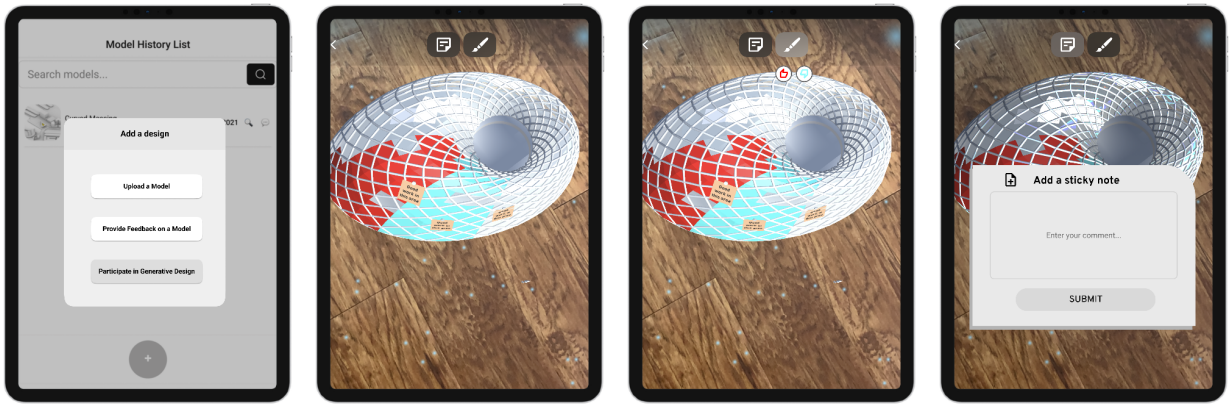


Figure 3.14. Single Massing Feedback Process.

As discussed earlier, experts and non-experts can participate in a generative design process using InsightXR. As designers (experts), users can participate by reviewing other people's shared site models and entering the specific site model using a code shared with them.

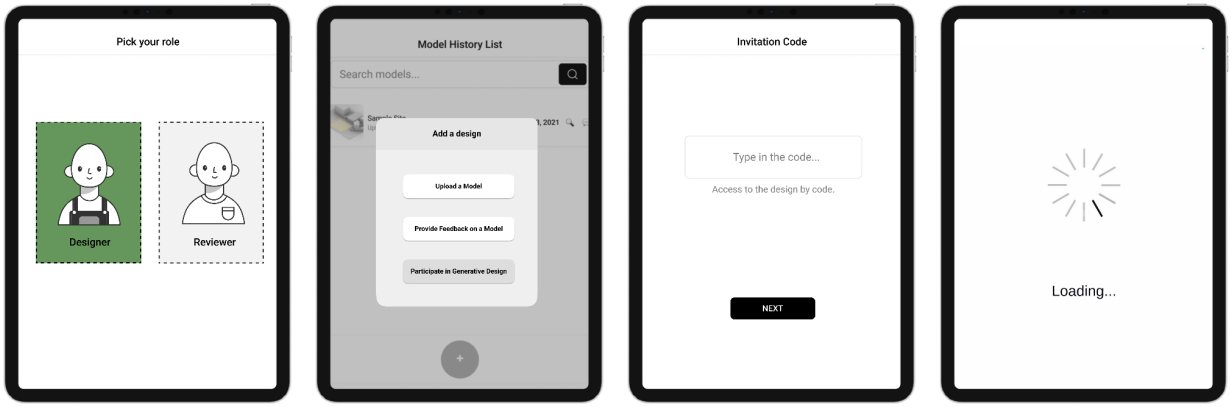


Figure 3.15. Initial Steps in Participating in a Generative Design Process as a Reviewer.

Users can also participate as reviewers in the generative design process, which will categorize them as non-experts. The main difference between designer and reviewer in this process is the ability of designers to upload a model to receive feedback on 3D designs. Three massing options are presented to users per generation when participating in the generative design process. Participants will view each design option and rate the design option. They can also leave a comment for the designer based on the design option reviewed. This process continues for ten generations. In addition to collecting users' feedback for each option, face transform, eye transform, object transform, and camera motion are collected and stored on the S3 bucket.

Overall, there are two types of experiments possible using InsightXR. The first option is conducted based on users' aesthetic preferences, where they can explore each design option in a bird's eye and first-person view and provide their rating for each design option. Figure 3.16 shows the first experiment option. In option two, the designers can provide performance metrics used by the algorithm and the definition of each performance metric (objective functions) and include that in the rating process. Users can view the values for each metric (objective values) and base their ratings on those values. Figures below show option two experiments conducted with users, which will be discussed in Section 4.3.

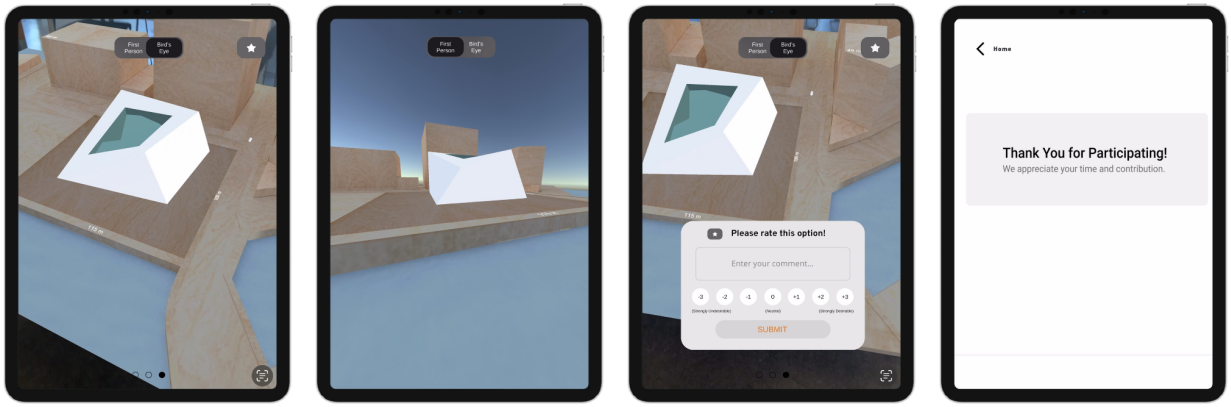


Figure 3.16. Option 1 Generative Design Process as a Reviewer.



Figure 3.17. Option 2 Generative Design Process as a Reviewer.

3.4 Results Visualization

In the early stages of this dissertation, there was an interest in understanding users' attention toward computationally designed 3D geometries and what spatial elements and features attract attention. The goal was to decipher if there is a relationship between those features of interest and users' targeted comments and whether they can be considered successful or unsuccessful elements within the 3-dimensional design. To accomplish this, InsightXR's final design aims to allow users to view eye tracking results using view vectors and 3D fixation points. The current platform includes a prototype for this feature which can be further explored in the future.

Subjective feedback, if used alone, can be limiting, and therefore, the use of objective measures such as the user's physiological studies can be used to "corroborate" these subjective measures [113]. Therefore, eye tracking data can be used as quantitative evidence to understand the user's visual and attentional process [74]. Assessing eye tracking to understand users' interests in the environment and understanding user's perceptions, preferences, and behaviors have been previously studied in the literature [202]. Such studies include Smith's research where users were asked to select specific points within the virtual environment [202]. These point locations are then recorded and mapped to correspond to the user's preference [202]. Based on Holmqvist et al. [108], during fixations, the eyes remain in the same location, lasting from tens of milliseconds or more. An average fixation can generally be assumed to last between 200 to 300 ms [108]. Blascheck et al. discussed major metrics for fixation, including the number of fixations known as fixation count, fixation duration, and fixation position [20]. InsightXR's results visualization focuses primarily on fixation counts and position. Figures 3.18 and 3.19 shows the results visualization feature and the ability to visualize both interactions with the 3D designs.

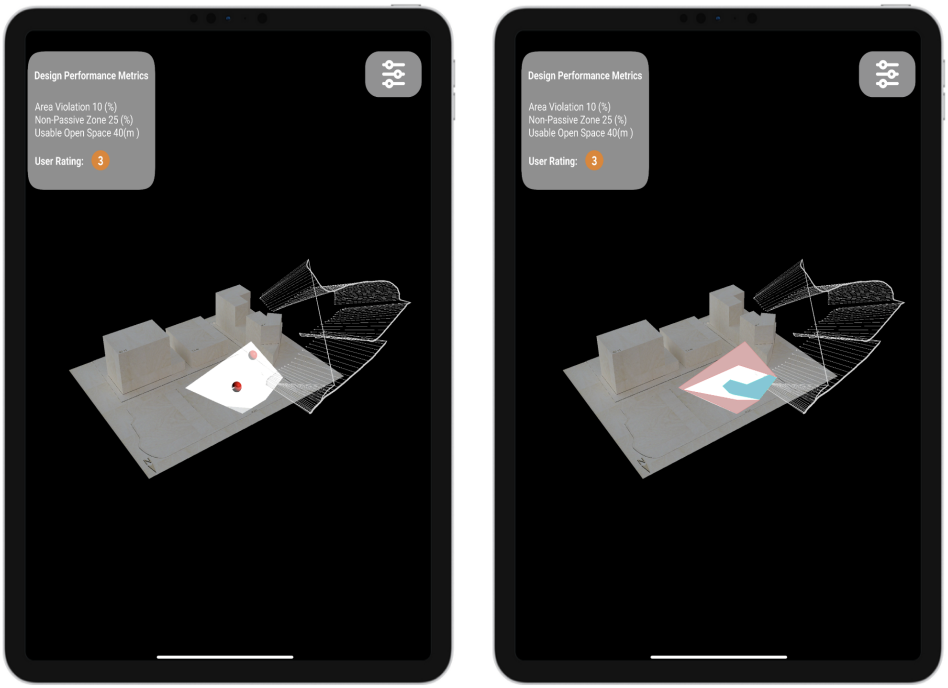


Figure 3.18. (left) Camera motion and fixation points. Figure 3.19. (right) User mark-up visualization.

Chapter 4: Experiments

4.1 Experiment 1: Operative Generative Design Using Non-Dominated Sorting Genetic Algorithm II (NSGA-II)

Massing studies during the early stages of architectural design play an essential role in determining the final building's performance across design objectives. This research aims to answer the question: How can early-stage architectural design workflows be translated into a generative design process to create valuable massing solutions? In response, a new application of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) using the Pymoo framework is proposed for the field of Operative Design. Nine experiments are discussed that test the algorithm's geometry optimization capabilities based on objective functions reflecting common architectural design goals, including Floor Area Ratio (FAR), Non-Passive Zone (NPZ), Roofs and Best Oriented Surfaces (RBOS), and Usable Open Space (UOS). Selected cases are visualized among non-dominated solutions in each experiment, demonstrating the trade-offs between different objectives while programmatically generating successful building designs. As a future step which will be discussed in section 4.2, the proposed generative design workflow will be implemented to run optimizations independently from other software within immersive environments [13].

4.1.1 Introduction

This research integrates volumetric operations within a generative design process in early-stage massing design [148]. These operations represent an evolution of volumes and can be used to accommodate different design program changes [148]. Operations in this context are based on different actions: Add, Displace, or Subtract [148]. In this research, a novel generative design workflow is proposed using a Non-Dominated Sorting Genetic Algorithm II (NSGA-II) [57] that allows for applying operative design methods to create massing options. For the optimization process, common objective functions in the architectural design process and performance-based rules-of-thumb are applied. The proposed generative design workflow is applied to early-stage massing scenarios. The goal is to optimize early-stage massing for a mixed-use building. The site is located at 74 Mission Rock St, San Francisco.

In this research, different experiments are proposed to test the algorithm's capabilities by introducing important design and shape optimization objective functions. The generated non-dominated design solutions in the Pareto fronts are then visualized and assessed [250]. In this research, the concept of constraint violation as an objective in the multi-objective optimization process is introduced for some of the proposed experiments. The objective functions, which will be further defined below, include: 1. Floor Area Ratio (FAR), 2. Non-Passive Zone (NPZ), 3. Roofs and Best Oriented Surfaces (RBOS), 4. Usable Open Space (UOS).

4.1.2 Method

In this generative design workflow, a new application of the NSGA-II algorithm is proposed that uses common architectural design operations. In the early stages of this research, the algorithm used simple GAs. The initial GAs did not result in diversity among the solutions, and due to the interest in working with multiple objectives for the future applications of this research, the NSGA-II algorithm was selected to explore further.

The introduction of design operations in generative design improves the proposed massing solutions. This process follows how architects design and, therefore, amplifies the possibility of creating more valuable massings using generative design. In addition, using the proposed application of the algorithm in this research, users can optimize massing options by considering critical objective functions in the early-stage design process (e.g., FAR). The users can decide on these objectives during the optimization process. As a proof of concept for integrating design operations in generative design using NSGA-II, carving and expansion are introduced as two primary operations. Carving is approached mainly as a courtyard concept and expansion by increasing the height of roof points.

The variables in the proposed algorithm define the movement of points comprising the overall volume. To produce the final geometry for each individual in the population, movement vectors are added to the initial points, which will produce the final geometry.

First, an initial set of points is defined, which represent an overall massing covering the proposed site. Although the algorithm can work independently from other software (i.e., independent from other 3D modeling software such as Rhinoceros), users still need to define the site dimensions and the upper and lower limits of the various variables. These variable values can be directly set inside the algorithm. Users can also set these numbers directly in a CSV format and use the content as input in the algorithm. Since the height limit is an essential part of the massing design, the upper and lower limits need to follow the height requirements for the site. Figure 4.1 shows the NSGA-II algorithm workflow.

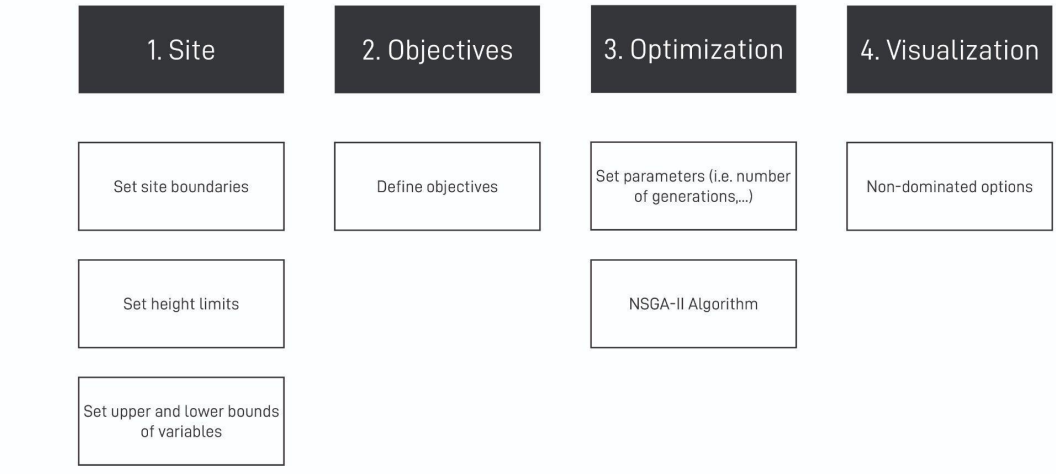


Figure 4.1. NSGA-II algorithm workflow.

Massing

The applied algorithm is designed using Pymoo [19]. Pymoo is a framework for multi-objective optimizations in Python, where algorithms can be modified and extended as needed [19]. When applying the NSGA-II algorithm using Pymoo, the multi-objective optimization is defined using several objective functions. The goal of this process is to find the best (fittest) solutions while satisfying any existing constraints associated with the objective functions [19]. Equation 1 [19] overviews a problem definition in NSGA-II as used in Pymoo. In this equation, the goal is to minimize $f_m(x)$ objectives, where N represents the number of variables, M represents the number of objectives, J represents the number of inequalities, and K represents a number of equality constraints. The lower and upper variable boundaries are represented as x_i^L and x_i^U [19].

$$\begin{aligned} & min\ f_m(x) && m = 1,..,M. \\ & g_j(x) \le 0, && j = 1,..,J. \\ s.t. \quad & h_k(x) = 0, && k = 1,..,K. \\ & x_i^L \le x_i \le x_i^U && i = 1,..,N \end{aligned} \tag{1}$$

In the proposed generative design workflow, the number of variables, the objective functions, inequalities, and the upper and lower limits of variables are decided by the designers (users) of the algorithm. To help understand this process better, this proposed process will be explored using a case study.

Since the interest of this study is the integration of design operations in NSGA-II algorithms, an important step of the optimization process is defining the variables themselves in addition to their count or limits.

To start the optimization process in the case studies, an initial set of points are defined, representing an overall massing covering the proposed site. These points are defined based on the proposed design operations and mesh creation requirements. As described earlier, the user can set these initial points directly within the algorithm.

The proposed site can be in non-rectangular shape, and the idea is for the initial defining points to maximize the coverage in the x and y direction (horizontal), with an initial height of 3 stories (10 meters high). The initial height of the massing can be any value as long as the upper and lower limits boundaries are defined correctly to result in meaningful geometries. These initial point positions are used to calculate the objective function results for generated geometries and additionally generate the 3D models of solutions for visualization purposes. This initial massing is not itself a parent used in the optimization.

To view the 3D geometries, an initial obj file is created based on the initial defining points, and the results of the NSGA-II algorithm are added to each point of the obj file and then written out to

create a new 3D model in obj format that is a delta from the initial geometry. This method can be useful when creating geometries without the need to recreate meshes in other software.

The variables used in the algorithms are the movements of these initial points that lead to resulting solutions. In designing the variables, the inspiration is based on the application of forces as vectors used in physics. Using the forces as variables made generating the new masses straightforward and intuitive. The variables defined in the NSGA-II include:

Overall defining points' movements:

- p1 (base point 1) movement
- p2 (base point 2) movement
- p3 (base point 3) movement
- p4 (base point 4) movement
- p5 (roof point 5) movement
- p6 (roof point 6) movement
- p7 (roof point 7) movement
- p8 (roof point 8) movement

Courtyard defining points' movements:

- p9, p10, p11, p12, p13, p14, p15, p16, p17, p18, p19, p20, p21, p22, p23, p24

Upper courtyard rim points' movements:

- p9', p10', p11', p12', p13', p16', p17', p20', p21', p22', p23', p24'

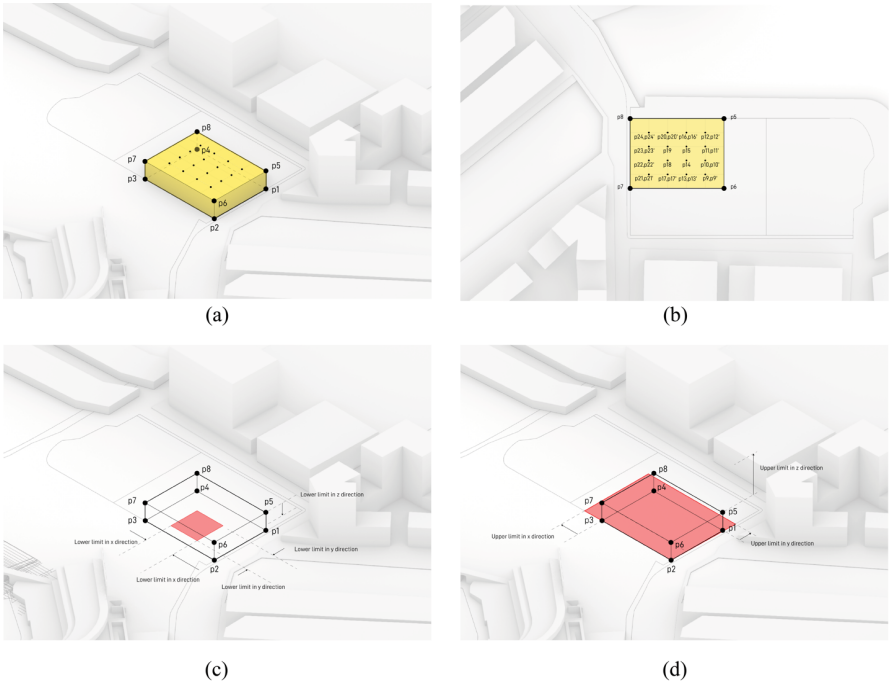
The defining points of the overall massing (initial massing) are shown in Figure 4.2. Figure 4.2a demonstrates the process of using a site to start the massing process. This site will be fully covered with an initial massing and become the basis for defining variables for the overall geometry. Once the initial site is defined, the variables are created for the massing optimization. Figure 4.2b shows how the variables are defined on the upper courtyard rim with a finer sampling. These points originally coincide with the courtyard points, and the upper rim points are denoted with the same number as courtyard points but prime). Figure 4.2c demonstrates the lower boundaries of variables, as defined by the user. The variables associated with courtyard points are dependent on the upper surface points (p5, p6, p7, p8). In the algorithm, rules are set for finding meaningful positions of the courtyard points. The courtyard points' x and y values are defined based on the position of p5, p6, p7, and p8. The courtyard points cannot go above the resulting roof surface vertically, and they can't go below the ground level. Figure 4.2d demonstrates the upper boundaries of the variables. Upper boundaries of the overall geometry are set to the boundaries of the site in the x and y directions, and the upper surface points (p5, p6, p7, p8) can go up to a maximum of 61 meters

high. The lower limit for the height of p5, p6, p7 and p8 is 10 meters (3 story building). The 61 meters height limit is selected to match the maximum height of the neighboring buildings. In addition, the initial height of this geometry can be set to any values, as long as the variables' upper and lower boundaries are defined accordingly to result in meaningful massings.

In the case studies in this research, carving (in the form of courtyard) and expansion are explored. These methods require the designer to define variables that can support a geometry with a courtyard and an expansion operation. This need necessitated the use of the upper rim points and corner points in the proposed case study. In addition, users need to define the upper and lower boundaries of the variables to create meaningful designs resulting from the proposed method. If users decide to use any additional design operations, further points need to be defined in the initial geometry to accommodate new design operations. In addition, new variables need to be defined in the algorithm to represent these new points, and lower and upper boundaries of these variables need to be decided by the designer.

In summary, the users will define a site, which will then be filled with an initial massing. The massing will include variables that need to be decided and defined by the designers, each with an appropriate upper and lower boundary. The initial massing generated in this step is called the initial parent (initialParent). In order to calculate the final position of defining points of the resulting massing (massing), the values of variables for each solution (results.X) can be added to the initial massing geometry point positions as shown in equation 2.

$$massing = results.X + initialParent \qquad (2)$$



(a) (b)

(c) (d)

Figure 4.2. (a) Initial site and variables definitions (b) Upper courtyard variables (c) Lower boundaries of variables (d) Upper boundaries of the variables.

NSGA-II Algorithm

In Pymoo, NSGA-II uses an overall GA approach with custom mating and survival functions, as described in section 2.3 [19]. The Pymoo framework in Python is used for the NSGA-II algorithm due to its customizability, focus on multi-objective optimizations for future research, and visualization capabilities [19]. In addition, Pymoo's NSGA-II algorithm is designed based on Deb et al.'s pioneering research on NSGA-II [57]. The algorithm has several components: design operation, the problem, constraints, and the optimization algorithm.

Pymoo includes submodules for mating selection, crossover, and mutation to create NSGA-II algorithms [19]. The NSGA-II algorithm in Pymoo is designed based on Deb et al.'s work on NSGA-II algorithms [57]. A crucial component of the NSGA-II algorithm is the crowded-comparison operator ($\preccurlyeq_n$). Among every two individuals, the individual with a better rank is preferred [57]. If they share the same front, the individual that belongs to the less crowded region is selected [57]. In equation 3 [57], $i_{distance}$ is the estimate of the crowding distance and i_{rank} and j_{rank} are the non-dominated ranks for two individuals [57].

$$\begin{gathered} 1. Non-dominated\ rank\ (i_{rank}); \\ 2. Crowding\ Distance(i_{distance}); \\ \text{3. Next a partial order } \preccurlyeq_n \text{ is defined where:} \\ i \preccurlyeq_n j\ if\ i_{rank} < j_{rank}\ ) \\ or\ ((i_{rank} = j_{rank}\ ). \\ and\ (i_{distance} > j_{distance})) \end{gathered} \tag{3}$$

In Pymoo, the survival selection submodule selects the surviving individuals for the next generation [19]. In this submodule, the RankAndCrowdingSurvival class is defined for survival selection within the algorithm, which is further described in equation 4 [19]. The class implements the concepts of non-dominated sorting, Crowding Distance calculation, and selection based on Crowding Distance [19].

To summarize, the equations in 4 [57] below describe the NSGA-II procedure. In equation 4, P_t represents the parent population, and Q_t represents the offspring population [57]. First, the parent and offspring populations are combined, and therefore the resulting population R_t has a size of two times of required population $(2N)$ [57]. Next, the resulting population R_t is ranked using NSGA-II's non-dominated sorting method [57]. F_1 represents the best non-dominated set, and the individuals belonging to F_1 are considered the best solutions and are prioritized in this process [57]. If possible, all individuals belonging to F_1 are included in population P_{t+1} [57]. Then the rest of individuals belonging to P_{t+1} are selected based on the rankings of their non-dominated fronts [57]. This will result in a population larger than the desired N population members. Next, crowded-comparison operator $\preccurlyeq_n$ is used to select exactly N individuals [57]. Then, the algorithm uses the operators of selection, crossover, and mutation on population P_{t+1} to create the resulting population Q_{t+1} using the make-new-pop method [57]. Finally, the generation counter is increased, and a new generation of individuals is initiated [57].

$$R_t = P_t \cup Q_t$$

$$F = fast - non - dominated - sort(R_t)$$

For ((i = 1, $P_{t+1} = \emptyset$), $|P_{t+1}| + |F_i| \leq N$, i ++)

$$P_{t+1} = P_{t+1} \cup F_i$$

$$crowding - distance - assignment(F_i) \quad (4)$$

$$sort\ (F_i, \preccurlyeq_n)$$

$$P_{t+1} = P_{t+1} \cup F_i[1:(N - |P_{t+1}|)]$$

$$Q_{t+1} = make - new - pop(P_{t+1})$$

$$t = t + 1$$

In this research, Simulated Binary Crossover (SBX) is used as the crossover method, which is based on the crossover of binary encoded variables [19]. The SBX in Pymoo was created by Deb et al.'s research on self-adaptive crossovers [62]. SBX simulates the single-point crossover on binary strings [62]. In this process, a spread factor β_i is defined as the ratio of offspring values to parent values [62]. In equation 5, offspring are shown as $x_i^{(1,t+1)}$ and $x_i^{(2,t+1)}$ and parents are $x_i^{(1,t)}$ and $x_i^{(2,t)}$ [62].

$$\beta_i = \left| \frac{x_i^{(2,t+1)} - x_i^{(1,t+1)}}{x_i^{(2,t)} - x_i^{(1,t)}} \right|. \quad (5)$$

Next, a random number between 0 and 1 defined as u_i is created and used to find the the ordinate β_{qi} [62]. The probability distribution $P(\beta_i)$ shown in equation 6 [62] is then used to generate the offspring using equation 7, where η_c is the distribution index [62]:

$$P(\beta_i) = \{(0.5(\eta_c + 1)\beta_i^{\eta_c}, if\ \beta_i \leq 1;, 0.5(\eta_c + 1)\ \frac{1}{\beta_i^{\eta_c + 2}}, otherwise. \quad (6)$$

$$x_i^{(1,t+1)} = 0.5\left[\left(1+\beta_{qi}\right)x_i^{(1,t)} + \left(1-\beta_{qi}\right)x_i^{(2,t)}\right],$$
$$x_i^{(2,t+1)} = 0.5\left[(1-\beta_{qi})x_i^{(1,t)} + (1+\beta_{qi})x_i^{(2,t)}\right]. \quad (7)$$

In this research, Polynomial Mutation (PM) is used as a mutation operator in NSGA-II [19]. In the PM selected in this research, a user-defined index parameter η_m is used [61]. Equation 8 [61] shows how to create a mutated solution ρ' from a given parent, using a random generated number u between 0 and 1 [61]. The two parameters δ_L or δ_R are calculated using equation 9 [61].

$$\rho' = \{\rho + \delta_L(\rho - x_i^{(L)}), for\ u \le 0.5, \rho + \delta_R(x_i^{(U)} - \rho), for\ u > 0.5. \quad (8)$$

$$\delta_L = (2u)^{\frac{1}{(1+\eta_m)}} - 1, for\ u \le 0.5,$$
$$\delta_R = 1 - (2(1-u)^{\frac{1}{(1+\eta_m)}}, for\ u > 0.5. \quad (9)$$

Design Operation. In order to implement the operations, a function is defined to apply carving or expansion randomly to each generated individual. In this process, carving or expansion are randomly selected for each individual. The amount of the expansion or carving is randomly generated. Based on the selected criteria, the operation is applied to the generated individuals in the population. The NSGA-II algorithm uses sampling to create the initial population. In this research, a random sampling method with a normal distribution is combined with the proposed design operations (carving and expansion) [19]. This sampling method returns the updated initial population by applying the specified Operative Designs.

The Problem. The problem in the algorithm is where the objective functions are defined. A problem in Pymoo includes the number of variables, number of objectives, number of constraints, lower bounds of the design variables, and upper bounds of the design variables [19].

Any relevant objectives can be defined within the algorithm, and the current algorithm has multiple objectives that users can choose from for the optimization process. Users can also define new objectives within the algorithm.

Constraints. There are different approaches to handling constraints in optimizations. Some standard methods include death penalties, static penalties, dynamic penalties, segregated GA, repair algorithms, and constraints as objectives. In the applied algorithms, constraints can be applied using Pymoo's integrated constraint handling methods, including constraint violation (CV) as an objective, CV as a penalty, and a parameterless approach [60].

The concept of handling constraints as objectives in evolutionary algorithms have been explored in the literature. In 1997, Surry and Radcliffe explored the constraint handling method called COMGA (Constrained Optimization by Multi-objective Genetic Algorithms) by using a vector-evaluated genetic algorithm (VEGA) and Pareto ranking [219]. Their approach viewed a constrained optimization problem as a constraint satisfaction problem and an unconstrained

optimization problem [219]. This approach dynamically changes the emphasis on constraint satisfaction and fitness value (objective function value) as the optimization progresses [219].

In 1998, Coello discussed using constraints as objectives in optimization problems [44]. Coello used a population-based multi-objective optimization method (VEGA) and treated the constraints as objectives [44]. Through this method, a fraction of the population was selected using the (unconstrained) objective function as its fitness, while another fraction used the first constraint as its fitness [44].

In 2000, Runarsson and Yao discussed an approach to balance objective and penalty functions stochastically [186]. They introduced a probability using only the objective function for comparisons in ranking in the infeasible regions of the search space [186]. Their technique was based on rank-based selection through a bubble-sort-like procedure without the self-adapting method proposed by Surry and Radcliffe [219].

The Optimization Algorithm. In the proposed algorithm, NSGA-II is used and the population size, the number of offspring, the sampling method, crossover, and mutation properties are defined.

Once the algorithm is defined, optimizations are run by minimizing the defined problems using NSGA-II. A termination criterion is set for the optimization based on the number of generations.

In order to calculate the final position of the points and create the final geometry for each massing, the individuals can be extracted from the minimization result of the defined problem. The values for each individual can be added to the initial parent values as introduced in section 3.1, using a simple summation operation to find the resulting massing. Overall, the optimization algorithm uses the steps below:

- Step 1. InitialParent is defined, which is a list representing the points defining the initial geometry created to cover the entire site.
- Step 2. The algorithm object is defined by an NSGA2 function that uses the population size $pSize$, number of offsprings n_off, and sampling which is a method generated by creating an initial sample from applying design operations to individuals. Crossover and mutation are other operators applied in the NSGA2 algorithm. Next, the results object is created by minimizing the defined objective functions as problem(), and using the defined algorithm in the previous step.
- Step 3. The results object is created by minimizing the defined objective functions as problem(), and using the defined algorithm in the previous step.
- Step 4. One of the outputs of the results objects are Pareto front solutions that can be extracted using results.X
- Step 5. In order to create the final points defining the resulting individuals (massings), results.X can be added to initialParent. It is important to note that access to the final solution positions is required in order to calculate spatial-related objective function values for every individual in the population.

The optimization procedure and the calculation method to find the final positions of massing points is shown below.

Step 1. Define the initial massing: $initialParent$

Step 2. Run the NSGA2: $NSGA2(pSize, n_off, sampling, crossover, mutation)$

Step 3. Calculate the results: $minimize(problem(), algorithm)$ *(10)*

Step 4. Extract the results values: $results.X$

Step 5. Calculate the final massing: $results.X + initialParent$

Experiments

The proposed generative design workflow is explored through case studies in early-stage massing scenarios. The goal is to optimize early-stage massing for a mixed-use building. The site is located at 74 Mission Rock St, San Francisco, CA 94158. This site is across from other mixed-use buildings and south of Oracle Park stadium in San Francisco. Figure 4.3 shows the site location.

Site limitations are important to consider during the massing design phase, and they were taken into consideration when setting the upper and lower boundaries of the variables, in addition to the selected objectives in the NSGA-II algorithm. The proposed case study in this research uses objectives associated with site and building geometry while taking into account the current needs of clients and the community in the urban context. The selected site is located in San Francisco, in the Mission Rock Mixed Use District (MR-MU) zone. There is currently no floor area ratio limit based on San Francisco zoning [294]. However, the maximum allowable height on the selected site is up to 240’ (approximately 73 meters) [294]. In the proposed case study, a maximum height of 61 meters has been used by the authors to match the maximum height of the neighboring buildings.

In addition, San Francisco’s housing demand has been a critical issue during the last few decades since the 1960s [153]. In addition to the shortage of affordable housing, NIMBY (Not In My Backyard) groups' interference results in delays in affordable housing projects [153]. Therefore, to respond to higher needs for housing and access to affordable housing in the San Francisco area and the need for higher density, the proposed research uses maximization of FAR as an objective to generate higher amounts of future housing in the massing stage, which might not be desirable in other sites.



Figure 4.3. Proposed Site in San Francisco.

Different experiments are designed to test the algorithm capabilities by introducing important design and shape optimization objective functions, then visualizing and assessing the generated non-dominated design solutions. The concept of constraint violation as an objective is introduced in the multi-objective optimization process for some of the proposed experiments using Area Violation (AV). The objective functions in these experiments include 1. FAR, 2. NPZ, 3. RBOS, 4. UOS, defined below in section 3.4.

Objective functions are important design forces that help shape the solution space, and their integration is a design decision. Selecting the objective functions, as well as variables and upper and lower boundaries of these variables, are driven by an architect's knowledge of the architectural design process combined with zoning and building regulations specific to the building site. Therefore, in the proposed method, designers need to select objectives based on project needs and context, and the applied NSGA-II algorithm discussed in this research can be adapted and utilized for different objective functions as needed. In addition, the decision on maximizing or minimizing objectives such as FAR and UOS are design decisions that need to be made by designers.

In the proposed research, sustainable design rules-of-thumb objective functions are selected, and the tradeoffs between these objectives and FAR, UOS and AV are explored. The role of these rules-of-thumb are similar to the objective functions used by Wortmann and Natanian's research where they maximized FAR and explored the tradeoff between FAR and zero energy goals [240].

Different combinations of these objectives in addition to different site dimensions are explored in multi-objective optimization processes. Two different site options are considered in the optimizations. Site 1 is 4800 m², and site 2 is 11,500 m². Figures 4.4 and 4.5 show the different site options in the experiments. Since FAR is an essential objective for an architectural design

representing the gross area of the building, FAR is used as an objective function in all the experiments. In these experiments, the goal is to maximize FAR and, where applicable, maximize the UOS. Table 4.1 shows the proposed experiments and their properties. In these experiments, a population size of 30, simulated binary crossover (SBX) with probability 0.9, and polynomial mutation (PM) are used. In 15 generations, the optimization is terminated. For maximization problems, the values for the calculations are negated.

UOS and FAR are conflicting objectives since FAR tries to maximize the built building area while UOS tries to maximize the unbuilt building area on the site. RBOS is defined based on San Francisco's climate, to maximize northern and southern facade combined with the roof area, which will be different based on other locations.

As discussed earlier, in sustainable design, different objective functions are used for estimating building performance as rules-of-thumb without using simulation tools (i.e., EnergyPlus, Radiance). These rules-of-thumb objective functions help with faster optimization processes (without simulations) and can be a useful starting point for the early-stage massing design. These objective functions achieve a fast response compared to simulations in the early stages of design [240].

Table 4.1. Experiments' definitions.

Experiment	Objectives	Site
1	FAR (11), NPZ (12)	Site 1 60 m x 80 m
2	FAR (11), RBOS (13)	
3	FAR (11), UOS (14)	
4	FAR (11), NPZ (12), AV* (15)	Site 1 60 m x 80 m
5	FAR (11), RBOS (13), AV (15)	
6	FAR (11), UOS (14), AV (15)	
7	FAR (11), NPZ (12), AV (15)	Site 2 100 m x 115 m
8	FAR (11), RBOS (13), AV (15)	
9	FAR (11), UOS (14), AV (15)	

*Area Violation (AV) in these experiments is the absolute value of 60,000 m²- Gross Floor Area.

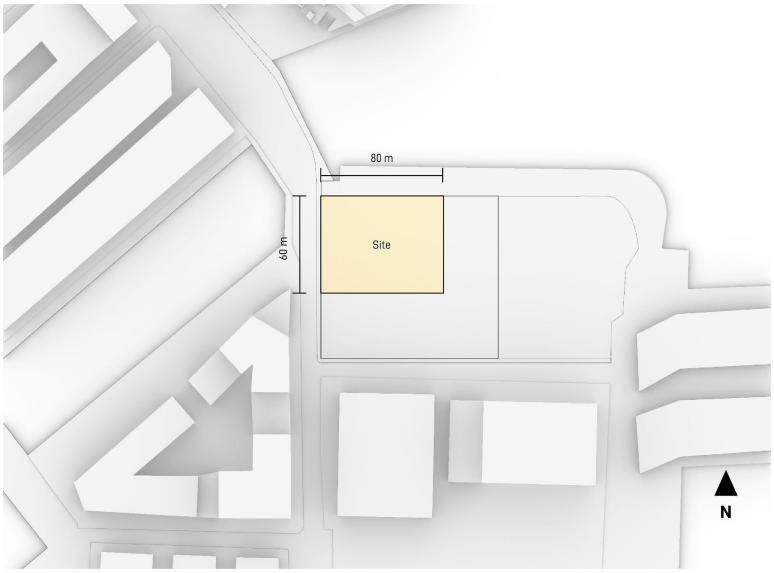


Figure 4.4. Site 1 dimensions (60x80m).

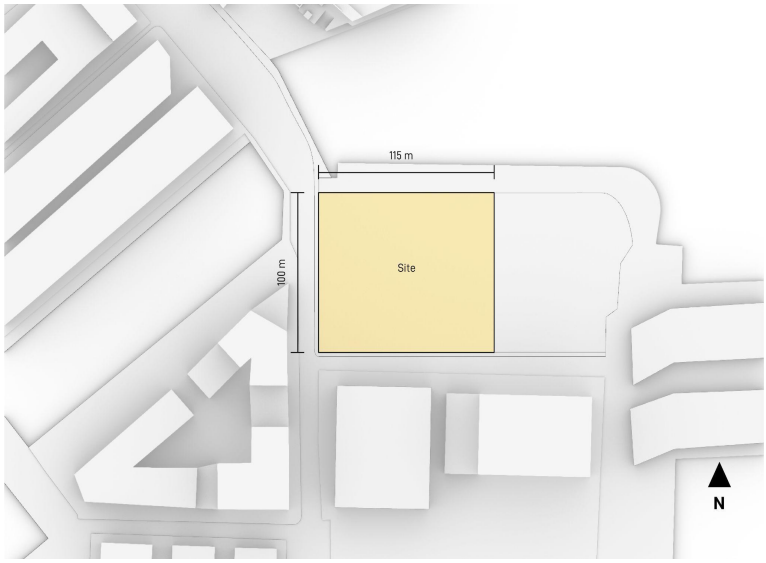


Figure 4.5. Site 2 dimensions (100x115m), a 139.6% increase in relation to site 1.

At the initial stages of this research, testing the integration of Operative Design in an NSGA-II algorithm was the main interest of our studies. Experiments 1-3 were designed to test the algorithm's capabilities and assess the generated massings to ensure the algorithm behaved as expected. To generate more diversity in the shape of generated massings at this stage, FAR was not limited in the design constraints. In the next stage of this research Area Violation (AV) was introduced as an objective, since building floor area is an important property to consider when designing massings. Using AV as an objective allowed for considering the intended floor area in experiments 4-9. Introducing AV helped make the problem definition more realistic. In future

iterations, instead of FAR, the difference between FAR and desired FAR can be used as an objective to make the results closer to the desired architectural requirements.

Objective functions

Below, the calculation methods for each objective function are defined.

1. **FAR.** FAR is calculated by adding the area of generated floor slab points in the algorithm, subtracting the area of courtyards (carving) if they exist, and dividing the result by the area of the lot. Since the goal is to maximize the FAR, the resulting values are negated in the optimization. In this objective function, GFA represents the Gross Floor Area of the building in square meters, and AL represents the area of the lot in square meters.

$$f11 = -1 * ((GFA)/AL) \quad (11)$$

2. **NPZ.** To calculate the NPZ, the floor slabs are defined on each floor and the amount of carving in the floor, if it exists, and then calculate the amount of non-passive area for each floor. In order to make the comparison of this objective function value easy to compare with other values, the result is turned into a ratio, dividing the resulting total non-passive area by the Gross Floor Area. The division results in the NPZ value.

   Figure 4.6 shows what a non-passive area (shown in red) looks like within the algorithm for the first floor, which can be applied to every floor. H represents the height of one level within the massing. In the calculations, the first floor is considered as 4 meters and the rest of the floors to be 3 meters (floor to floor). Users can change these numbers within the algorithm. In this objective function, Non-PassiveArea represents the total non-passive areas in the building in square meters, and GFA represents the Gross Floor Area in square meters.

$$f12 = (Non - PassiveArea)/(GFA) \quad (12)$$

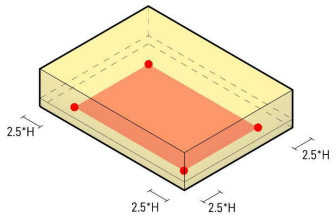


Figure 4.6. Non-passive area on the first-floor representation.

3. **RBOS.** In order to calculate the RBOS, it is crucial to consider the climate where the site is located and how it impacts the building design. San Francisco is in climate zone 3C, with a Warm Marine climate [4]. This location receives high solar radiation throughout the year, with an average annual amount of 1,832 kWh/m^2. The access to solar radiation and the

need for annual heating makes solar heat gain a useful strategy to reduce heating needs and the ability to use the radiation for renewable energy (e.g., photovoltaics) possible. San Francisco is located at 37.7° N, and 122.4° W, where the northern and southern facades combined have a higher potential for solar insolation absorption than other facades [34]. Therefore, RBOS is defined using the roof area combined with the northern and southern facade areas, subtracting the area of the courtyard from the roof [34]. Since this is a maximization problem, the result is negated in the algorithm. In this objective function, roofArea represents the roof area in square meters, and courtyardArea represents the amount of carved courtyard from the roof in square meters. In addition, surfSArea is the area of the southern facade in square meters, and surfNArea is the northern facade area in square meters.

$$f13 = -((roofArea - courtyardArea) + surfSArea + surfNArea) \quad (13)$$

4. **UOS.** In these experiments, the amount of unbuilt area is calculated on the site as UOS, and the goal is to maximize this area for improved user experience. The results are negated for this maximization objective function. Here, AL is the area of the lot in square meters, and footprint is the first-floor area of the building in square meters.

$$f14 = -(AL - footprint) \quad (14)$$

5. **AV.** The experiments introduce AV as an objective function within the algorithm. The AV objective function is the difference between the goal area (60,000 m²) and the total floor area of each individual in each generation. The absolute value is used in the calculations. However, in the visualizations of the experiments, the real numbers are shown to better understand the difference between the goal area and the resulting area. In this objective, 60,000 square meters is the goal area, and GFA is the Gross Floor Area of the building generated in square meters.

$$f15 = |60000 - GFA| \quad (15)$$

4.1.3 Results

Experiment 1. FAR and NPZ (Site 1)

The first experiment intends to maximize the FAR and minimize the NPZ in the proposed massings. In each experiment, 3-4 non-dominated solutions are selected from the results to visualize the extreme Cases of the solutions and compromises among objective functions. In experiments 1-9, Cases 1 and 3 are solutions with extreme values of FAR (Case 1 results in maximum FAR, and Case 3 has minimum FAR). In contrast, Case 2 is selected as a compromise between the FAR objective function and other studied objective functions in experiments. In experiment 1, Case 1 is one of the extreme Cases with the highest FAR and highest (worst) NPZ percentage. Case 1 is a massing with no courtyard, selected from the first generation in the

population by the algorithm. On the other hand, Case 2 has a large courtyard that reduces the NPZ percentage, while the FAR is lower than Case 1. Case 2 shows a good compromise between FAR and NPZ objectives. Case 3 is another extreme Case, where the FAR is at its lowest, and the NPZ is at its minimum. Case 3 has a large courtyard, in addition to its smaller first-floor area (footprint) that leads to the low FAR and the NPZ reduction. Case 3 is a hybrid (combination of carving and expansion), while Case 2 represents carving operation only. Among the visualized Cases, Case 2 is a successful massing with a high potential of providing access to natural light using its courtyard while providing a large area for building programs.

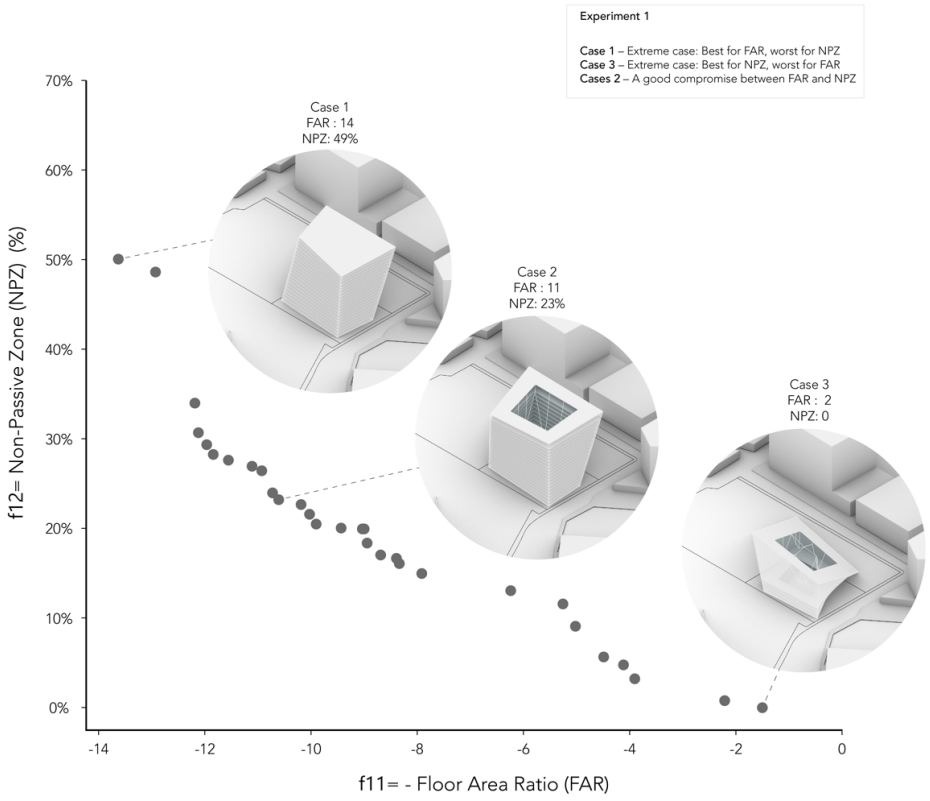


Figure 4.7. Experiment 1 non-dominated solutions.



Figure 4.8. Experiment 1 non-dominated solutions.

Experiment 2. FAR, RBOS (Site 1)

Experiment 2 explores maximizing the RBOS (the summation of the roof area, north facade area, and south facades) and maximizing the FAR. Cases 1, 2, and 3 result in a high number of floors, increasing the potential amount of facade exposure.

Case 1 has large floor plates with less carving compared to other Cases. In Cases 1, 2, and 3, the buildings are east-west oriented and therefore have more prominent northern and southern facades useful for RBOS goals. Case 2 is a good compromise between RBOS and FAR values. Case 3 is an extreme result visualized where the northern and southern facades are maximized, and the roof has a larger overall boundary compared to the first-level footprint, which results in the highest RBOS value. On the other hand, Case 3 has the lowest FAR due to its footprint on each level and the large courtyard.

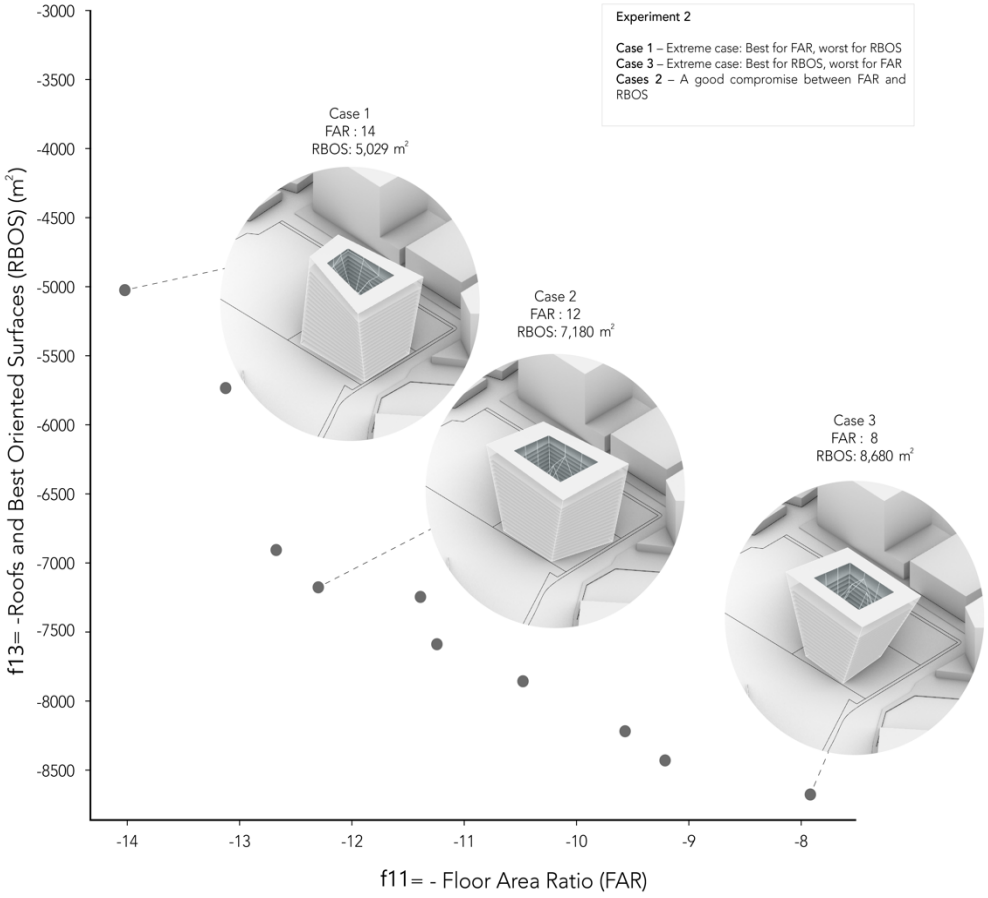


Figure 4.9. Experiment 2 non-dominated solutions.



Figure 4.10. Experiment 2 non-dominated solutions sections.

Experiment 3. FAR, UOS (Site 1)

Experiment 3 aims to maximize the UOS and the FAR. Case 1 is an extreme Case with the highest FAR compared to other solutions with a small carving (courtyard), while the UOS value is the lowest compared to other non-dominated solutions.

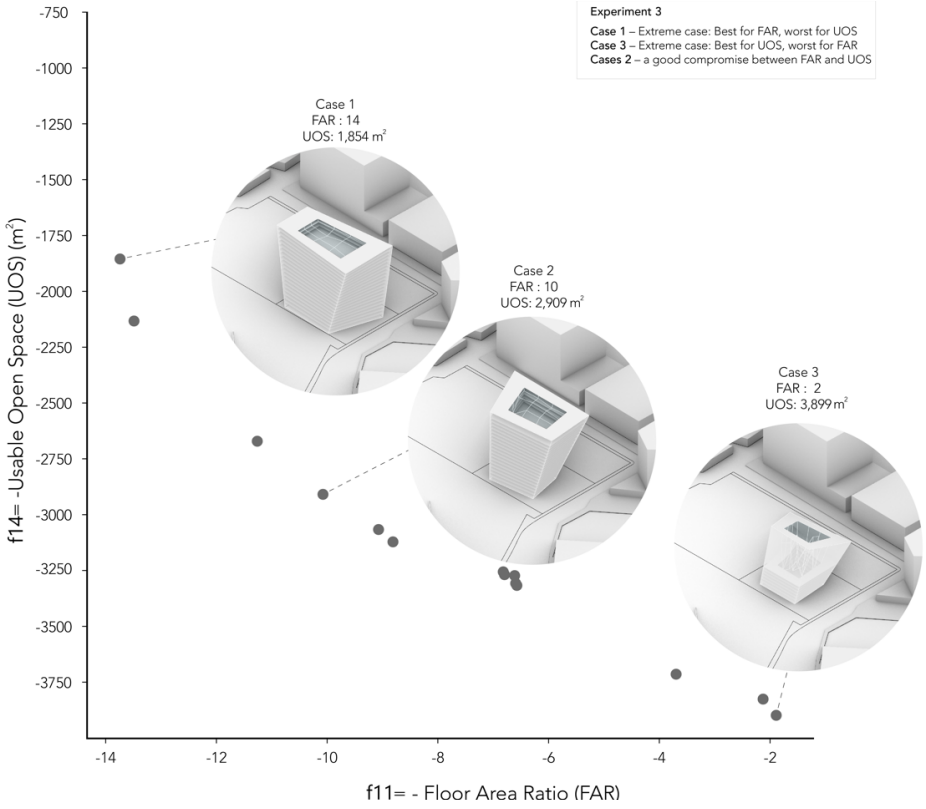


Figure 4.11. Experiment 3 non-dominated solutions.

Case 2 has a larger UOS compared to Case 1, and Case 3 has the highest UOS value with the lowest FAR value with a large courtyard. As expected, non-dominated solutions in this experiment try to provide large UOSs by creating a smaller footprint (in Cases 2 and 3). Case 2 introduces more floors within the massing despite the first-floor area (footprint) and is a good compromise

between FAR and UOS objectives. Case 1 and Case 2 are both results of carving operations, while Case 3 is a hybrid solution.



Figure 4.12. Experiment 3 non-dominated solutions sections.

Experiment 4. FAR, NPZ, AV (Site 1)

Experiment 4 uses three objective functions: FAR, NPZ, and AV. AV, as discussed earlier, represents the difference between the goal area (60,000 m²) and the Gross Floor Area (GFA). To better understand the relationship among solutions, the solutions are visualized in a 2D format using a color gradient instead of the third objective function (the AV).

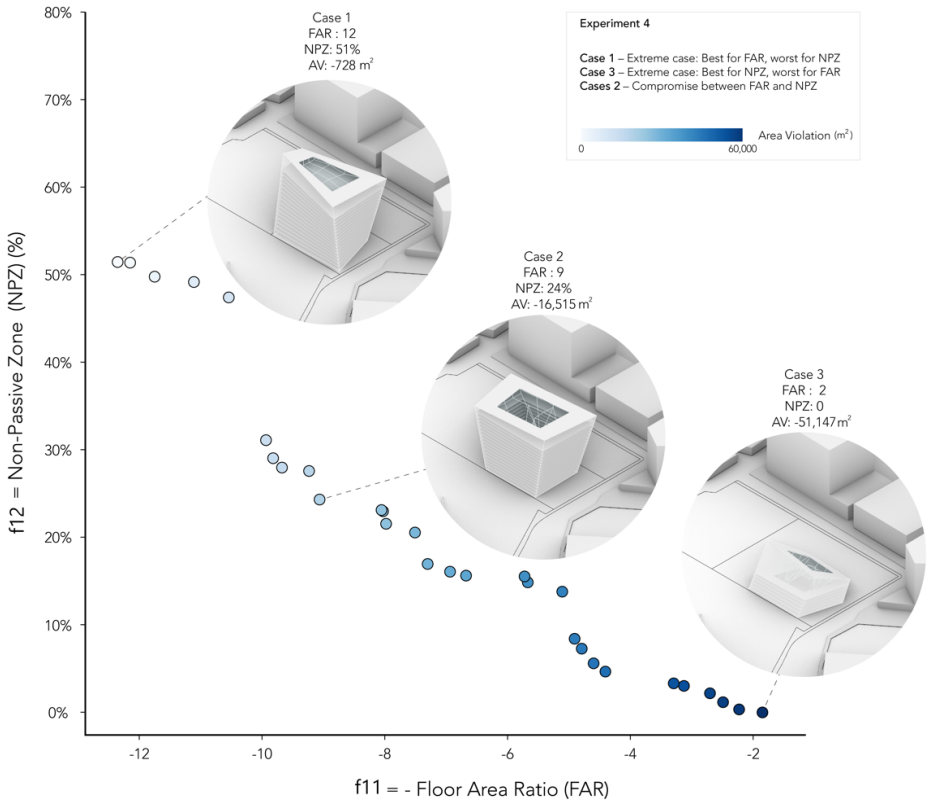


Figure 4.13. Experiment 4 non-dominated solutions. Color represents the third objective function, with dark blue meaning a higher AV and white representing cases where the total building area is closer to the initial goal.

Based on these results, Case 1 has the highest FAR and the lowest AV compared to other non-dominated solutions. This Case has large floor plates with a small courtyard with the highest percentage of NPZs. Case 2 is a compromise among all three objectives. Case 3 has the lowest FAR and NPZ values while showing a high AV from the goal area. In this experiment, increasing FAR leads to a decreased AV. All three visualized Cases in this experiment are a combination of carving and expansion operations (hybrids).



Figure 4.14. Experiment 4 non-dominated solutions sections.

Experiment 5. FAR, RBOS, AV (Site 1)

Experiment 5 explores FAR, RBOS, and AV as objective functions. Case 1 has the highest FAR and the lowest RBOS, while case 2 has the best performance in terms of AV. Case 2 is a good compromise among all objectives. Among the non-dominated solutions, case 3 has the highest RBOS and the lowest FAR value due to the carving's penetration within the massing. Case 3 also has the highest AV due to its small GFA. Case 3's facades have a larger perimeter on the roof compared to the ground floor, which helped increase the roof area. Cases 1, 2, and 3 are east-west oriented and therefore have larger northern and southern facades, which help maximize the RBOS in this experiment. As shown in Figure 4.15 and the color coding of AVs, solutions with higher FAR values have lower AV values.

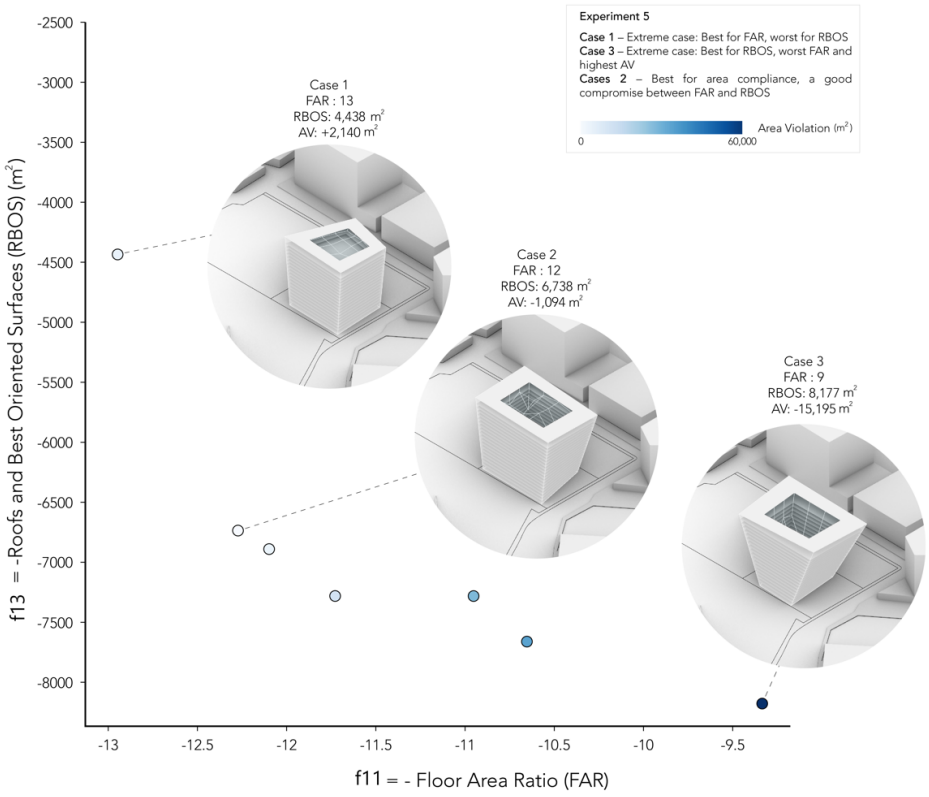


Figure 4.15. Experiment 5 non-dominated solutions. Color represents the third objective function.



Figure 4.16. Experiment 5 non-dominated solutions sections.

Experiment 6. FAR, UOS, AV (Site 1)

Experiment 6 explores FAR, UOS, and AV as objective functions. Among the non-dominated solutions, Case 1 is an extreme case with the highest FAR and the lowest AV value. Case 1 has the smallest UOS, which is expected due to its low FAR. Case 2 has a larger UOS compared to Case 3 but a lower FAR value than Case 1, while case 3 has the highest UOS and lowest FAR among the non-dominated solutions. Case 1 in this experiment is the result of expansion operation,

while case 2 is the result of carving. Case 3 is a hybrid resulting from carving and expansion operations. Solutions with the highest FAR values have better area compliance (lower AV).

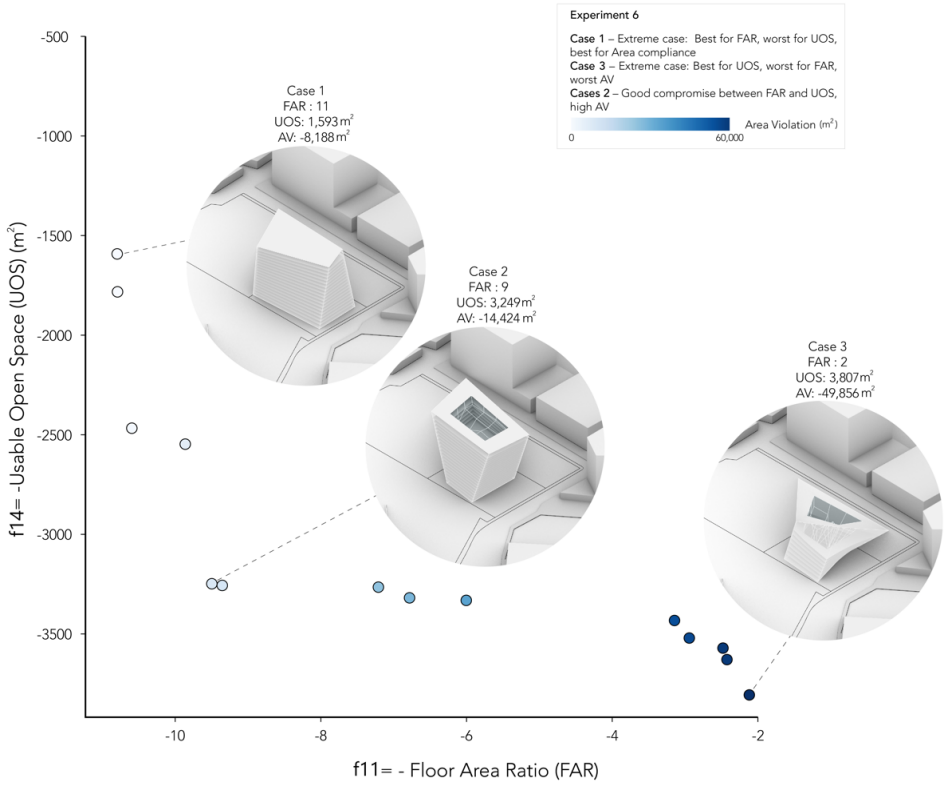


Figure 4.17. Experiment 6 non-dominated solutions. Color represents the third objective function.



Figure 4.18. Experiment 6 non-dominated solutions sections.

Experiments 7-9 introduce a new variable, site area, to understand the impact of the site area on the solution space. Experiments 7-9 are thus performed using site 2, with an area of 11,500 m^2 (115 m x 100 m), a 139.5% increase in relation to site 1, which had 60 m x 80 m as dimensions.

Experiment 7. FAR, NPZ, AV (Site 2)

In experiment 7, Case 1 is an extreme solution with the highest FAR and a high NPZ value due to its large slabs and small courtyard. Case 2 has a lower NPZ value due to its large courtyards and reduced slab sizes. Case 3 has a high AV and a low FAR while having a low NPZ value. This high AV is due to the small footprint of each slab and the large courtyard within the building. In experiment 7, another solution is picked for visualization, Case 4, with the highest area compliance in the non-dominated solutions. In addition, Case 4 has a large courtyard that helps reduce the NPZ value. Case 2 and Case 4 are similar; both Cases provide a good compromise between FAR, NPZ, and AV. Both Cases 2 and 4 are hybrids of carving and expansion.

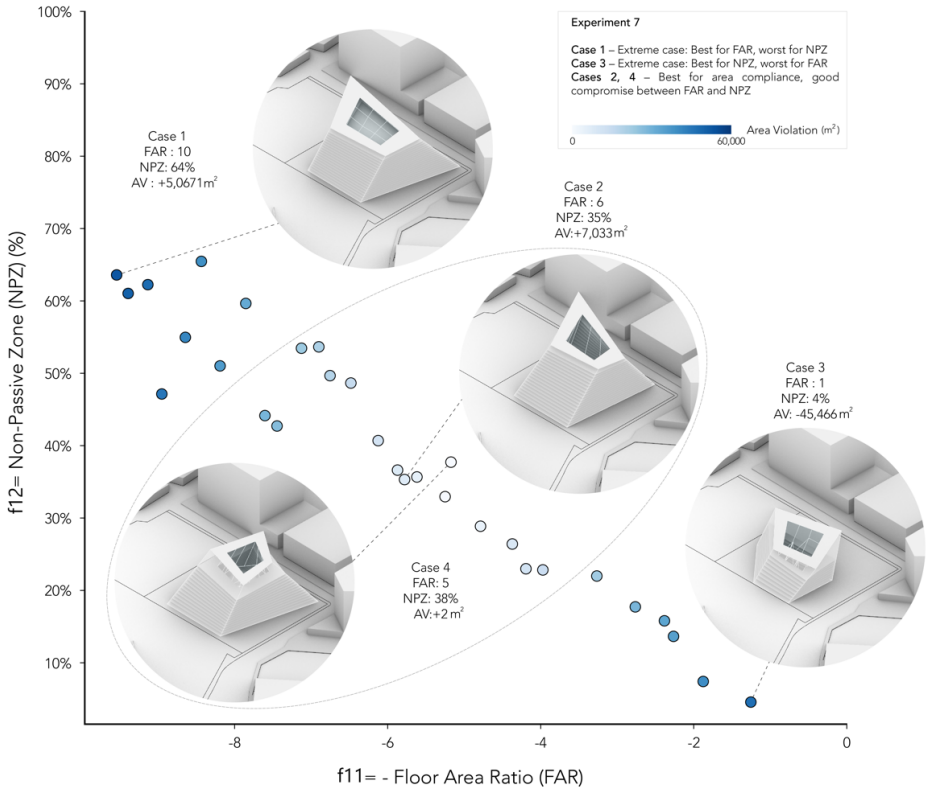


Figure 4.19. Experiment 7 non-dominated solutions. Color represents the third objective function.

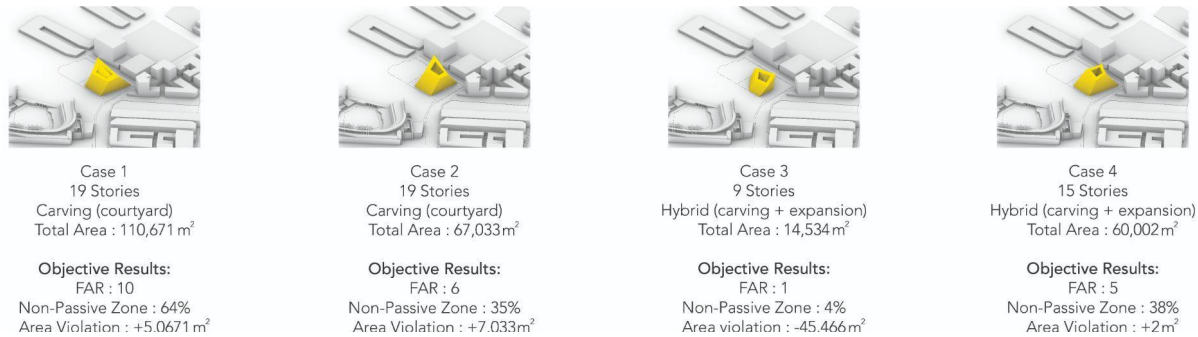


Figure 4.20. Experiment 7 non-dominated solutions sections.

Experiment 8. FAR, RBOS, AV (Site 2)

Experiment 8 explores FAR, RBOS, and AV as objective functions to find optimized massings for site 2. The goal is to maximize FAR and RBOS and minimize the AV. Case 1 has the highest FAR and a low RBOS value with a high AV value. Case 2 compromises FAR and RBOS while having a high AV. Case 3 has the lowest FAR value and a high AV while having the highest RBOS value. Case 4 is selected since it has the lowest AV value among the non-dominated solutions. Case 3 and 4 are results of expansion and carving operations, while Case 1 does not include carving and Case 2 includes carving only.

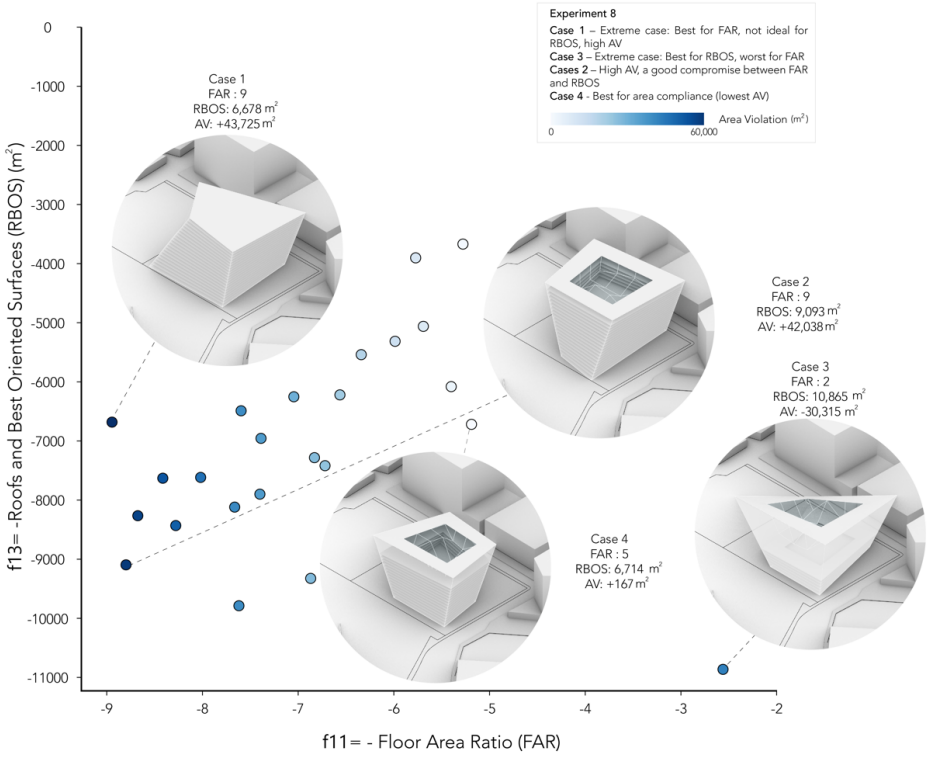


Figure 4.21. Experiment 8 non-dominated solutions. Color represents the third objective function.

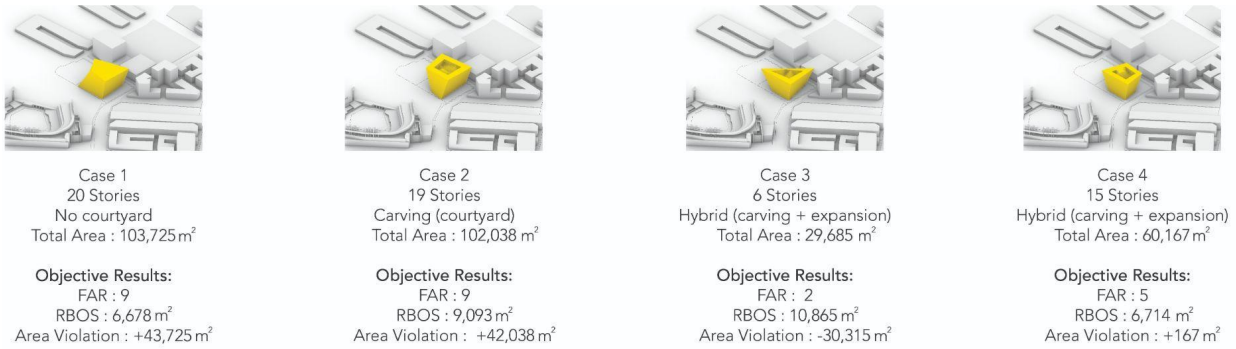


Figure 4.22. Experiment 8 non-dominated solutions sections.

Experiment 9. FAR, UOS, AV (Site 2)

In experiment 9, we explore FAR, UOS, and AV as objective functions to optimize massing options on site 2. The goal is to maximize FAR and UOS and minimize AV. Based on these results, Case 1 is an extreme case that provides the lowest UOS value and a high AV while providing the highest FAR value. Case 1 uses the opportunity of the larger site compared to experiments 1-6 by maximizing the FAR beyond what was needed by the area goals. On the other hand, Case 2 has a lower AV and a higher UOS value. Case 3 has the lowest FAR and the highest UOS value, with the highest AV. Case 3 is the result of expansion and carving operations combined through the evolution of the individuals. Cases 1 and 2 are the results of carving only. Among the solutions, Case 2 provides a good compromise among the objectives of experiment 9. Unlike the similar experiment on site 1 (experiment 6), the solutions with FAR between 4-6 have a higher area compliance (lower AV value).

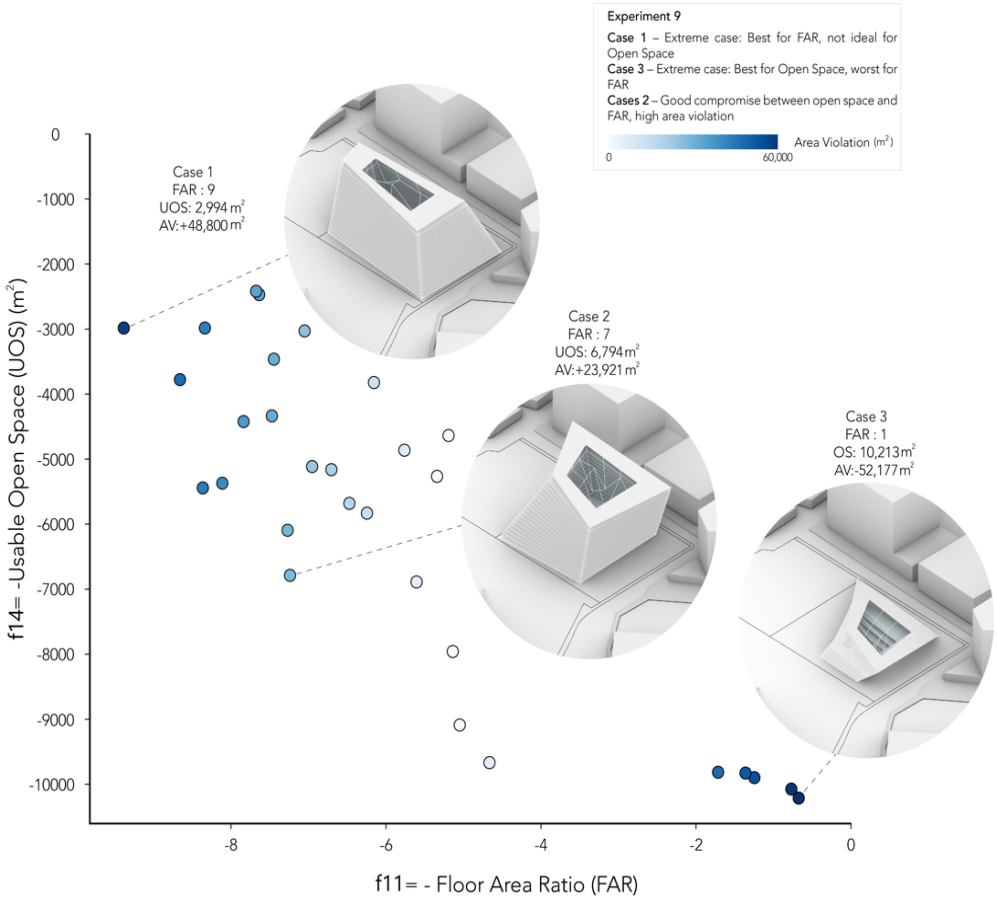


Figure 4.23. Experiment 9 non-dominated solutions. Color represents the third objective function.



Figure 4.24. Experiment 9 non-dominated solutions sections.

4.1.4 Discussion

The results of experiments will be discussed by dividing them into three categories based on the studied objective functions in each experiment. The three categories are 1. NPZ objective results, 2. RBOS objective results, and 3. UOS objective results. Case 1 in most experiments provides small or no carvings, which leads to the least passive options based on the discussed rules-of-thumb. Overall, in all experiments, Case 3, with the lowest FAR value, results from carving and expansion operations. Due to the addition of expansion, the highest floor plate in these cases (Case 3 across experiments 1-9) can result in a large volume, creating inefficiency in the space. Case 2 solutions in most experiments create efficient building volumes closest to how architects would design an efficient courtyard building. Further investigations using shape similarity techniques will be used to compare the generated design options among different experiments.

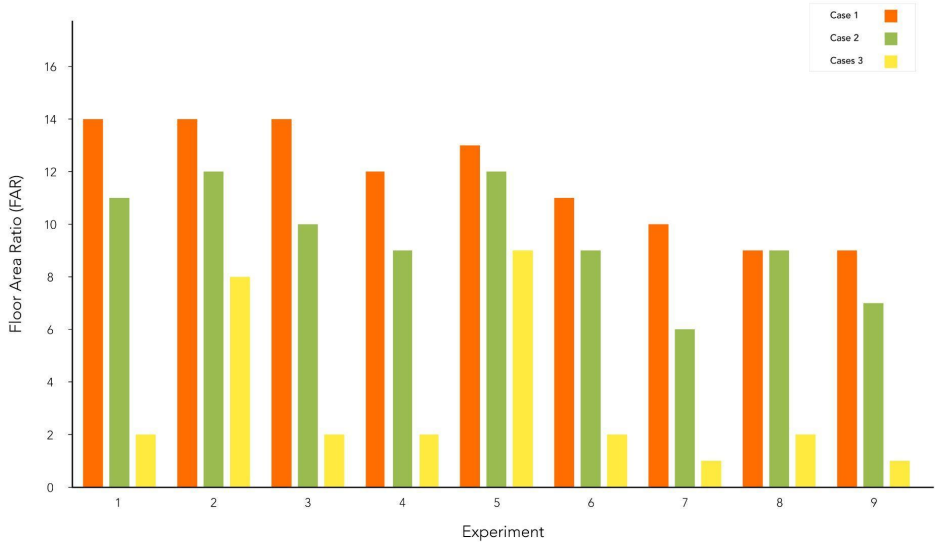


Figure 4.25. Comparison of FAR among Cases 1-3 in each experiment in experiments 1-9.

In addition, in experiments that use site 1 with AV as an objective function (experiments 4-6), the non-dominated solutions with the maximum FAR have lower AV values, while in site 2, solutions with FARs between 4-6 result in the least AV values (best area compliance). Figure 4.25 compares FAR objective functions among Cases 1-3 in experiments 1-9. Overall, in all experiments, Case 1 has a higher or equal value of FAR, and Case 3 in all experiments results in a lower FAR value in every experiment. Figure 4.26 shows experiment 4-9 Cases 1-3, where AV is used as objective functions. In experiments 4-6, where site 1 is used, Case 1 has the minimum AV in all three experiments. In experiments 7-9, where site 2 is used with a larger area, Case 2 and 3 have smaller AV values (except for experiment 9, where Case 3 results in a significantly smaller building and a higher AV value compared to Case 1).

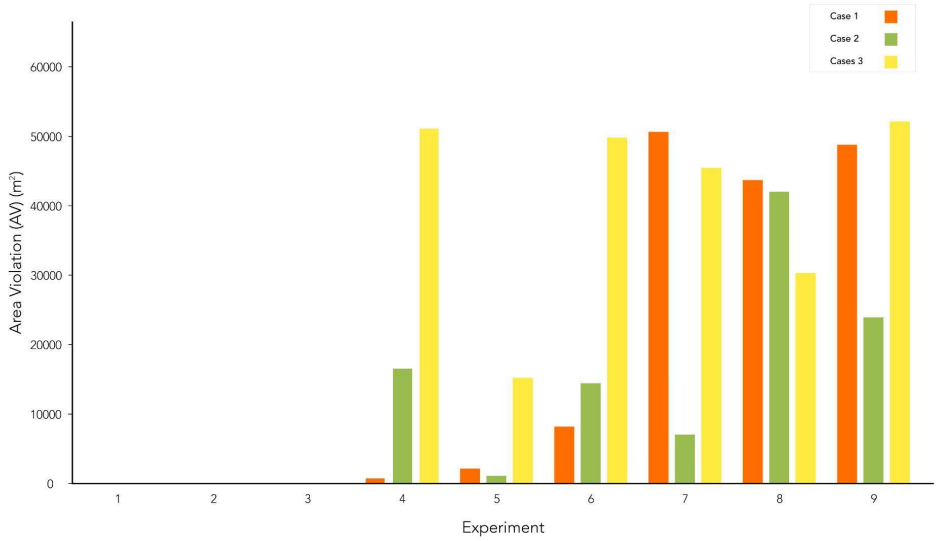


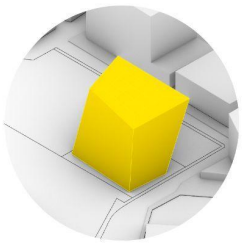
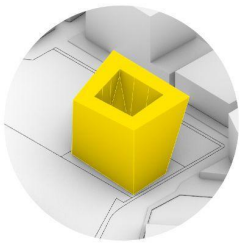
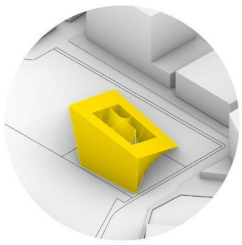
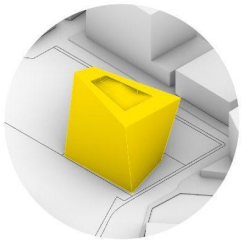
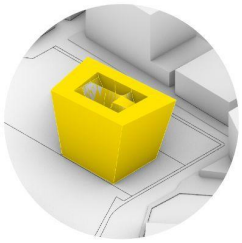
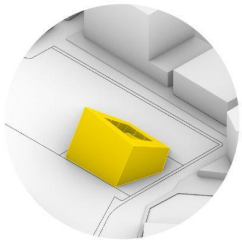
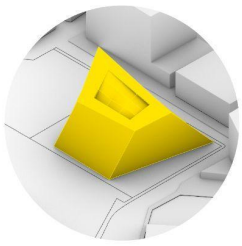
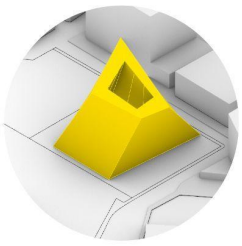
Figure 4.26. Comparison of AV among Cases 1-3 in experiments 4-9.

NPZ objective results

This section compares experiments 1, 4, and 7, where NPZ is one of the applied objectives. Among the non-dominated solutions that explore FAR and NPZ objectives, experiment 1 and experiment 4 result in similar massing solutions, and Cases 1, 2, and 3 in both experiments have similar NPZ values. Experiment 7, on the other hand, provides options with the lowest AV Values (lowest AV in Case 4), which is due to the larger size of site 2 that allows for a larger first-floor area explored in the solutions and therefore decreases the amount of AV.

In experiment 4, where site 1 is used with the smaller footprint, non-dominated solutions with higher FAR values (more than 10) provide lower AVs. In contrast, in experiment 7, solutions with FAR between 4-6 are more desirable due to their lower AV. All three experiments minimize the NPZ and experiment 1 and 4 provide an NPZ value of zero. Figure 4.27 shows the comparison of Cases in experiments 1, 4, and 7 that all relate to the NPZ objective function. These results show that the cases follow a similar pattern where Case 1 has the highest NPZ value, and Case 3 has the minimum NPZ value. Table 4.2 compares Cases 1,2, and 3 from experiments 1, 4, and 7.

Table 4.2. NPZ objective results.

Experiment 1	Case 1 FAR : 14 NPZ: 49% 	Case 2 FAR : 11 NPZ 23% 	Case 3 FAR : 2 NPZ: 0 
Experiment 4	Case 1 FAR : 12 NPZ: 51% AV: -728 m^2 	Case 2 FAR : 9 NPZ: 24% AV: -16,515 m^2 	Case 3 FAR : 2 NPZ: 0 AV: -51,147 m^2 
Experiment 7	Case 1 FAR : 10 NPZ: 64% AV: +50,671 m^2 	Case 2 FAR : 6 NPZ: 35% AV:+7,033 m^2 	Case 3 FAR : 1 NPZ: 4% AV: -45,466 m^2 

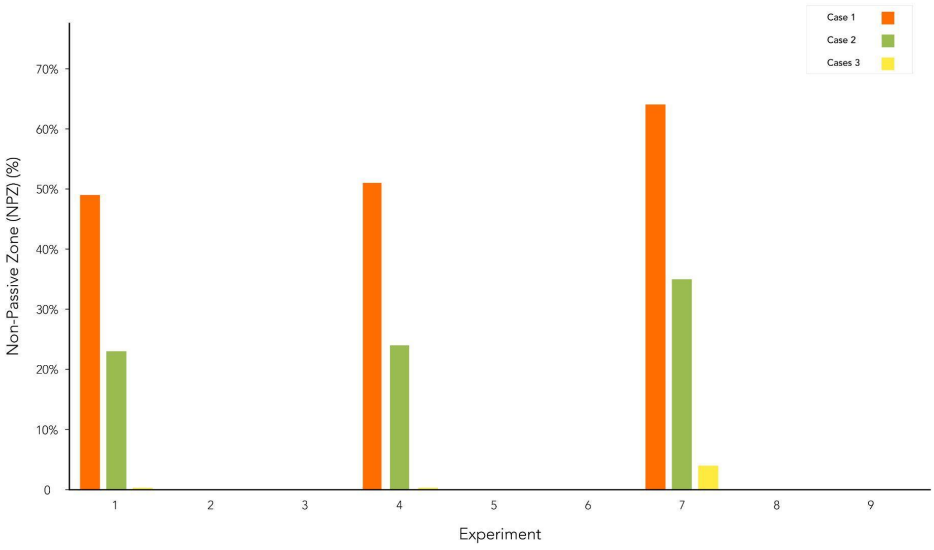
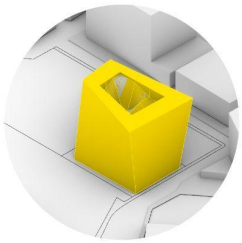
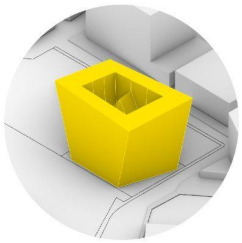
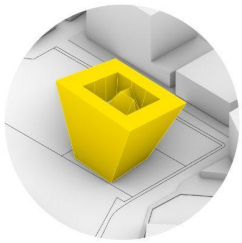
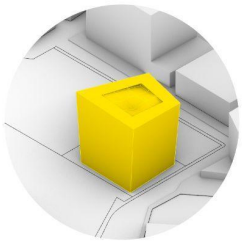
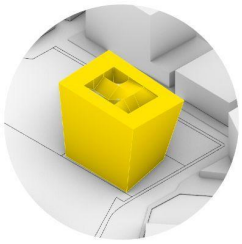
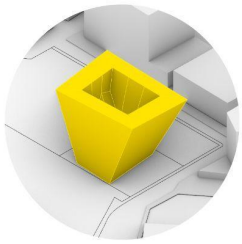
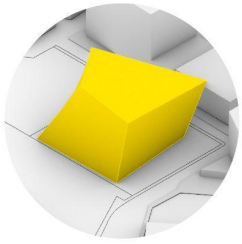
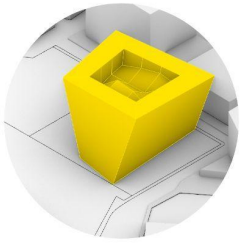


Figure 4.27. Experiments 1, 4, and 7 comparisons of NPZ values among Cases 1-3.

RBOS objective results

This section discusses experiments 2, 5, and 8, where RBOS is one of the applied multi-objectives. Similar to NPZ experiments, experiments 2 and 5 that use site 1 result in similar shapes, where Cases 1, 2, and 3 in both experiments have similar values and overall shapes. Experiment 8 provides the lowest value of AV and RBOS compared to experiments 2 and 5 due to site 2 size. The Case with the highest RBOS value in experiment 8 (Case 3) has a maximized roof area in addition to the ability to provide maximized facades and therefore has the highest RBOS among all solutions. In all experiments 2, 5, and 8, buildings are east-west facing, which helps with the RBOS values with higher northern and southern facade areas. Case 3 in all three experiments results in a tapered form with a smaller ground floor perimeter. In experiment 5, AV decreases as the FAR increases, while in experiment 8, the AV is at the minimum within FAR 5-6. Figure 4.28 compares RBOS objective functions in experiments 2, 5, and 8. All three experiments' cases follow a similar pattern in their RBOS value, where Case 1 has the lowest RBOS value, and Case 3 has the highest value. Table 4.3 compares selected Cases from experiments 2, 5, and 8.

Table 4.3. RBOS objective results.

Experiment 2	Case 1 FAR : 14 RBOS: 5,029 m² 	Case 2 FAR : 12 RBOS: 7,180 m² 	Case 3 FAR : 8 RBOS: 8,680 m² 
Experiment 5	Case 1 FAR : 13 RBOS: 4,438 m² AV: +2,140 m² 	Case 2 FAR : 12 RBOS: 6,738 m² AV: -1,094 m² 	Case 3 FAR : 9 RBOS: 8,177 m² AV: -15,195 m² 
Experiment 8	Case 1 FAR : 9 RBOS: 6,678 m² AV: +43,725 m² 	Case 2 FAR : 9 RBOS: 9,093 m² AV: +42,038 m² 	Case 3 FAR : 2 RBOS: 10,865 m² AV: -30,315 m² 

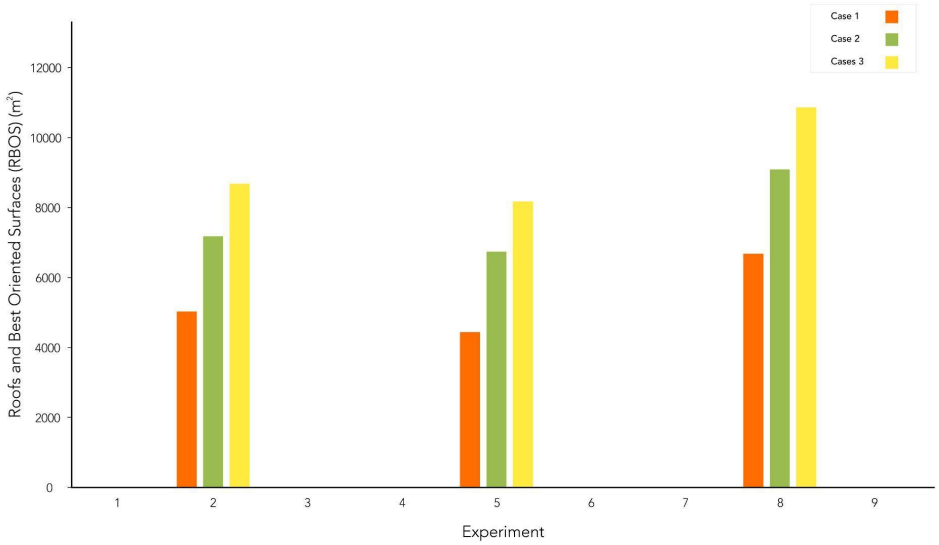
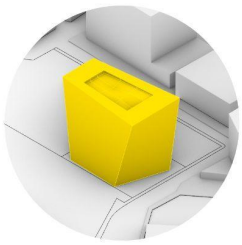
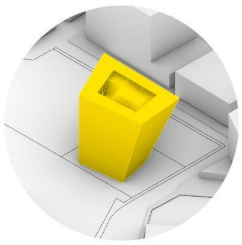
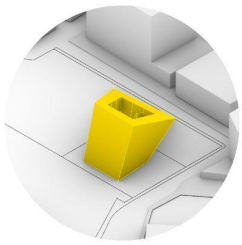
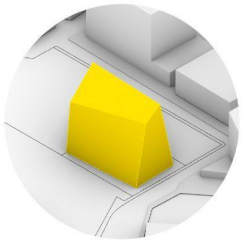
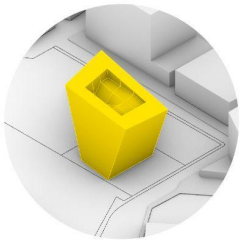
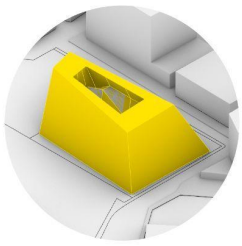
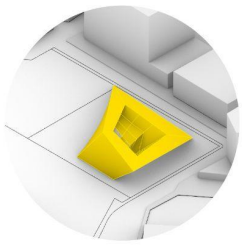


Figure 4.28. Experiments 2, 5, and 8 comparisons of RBOS among Cases 1-3.

UOS objective results

In this section, experiments 3, 6, and 9 are compared, where UOS is one of the applied multi-objectives. Among experiments focused on the UOS, experiments 3 and 6 resulted in cases with similar FAR and UOS values. In both experiments, Case 3 is a hybrid massing (a combination of expansion and carving). In all three experiments, solutions with less carving were found compared to the other experiments in this project due to the focus of these experiments on providing maximum FAR and UOS, and carving can mainly help the AV Objective. Experiment 9 achieves the maximum UOS value. Similar to other experiments, in experiment 6 on site 1, non-dominated solutions with the lowest AVs are located where the FAR is maximized. In experiment 9, designed on site 2, the solutions with FAR between 4-6 have lower AV values. Figure 4.29 compares UOS objective functions in experiments 3, 6, and 9. In experiments 3, 6, and 9, all Case 3 solutions have the highest value of UOS, and the UOS is at its highest in experiment 9, where site 2 is used. The ability to reach higher UOS values is expected on site 2 as it provides a larger area for generating massings. Table 4.4 compares selected Cases from experiments 3, 6, and 9.

Table 4.4. UOS objective results.

Experiment 3	Case 1 FAR : 14 UOS: 1,854 m^2 	Case 2 FAR : 10 UOS: 2,909 m^2 	Case 3 FAR : 2 UOS: 3,899 m^2 
Experiment 6	Case 1 FAR : 11 UOS: 1,593 m^2 AV: -8,188 m^2 	Case 2 FAR : 9 UOS: 3,249 m^2 AV: -14,424 m^2 	Case 3 FAR : 2 UOS: 3,807 m^2 AV:-49,856 m^2 
Experiment 9	Case 1 FAR : 9 UOS: 2,994 m^2 AV:+48,800 m^2 	Case 2 FAR : 7 UOS: 6,794 m^2 AV:+23,921 m^2 	Case 3 FAR : 1 UOS: 10,213 m^2 AV:-52,177 m^2 

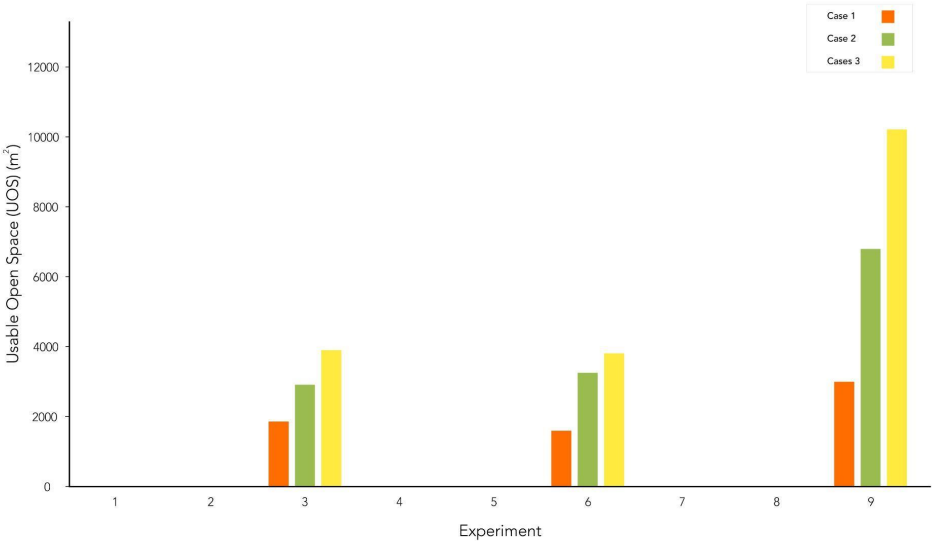


Figure 4.29. Experiments 3, 6, and 9 comparisons of UOS among Cases 1-3.

In the future, the proposed generative design workflow can be implemented for massing optimization within immersive environments (e.g., augmented reality (AR)). Using the proposed workflow in this research, users can run optimizations independently from other software (e.g., Rhinoceros) using their head-mounted devices (HMDs) and handheld devices and using cloud computing. Generating massing solutions can be explored on a third site smaller than site 1, which can result in different applications of the design operations (i.e., carving and expansion) and different massing proportions compared to the current experiments' results.

User integration in the generative design process can be explored, where user feedback can be used as an objective in the multi-objective optimization, or used to find feasible design solutions to train machine learning models [223]. As an example, SplineLearner, a generative learning framework for B-spline surfaces used user feedback to label feasible or infeasible surfaces generated [223]. User studies will need to be conducted to understand the impacts of integration of users with the proposed generative design workflow discussed in this research. In the future, other algorithms such as multi-objective particle swarm (MOPSO), and multi-objective simulated annealing (MOSA) can be explored to compare our case study results.

4.1.5 Experiment Conclusions

This experiment explores a new generative design method inspired by architectural design operations, a fundamental process in massing design. A new application of NSGA-II algorithm is proposed that uses Operative Design to help optimize geometries without the need for other 3D platforms. Such algorithms can be helpful for geometry optimizations in immersive technologies with the ability to run cloud-based computations without the need for 3D modeling platforms.

Nine experiments were designed to test the algorithm's capability to optimize geometries based on objective functions reflecting common architectural design goals, including FAR, RBOS, NPZ, and UOS. In experiments 4-9, a fifth objective was introduced, the AV, in relation to target square

footage. Due to its importance in architectural design practice, FAR was used as a consistent objective function in all experiments. There are shape similarities among cases with and without the AV as an objective, where two objective functions' values are similar, as expected. Among the non-dominated solutions visualized in all experiments, there are hybrids (combinations of carving and expansion) and solutions with only either carving or expansion as the operation. All experiments were able to achieve their minimization/maximization goals.

Case 1, in most experiments, provides the smallest carvings, leading to the least passive options based on the proposed rules-of-thumb. Case 2, on the other hand, in most experiments, creates efficient building volumes closer to how architects would design an efficient courtyard building. Overall, Case 3 (with the lowest FAR value) is always a hybrid solution in all experiments. In addition, for experiments that use site 1 with AV as an objective, the non-dominated solutions with maximum FAR have lower AV values. In contrast, in site 2, solutions with FARs between 4-6 result in better area compliance.

Further investigations using shape similarity techniques can be used to analyze the generated solutions and extract additional conclusions. In the future, other design operations can be integrated within the algorithm to create more complex design solutions based on different programs or aesthetic needs of designers and stakeholders. Another promising research path is the integration of Operative Generative Design algorithms in immersive technologies like augmented reality, allowing user participation in design optimization during early-stage architecture design by introducing user preference as one of the objective functions.

4.2 Experiment 2: InsightXR: Operative Generative Design in Augmented Reality

This section introduces InsightXR, the augmented reality (AR) application for early-stage massing studies in architecture for this dissertation. InsightXR uses a novel operative generative design algorithm using Non-Dominated Sorting Genetic Algorithm II (NSGA-II) in the Pymoo framework [19]. To test InsightXR's capabilities, a case study is introduced to generate 3D design alternatives in real-time for a mixed-use project located in San Francisco, USA. Design objectives explored in this research are Roof and Best Oriented Surfaces (RBOS), Usable Open Space (UOS), and Area Violation (AV). Design alternatives are generated for 5 generations and visualized in InsightXR. In Experiment 3, an interactive optimization process will be introduced to generate subjectively desirable design options based on users' interactive feedback.

4.2.1 Introduction

In the proposed workflow, the operative NSGA-II algorithm is accessed using Amazon Web Services, and the users can visualize selected design alternatives from each generation using InsightXR [293]. To reduce user fatigue during this interactive process, InsightXR only shows selected design options per generation (3 design options) sampled based on the shape similarity of meshes over the five generations [217], [16].

In this research, the proposed generative design process in augmented reality is introduced using a case study. This case study endeavors to generate design alternatives using the proposed operative generative design algorithm for a site located at 74 Mission Rock St, San Francisco, CA 94158 [13]. To assess the capabilities of the proposed algorithm in augmented reality, design alternatives are generated using the following objective functions: Roof and Best Oriented Surfaces (RBOS), Usable Open Space (UOS), and Area Violation (AV). RBOS explores maximizing the total area of the best-oriented facades, which can result in an increase in the potential for renewable energy in addition to an increase in daylighting [13]. UOS can be defined as the total unbuilt area of the site, which can be calculated by subtracting the ground floor area from the lot area. The goal of this research is to maximize the UOS value. AV is a constraint violation used as an objective function in this research [13].

4.2.2 Methods

Implementing generative design in a participatory environment can facilitate the introduction of novel design ideas and foster creative work that couldn't have been conceived of by the designers or design reviewers in a standard design process. Integrating immersive technology such as augmented and virtual reality into a user's interaction in the design process can help their immersion and better promote understanding of the project's scale and overall geometry. This is in addition to the benefits of being able to interact with the design remotely.

InsightXR's proposed Operative Generative Design workflow uses AR technology that enables users to visualize generated design alternatives using the applied NSGA-II algorithm. In this workflow, Pymoo's framework in Python is used to run the applied NSGA-II algorithm in AR [19]. The algorithm runs on a local server, and using Amazon Web Services (AWS), generates design alternatives that can be visualized in AR in real-time. In the proposed workflow, the designers will have access to the generated database of different design alternatives produced using InsightXR since they are saved into AWS storage, and they can decide on the design outcome with full access to that knowledge. InsightXR is created using Unity Game Engine, and the AR capabilities of InsightXR use the AR Foundation framework within Unity. In addition, since InsightXR is designed for iOS devices, the ARKit XR Plugin is used within Unity to allow for the use of iOS devices' features and capabilities [291].

In this section, a case study is introduced to test the capabilities of InsightXR with respect to applying Operative Generative Design in augmented reality during early-stage massing studies. In addition, the proposed case study is used to assess the proposed clustering methods for visualizing selected design options [13]. This case study explores using the operative generative design using the NSGA-II algorithm by generating design solutions for early-stage massing studies for a mixed-use building on a site located at 74 Mission Rock St, San Francisco, CA 94158. The site covers an area of 4800 m^2 and is 80 m by 60 m [13]. In this case study, the objectives related to rule-of-thumb sustainable design and early-stage design are used. The explored objectives, as discussed in the Introduction, include: 1. Roofs and Best Oriented Surfaces (RBOS), 2. Usable Open Space (UOS). In addition, Area Violation (AV) is used as a constraint violation in this case study. The goal of this research is to maximize the RBOS and UOS while minimizing the AV.

The population size in the NSGA-II algorithm is f 30 individuals, with the simulated binary crossover set to 0.9. After 5 generations, the optimization is terminated. Design operations explored in this case study are Carving and Expansion (applied by adjusting the roof mesh points). In each generation, a selection of design alternatives (solutions) is presented to the users. At the end of the process, the Pareto Front can be visualized for further investigation of the solution space. In order to reduce user fatigue, three design alternatives are selected per generation using k-means clustering to be shown to the users [176].

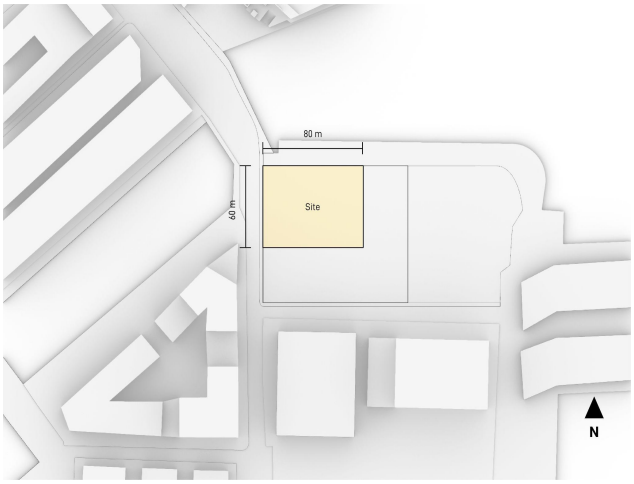


Figure 4.30. The case study site used in this research [13].

The overall generative design workflow in InsightXR can be described as follows:

1. The applied NSGA-II algorithm runs on a local server for the initial generation and creates the first generation of individuals.

2. A k-means clustering algorithm is applied to the initial population. The k-means clustering algorithm used in this research is based on the Scikit-learn Python module [176]. In this process, individuals are separated into groups with equal variance while minimizing the inertia within the cluster [176]. In each generation, 3 clusters are created. The k-means clustering in this process is defined based on the fitness values (objective function results for all three objectives).

3. In the next step, each cluster's design alternatives are analyzed using a shape similarity algorithm. This algorithm is designed based on the sum of Euclidean distances between each vertex of meshes, as discussed by Beham et al. [16]. The calculated shape similarity values are then used to rank each cluster's design alternatives from most similar to least.

4. Next, from each cluster, the first design alternative based on the shape similarity ranking is selected to be viewed in AR. The selected design alternatives from each cluster are the most representative of each cluster (least dissimilar based on shape similarity algorithms).

5. The algorithm then creates OBJ files representing each selected design alternative for the three clusters and then uploads them to the defined storage on AWS.

6. On InsightXR's application, the users are provided an interface to visualize the first generation's design alternatives.

7. In the next step, the next generation of individuals is created using the applied NSGA-II algorithm, and the discussed clustering method selects three design alternatives from the new generation to be visualized by the users in each generation.

8. These steps will repeat for each generation until the termination criterion is satisfied (defined as a generation limit of 5).

9. In the end, the resulting Pareto-front is generated using the algorithms, and users can visualize this result on InsightXR's application. Users can visualize design alternatives in real-time or view the results after all generations are finished running.



Figure 4.31. The application overview; a. Using QR code to visualize the site as the first step; b. Starting the generative design process; c. the site and a visualized 3D design alternative from the generation; d. Overview of all designs in generation 1.

During the experience, the 3D site model is first visualized using the Image Tracking method in the real world. To achieve that, a predefined QR code is placed in a fixed position in the physical world, and the site will get visualized and placed at the center of the QR code once it's detected by the AR camera [295]. Each design alternative's 3D model is created by generating an obj file format from the NSGA-II algorithm and uploading the design alternative on an S3 bucket, a dedicated storage space for the algorithm on AWS. Selected design options using the clustering method described earlier are downloaded from AWS throughout the experience and are visualized on the AR app. Figure 4.31 shows the interface used in InsightXR from starting the NSGA-II optimization process through visualizing the design options.

4.2.3 Results

In this section, the design alternatives generated using InsightXR will be discussed. First, the final non-dominated solutions will be presented. Next, the generated design solutions in each generation will be introduced while assessing the selected solutions visualized in InsightXR. To discuss the characteristics of the non-dominated solutions, 4 design options are visualized. Figure 4.23 visualizes the non-dominated solutions using AV and UOS on the horizontal and vertical axes and RBOS as the third objective using color where gray means higher RBOS values and light gray means lower RBOS values. As can be seen in the non-dominated solution results visualized in the figure, as the AV increases, UOS increases due to the smaller overall gross floor area in these results.

In Case 4, the AV value is at its lowest among all solutions, with one of the highest values for RBOS. This design option is a courtyard design, with only carving as the applied operation on the massing. Case 4 reaches the maximum height allowed on the site (61 m) while creating a deep courtyard that will provide daylighting. This case aims to reach the maximum areas for the southern and northern facades while reducing the first floor (ground level) floor plate. This case has one of the lowest UOS values while providing the lowest AV value.

Case 1 is a hybrid design option (carving and expansion), with the roof facing south and maximizing the roof area. This design option has an increased unbuilt area on the lot, allowing an increased level of UOS to 64% of the site. Case 1 provides the maximum RBOS value. Case 2 has an increased height with a reduced value for RBOS compared to Cases 4 and 1 while increasing the UOS value. Finally, Case 3 has the highest UOS value while increasing the AV by reducing the gross floor area of the building. This case is a hybrid design option (carving and expansion). Figure 4.33 visualizes the same non-dominated solutions but uses RBOS and UOS on the horizontal and vertical axes and AV as the third objective using color where dark blue means higher AV values and light blue means lower AV values. This visualization only shows 3 selected Cases, with Case 1 having the highest RBOS value, Case 2 having a good trade-off between RBOS and UOS objectives, while Case 3 having the lowest RBOS and highest UOS values.

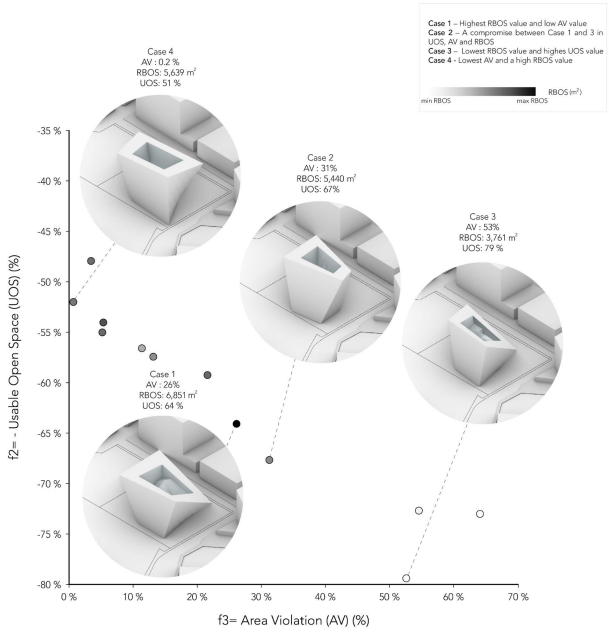


Figure 4.32. Non-dominated solutions. The third objective function, RBOS, is represented using color. Dark Grey represents higher RBOS values, and light gray means lower RBOS values.

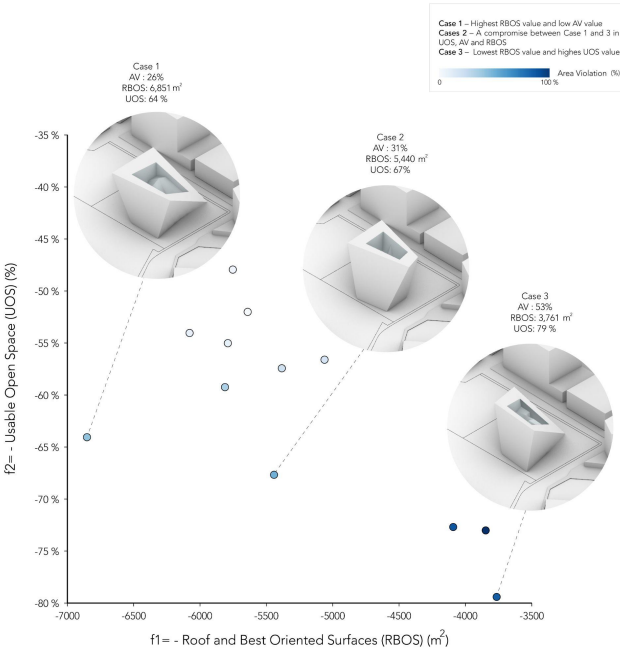


Figure 4.33. Non-dominated solutions. The third objective function, AV, is represented using color. Dark blue represents higher AV values, and light blue means lower AV values.

Next, each generation's solutions and the selected design options for visualization in InsightXR will be discussed in further detail.

Generation 1

At this step, the initial population of solutions is generated using the applied NSGA-II algorithm. The solutions are then divided into three clusters using the k-means clustering algorithm discussed in section 3. Using a shape similarity algorithm, the shapes in each cluster are then ranked, and the most representative shape of each cluster is automatically selected and turned into an OBJ file and uploaded to AWS to be viewed by the user in real-time. From a user's experience point of view, the user will select to start the optimization process and start with generation one. They will be presented with three design options, as discussed in previous sections, and they can view each of these design solutions. Figure 4.34 shows the three design options visualized in InsightXR.

As can be seen in Figure 4.35, cluster 1's representative solution, called Case 1, is a result of expansion operation, with a low AV value and a high RBOS value. Case 2, on the other hand, is a hybrid solution, resulting from expansion and carving in the form of a courtyard. This case maximizes the southern facade and roof area while reducing the unbuilt area on the ground level (increasing the UOS). Case 3 is an expansion design option with a lower value of RBOS and a lower value of UOS. Overall, cluster 1 includes options with higher RBOS values and lower AV values, cluster 2 is a trade-off between cluster 1 and 3 in terms of AV and RBOS values, and cluster 3 includes options with higher AV and lower RBOS values. Cases 1 and 3 have shallow skylights, and in this research, shallow skylights (under one story height) are not considered a courtyard.



Figure 4.34. Generation 1 solutions. A. Case 1, b. Case 2, c. Case 3.

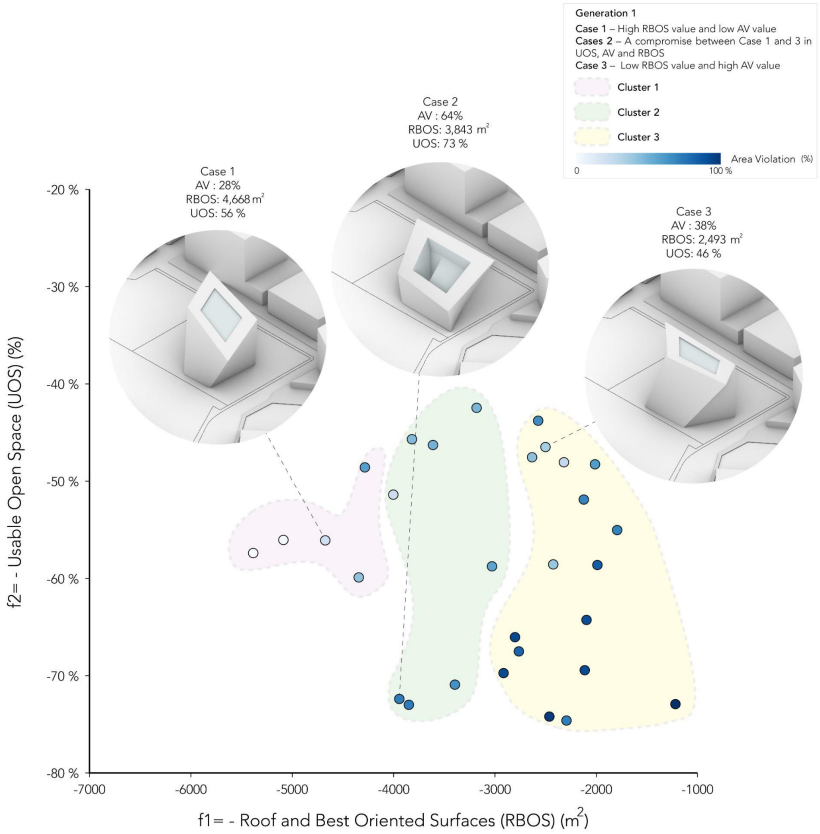


Figure 4.35. Generation 1 solutions. The third objective function, AV, is represented using color. Dark blue represents higher AV values, and light blue means lower AV values.

Generation 2

As can be seen in Figure 4.37, generation two's clusters are defined similarly to the previous generation, where cluster 1 includes options with higher RBOS values and lower AV values, which is desired in this case study. Cluster 2 includes design alternatives that show a tradeoff between RBOS and UOS values, and cluster 3 options have the lowest RBOS values and highest UOS values. Cluster 3 design options have the highest AV values among all options, which is due to their deep courtyards (carving) resulting in smaller floor plates.

Similar to generation one, three design options are selected using the algorithm and visualized in InsightXR. Case 1 is a hybrid design option (expansion and carving) with a high RBOS value. Case 2 is a tradeoff between Case 1 and 3 in RBOS and AV values. Cluster 3 had the widest variety of UOS values. Case 3 has a lower UOS value and RBOS value, and high AV values.

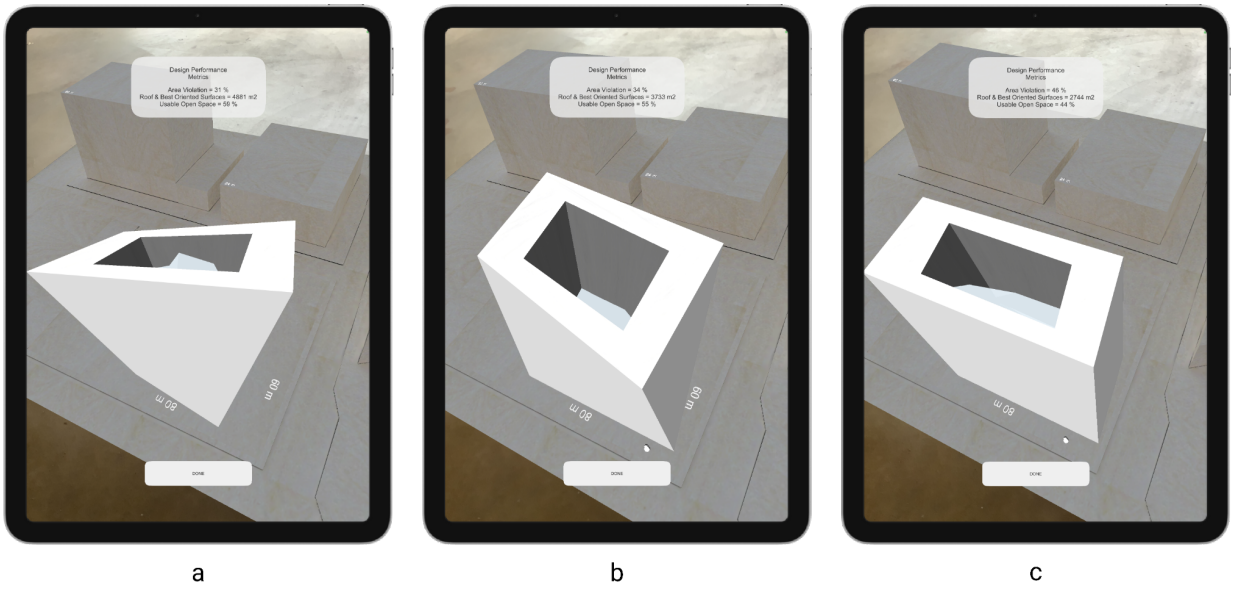


Figure 4.36. Generation 2 solutions. A. Case 1, b. Case 2, c. Case 3.

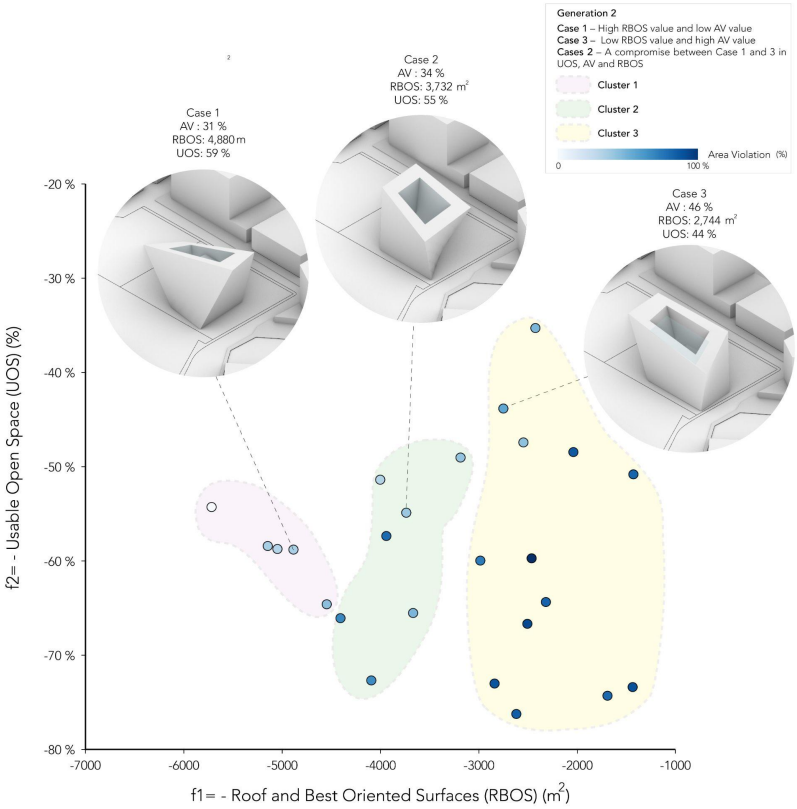


Figure 4.37. Generation 2 solutions. The third objective function, AV, is represented using color. Dark blue represents higher AV values, and light blue means lower AV values.

Generation 3

As can be seen in Figure 4.39, in generation three, cluster one includes options with the highest RBOS and lowest AV values among all solutions while having UOS values between 50 and 60 percent. On the other hand, cluster two includes solutions with tradeoffs between RBOS, UOS, and AV values compared to the other two clusters. Cluster 3 includes solutions with lower RBOS values and higher AV values, which are not desired, while providing higher values of UOS, which can result in smaller floor plates or larger courtyards.

Case 1 is a hybrid design option (carving and expansion) with a large UOS value, tilting towards the south and providing a large RBOS value. Case 1 provides a better AV value compared to Cases 2 and 3. Case 2 is a tradeoff between Case 1 and Case 3 in terms of RBOS and AV values. Case 3 provides the highest UOS among the three selected cases while providing the lowest RBOS. Case 3 is a hybrid design option (carving and expansion) with a high AV value that doesn't provide the desired gross floor area. Figure 4.38 shows the visualized Cases 1, 2 and 3 in InsightXR.



Figure 4.38. Generation 3 solutions. A. Case 1, b. Case 2, c. Case 3.

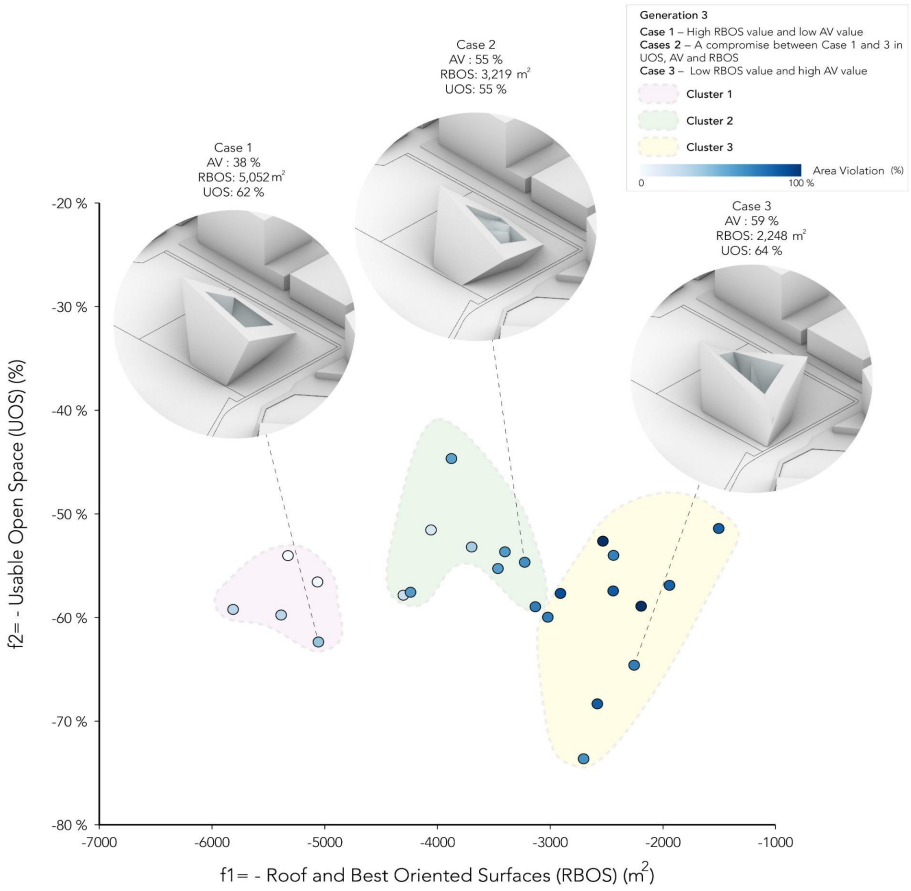


Figure 4.39. Generation 3 solutions. The third objective function, AV, is represented using color. Dark blue represents higher AV values, and light blue means lower AV values.

Generation 4

Generation 4 design options are more diversified compared to the previous generations. Cluster 1 includes solutions with the highest RBOS values and the lowest AV values, and as can be seen in Case 1, this option is a carving solution (courtyard) that includes large floor plates with a shallow courtyard. Cluster 2, on the other hand, provides a compromise between clusters 1 and 3, where its solutions provide lower RBOS values compared to cluster 1 but higher RBOS values compared to cluster 3. Cluster 3 includes design options with lower RBOS values and higher AV values, and as can be seen in Case 3, smaller facade and roof areas result in a lower RBOS value compared to Case 1 and Case 2. Case 3 is a hybrid solution with a high AV, which is due to its small floor plates. Case 2 provides a tradeoff between Case 1 and 3 in terms of RBOS and AV values and is a hybrid design solution. Figure 4.40 shows the visualized Cases 1, 2 and 3 in InsightXR.

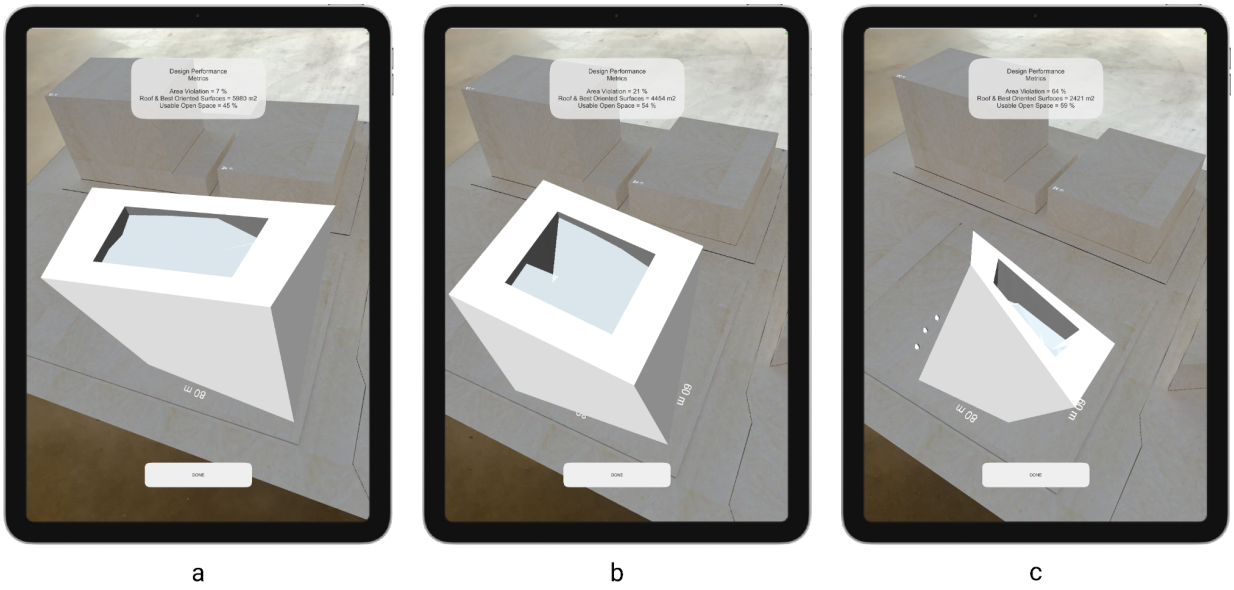


Figure 4.40. Generation 4 solutions. A. Case 1, b. Case 2, c. Case 3.

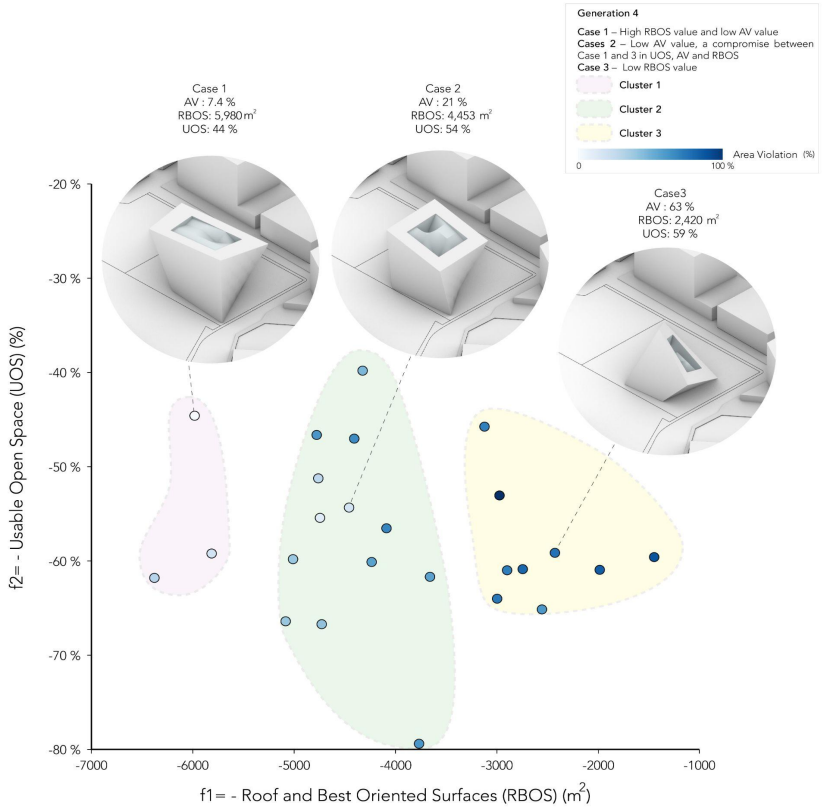


Figure 4.41. Generation 4 solutions. The third objective function, AV, is represented using color. Dark blue represents higher AV values, and light blue means lower AV values.

Generation 5

Generation 5's clusters are similar to previous generations, where cluster 1 includes design options with higher RBOS and lower AV values. Case 1 is the result of a carving operation with a deep courtyard while providing the lowest AV value among solutions. Cluster 2 includes solutions with a tradeoff between RBOS and AV values compared to the other two clusters. Cluster 3 includes design options with lower RBOS and higher AV values. Case 2 is a hybrid solution (carving and expansion) with a tilted roof orientated West. Case 2 has the lowest UOS value. Case 3, on the other hand, provides a higher UOS value with a lower RBOS value compared to the two other cases. Figure 4.42 shows the visualized Cases 1, 2 and 3 in InsightXR.



Figure 4.42. Generation 5 solutions. A. Case 1, b. Case 2, c. Case 3.

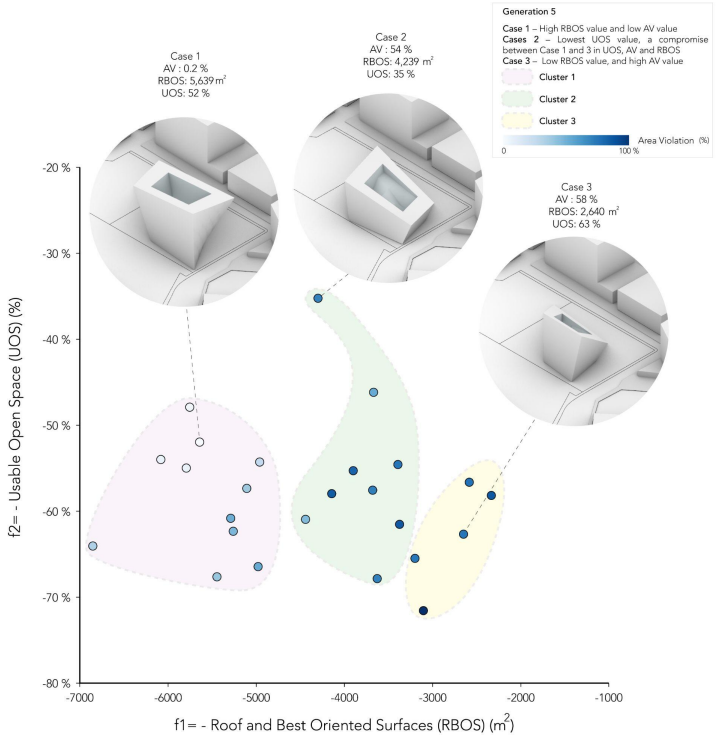


Figure 4.43. Generation 5 solutions. The third objective function, AV, is represented using color. Dark blue represents higher AV values, and light blue means lower AV values [13].

4.2.4 Discussion

Design solutions in each generation were successfully clustered into three groups, where each cluster had distinct behaviors in terms of all three objective functions of RBOS, UOS, and AV. In all 5 generations, cluster 1 included design solutions with lower AV values and higher RBOS values, which is due to design options that provide larger floor plates. As the generations progressed, design options in cluster 1 provided an improved RBOS value compared to previous generations.

Cluster 1 in generations 4 and 5 had selected cases with only carving as the design operation. Cluster 3 in all generations included solutions with the lowest RBOS and highest AV values. Case 3 in most generations was a hybrid solution (except for generation 1, where Case 3 is an expansion operation result). Cluster 2 in all generations provides a compromise between RBOS and AV values, where the selected Case 2 solution from each generation was a hybrid option. The algorithm was able to successfully cluster each generation's solutions and select a representative case among the solutions to be viewed by the users in InsightXR. In addition, InsightXR's application was able to successfully view each generation's selected design solutions in real-time. The proposed generative design process can also be non-realtime to allow users to visualize design alternatives.

In this case study, even though the goal is to maximize RBOS and minimize AV, it is essential to note that creating design options with smaller courtyards while having a deep floor plate (e.g., case 1 in generation 4) can reduce the access to daylighting and natural ventilation, which can result in non-desirable architectural design. It is ultimately the designer's role to evaluate the proposed non-dominated solutions provided by InsightXR to choose the most applicable design for the design at hand.

Future research can investigate the inclusion of user feedback in the optimization process using an interactive NSGA-II algorithm in InsightXR. In addition, research needs to be conducted to understand the perception of users regarding InsightXR and understand the opportunities and pain points of using InsightXR. In addition, this research selected five generations as the termination criterion for the algorithm, which can also be tuned or defined with respect to other criteria in future research.

4.2.5 Experiment Conclusions

InsightXR enables designers and non-designers to participate in an operative generative design process using a novel NSGA-II algorithm in real-time, remotely, or in person [13]. The proposed generative design process can also be non-realtime to allow users to visualize design alternatives. Using InsightXR, users can visualize representative solutions for each generation and participate in the generative design process through subjective evaluation of proposed designs over the generations. This research successfully introduced a clustering method based on k-means clustering using the Scikit-learn Python module [176]. This was combined with a shape similarity algorithm that allows for selecting representative 3D massing solutions to be presented to the users for further investigation, reducing user fatigue that can be caused by the presentation of excessive marginally differing solutions per generation [176]. To understand the capabilities of InsightXR in selecting and visualizing representative design solutions per generation, a case study experiment was conducted using three objective functions: RBOS, UOS, and AV, to generate early-stage massing design options on a site in San Francisco, USA. Each cluster per generation successfully created a set of unique options with respect to the constraints and metrics. At each generation, the three presented options spanned the spectrum of the different metrics, RBOS, AV, and UOS, presenting one at each extreme as well as a hybrid option. This would allow a designer to properly choose a design massing that will meet both objective and subjective goals, exploring the space of options across generations of the genetic algorithm.

4.3 Experiment 3: Human Subject Experiments

4.3.1 Introduction

Participation in the design process can help a community reach democratic decisions for their community, improve the quality of proposed architectural designs, and avoid early design mistakes by including people who are impacted by those decisions [213]. However, as discussed by Yan and Tamke, architects rarely seek feedback until they have already completed an initial design and associated visualization [243]. The inclusion of user feedback is traditionally ambiguous and in direct conflict with architects' expectations for their role in the process and the culture that pervades the architecture industry [17], [238], [243]. In addition, the theories associated with the impact of participatory design question power dynamics and the political essence of architecture [17].

Analysis of 2D images using visual quality metrics has been explored significantly more extensively than 3D designs [18]. Quality metrics for images can be categorized into three main categories, including full-reference (FR), reduced reference (RR), and no-reference (NR) evaluation [48]. Some FR metrics include Signal-oriented metrics and Single-channel modeling of human visual system (HVS) [48]. When working with static 3D meshes, it is possible to categorize approaches associated with human perception metrics into image-based and model-based metrics [48]. Image-based metrics are view-dependent, where images of multiple views of the 3D objects are analyzed using image-based metrics [48]. However, this approach is not reliable due to previous experiments' findings on differences in visual perceptions between 3D objects and 2D images from different views [48]. In the view-independent category, geometry is used to predict perceptual metrics such as impairments [48]. An example of an existing method in mesh visual quality assessment consisted of generating 2D images of different views and then using convolutional neural networks to analyze each image [1]. In 2019, Bhat et al. explored machine learning algorithms such as Support Vector Machines, Gradient Boosting, and Deep Neural Networks to explore replicating a human's subjective preferences for a visual quality assessment of a 3D chair object [18]. An important distinction between this research area and architectural massing is the great increase in complexity of such massings. In architectural massing, preferences cannot be effectively assessed by creating 2D images of the massing and gathering ratings or using the existing machine-learning models for standalone 3D objects. Architectural massings are complex and represent a program, respond to the context, and have environmental impacts that architects consider when assessing massings and, therefore, accordingly, need a more in-depth assessment strategy.

This section assesses the inclusion of participants in the generative design process using InsightXR. Despite architects' expectations about their roles as exclusive designers and decision-makers in the design process, the early assumption in this research contradicts this concept [17]. The hypothesis is that including non-experts in the early-stage design process can provide valuable feedback for the final design solutions, however, there will be significant differences in the final solutions' performance between expert and non-expert participants.

To assess the impact of including users in the generative design process using InsightXR, human-subject experiments and post-study experiments were conducted. The human-subject

experiments are titled Experiment 1, and the post-study experiments are titled Experiment 2. Experiment 1 was conducted in person at the XR Lab at UC Berkeley, lasting one hour per study. This experiment was conducted among experts and non-experts using two interventions. Note the initial generations provided to the users for both interventions were independent of each other. Experiment 2 was conducted after Experiment 1 using only the algorithm without including users, using the initial population of Experiment 1 as seeds. Below is the breakdown of each experiment and its interventions.

Experiment 1: InsightXR Human Subject Experiments

1. Introduction and initial survey
    a. Intervention 1: Generating and assessing massing options using InsightXR without providing performance metrics to users.
    b. Intervention 2: Generating and assessing massing options using InsightXR with providing performance metrics to users.
2. Final interview to collect subjective feedback on the massings, interventions, and platform.

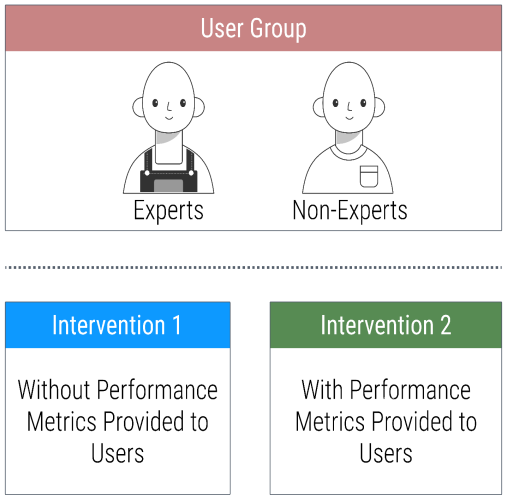


Figure 4.44. Experiment 1.

Experiment 2: Operative NSGA-II Algorithm Assessment Based on Users' Initial Populations

a. Intervention1: Generating and assessing massing options using InsightXR from the initial population seeds from users who **have not** been provided performance metrics.
b. Intervention 2: Generating and assessing massing options using InsightXR from the initial population seeds provided to users that **have** been provided with performance metrics.

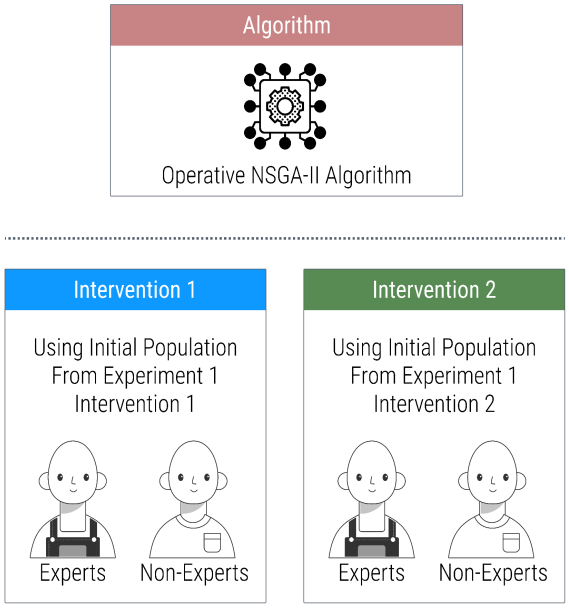


Figure 4.45. Experiment 2.

For both experiments, the following analysis is performed:

1. Hypervolume indicator assessment
2. Shape analysis focusing on height analysis to evaluate the distribution of massings:
3. Objective Function Analysis (Performance Metrics):
    a. Area Violation (AV)
    b. Non-Passive Zone (NPZ)
    c. Usable Open Space (UOS)

Figures 4.46 and 4.47 show the breakdown of both experiments.

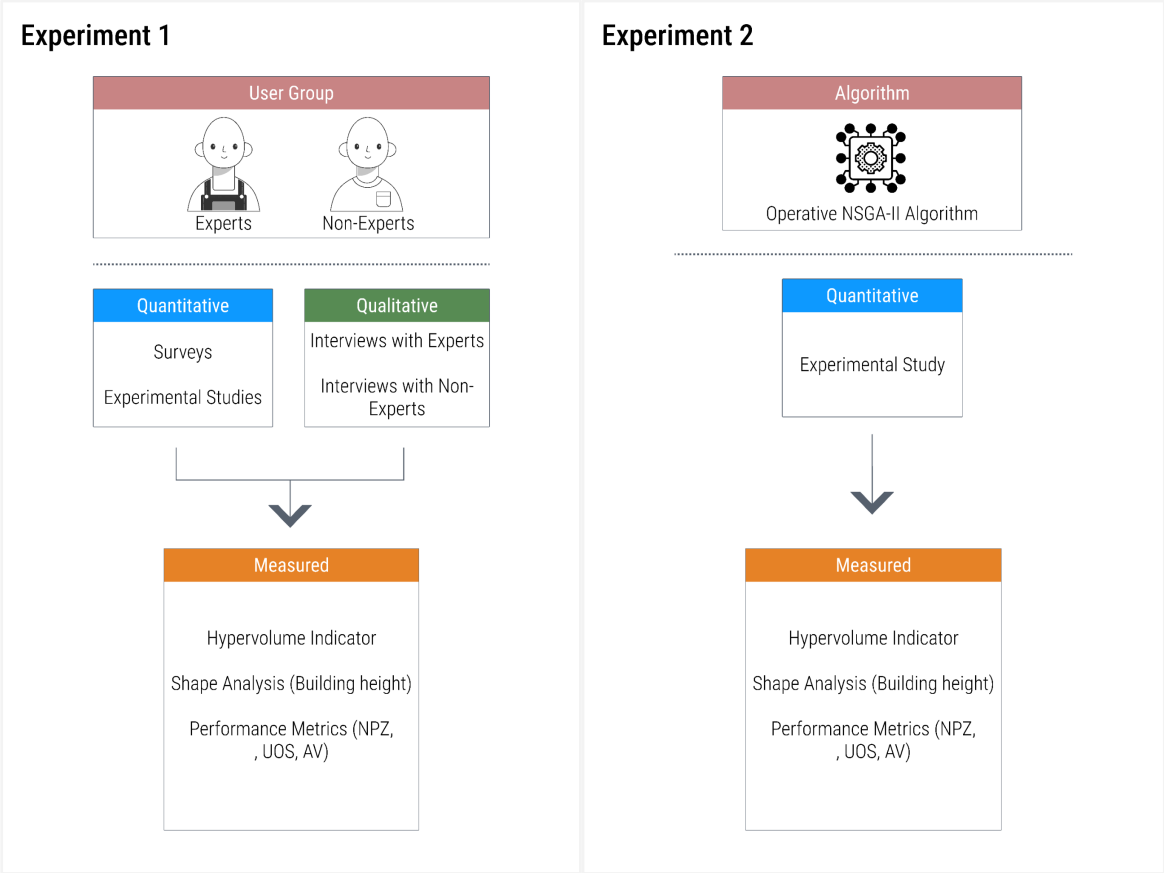


Figure 4.46. (left) Experiments 1 overview. Figure 4.47. (right) Experiments 2 overview.

The goal of the conducted experiments can be categorized into the following main subjects:

1. **Topic 1:** Assessing InsightXR's Operative Non-dominated Sorting Genetic Algorithm II's (NSGA-II) performance.

   Topic 1 Research Questions:

   a. Are the geometries optimized correctly (technical assessment of the tool)?

2. **Topic 2:** Understanding the impacts of including users (experts and non-experts) in the generative design process.

   Topic 2 Research Questions:

   a. How does the algorithm's performance compare with and without users?

3. **Topic 3:** Understanding the differences in highly rated generated massing options among experts and non-experts.

Topic 3 Research Questions:

a. How do the final non-dominated solutions compare among experts and non-experts?

4. **Topic 4:** Understanding the impacts of providing performance metrics in the early-stage generative design process to users.

Topic 4 Research Questions:

a. Is there a difference between experts' and non-experts' NSGA-II optimization performance with and without providing explicit performance metric information to users?

b. What are the experts' and non-experts' perceptions of the proposed objective functions?

4.3.2 Experiment 1

Experiment 1 uses the operative NSGA-II algorithm discussed in section 1.3 as the framework to generate massing options in InsightXR. Experiment 1 uses the following objective functions: 1. AV, 2. NPZ, 3. UOS, 4. User Rating. These experiments use a population size of 30, simulated binary crossover (SBX) with probability 0.9, and polynomial mutation (PM) with eta 10. The optimization is terminated after ten generations. The calculations' values are negated for maximization problems (UOS and Rating).

The selected users were a combination of architectural design experts and non-experts. They were recruited from undergraduate and graduate students at UC Berkeley. The experts (n=21) were selected from students with a strong architectural design background. To qualify to participate as experts in Experiment 1, the participants had to be architecture or city planning graduate students who had already completed an undergraduate degree in architecture. For graduate students without an undergraduate degree in architecture, at least two years of professional experience and architectural education were required to participate. This selection criteria are due to the importance of expertise in massing in architecture for experiment 1. The selected experts were two PhD students in architecture, eleven Master of Architecture II (MArch II) students, four Master of Architecture III (MArch III) students (with at least two years of experience in architectural design), three Master of City Planning (MCP) students, and one MS in Architecture Student. Figure 4.48 shows the breakdown of experts.

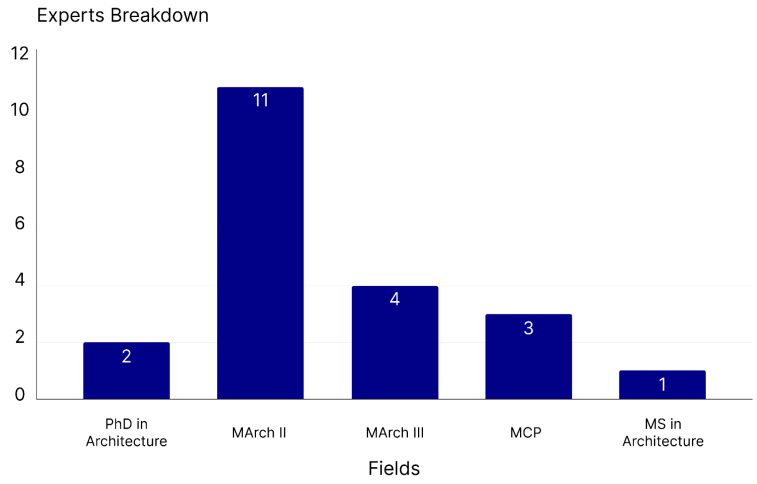


Figure 4.48. Experiment 1 experts' background.

To qualify to participate as a non-expert in Experiment 1, the participants had to be current students at UC Berkeley from any field of study without a strong background in architecture. This classification includes undergraduate students in architecture and other fields of study at UC Berkeley. The selected non-experts (n=21) were one PhD in architecture student (without an architecture background), four Bachelor of Arts in Architecture (BA in Architecture) students, three Master of Landscape Architecture (MLA) students, one Bachelor of Science in Computer Science (BS in CS) student, one first-year MArch III student (without the two-year requirement background in architecture), four Bachelor of Urban Studies students and one Bachelor of Physics student. Figure 4.49 shows the breakdown of non-experts. The Bachelor of Urban Studies and Physics fields are categorized as other fields. The total number of subjects, including experts and non-experts, was 42 (n=42).

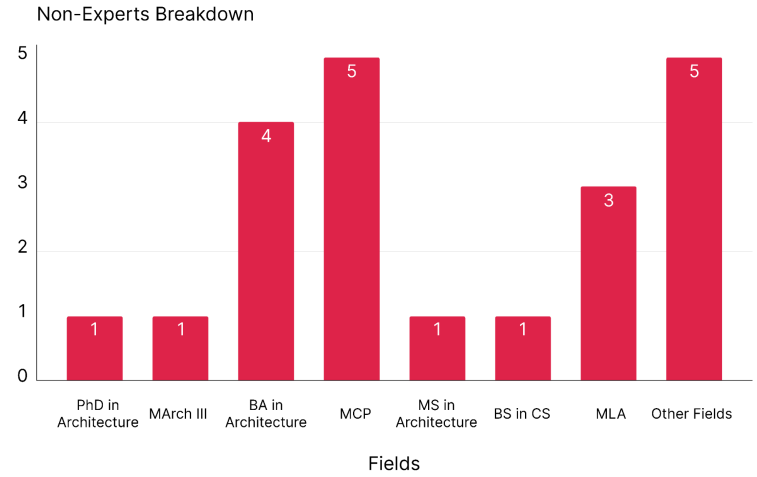


Figure 4.49. Experiment 1 non-experts' background.

From the overall participant population, twenty one percent of participants were very familiar with the Zoning Codes, and thirty one percent were very familiar with the building codes, all of whom came from the expert participants. On the other hand, fourteen percent of the population were not familiar at all with the zoning codes, and seventeen percent were not familiar at all with the building codes, all of whom belonged to the non-expert participants. Among participants, fifty-seven percent were not familiar at all with mixed reality, twenty-one percent were not familiar at all with virtual reality, and thirty-six percent were not familiar at all with augmented reality. Overall, most participants had little experience with virtual and augmented reality technology.

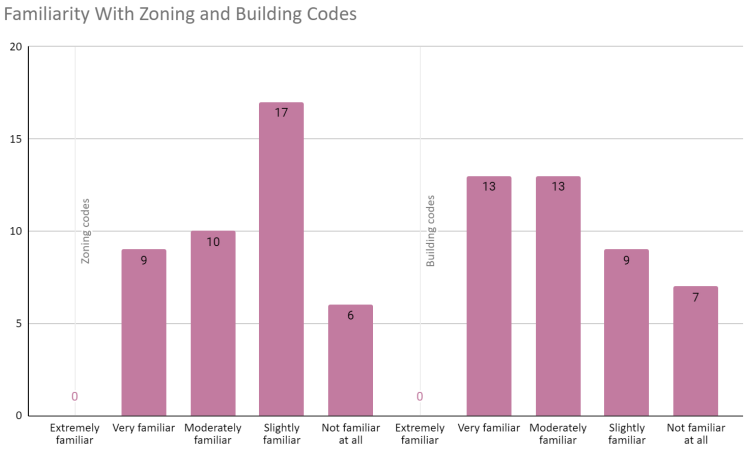


Figure 4.50. Experiment 1 participants' familiarity with zoning and building codes.

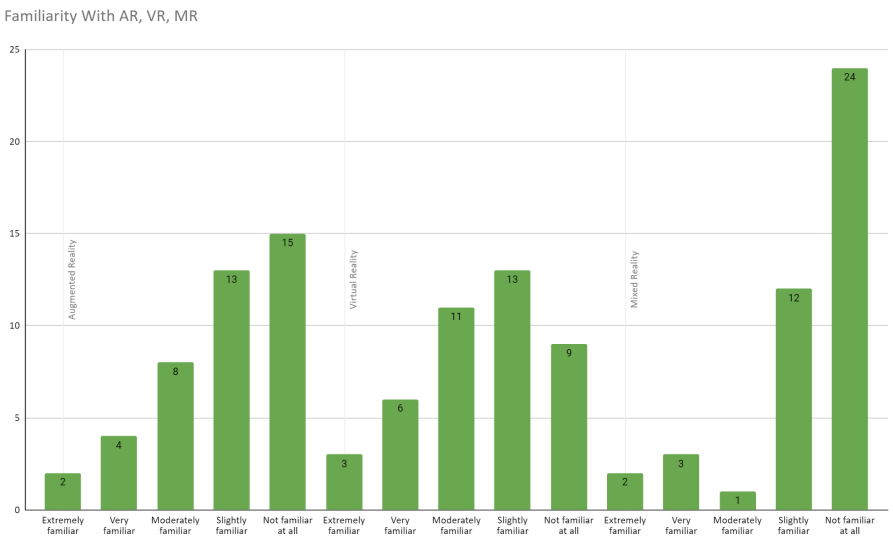


Figure 4.51. Experiment 1 participants' familiarity with AR, VR, and MR.

Experiment 1 Set Up

During the human subject experiments, a case study was introduced to users as a framework for generating massing options using InsightXR. The project's case study site is located at 74 Mission Rock St, San Francisco, CA 94158, as discussed in sections 4.1 and 4.2. A design prompt was shared with participants for the case study. This case study aims to design early-stage massing options for a new mixed-use building in San Francisco.

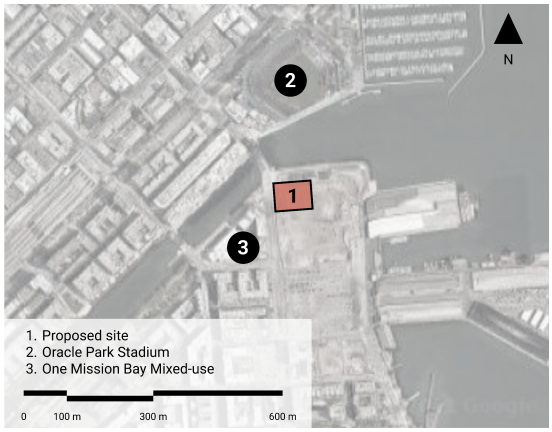




Figure 4.52. (left) Building Site. Figure 4.53. (right) Site surrounding.

The provided site is 100 by 115 meters, with a maximum height limit of 60 meters for the building. The goal is to design a mixed-use project that provides a minimum of 150 residential units, at 30,000 m^2 and 10,000 m^2 of commercial space. The maximum desired built area of the building is estimated at 40,000 m^2. Figure 4.54 shows the site used in this case study.

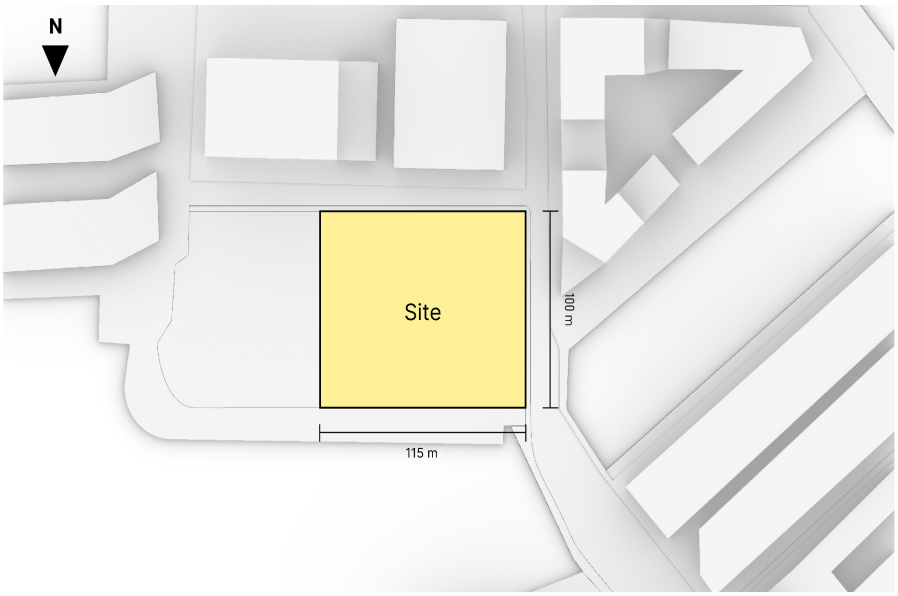


Figure 4.54. Building Site.

Table 4.5. Building Program.

Program	Area
Residential Units	150 units needed (30,000 m^2)
Commercial	10,000 m^2
	40,000 m^2

Massing

The generative design process in this experiment uses operative design in volumes, as discussed in Operative Generative Design Using Non-Dominated Sorting Genetic Algorithm II (NSGA-II) [13]. The operations selected in this experiment are carving and expansion. Carving is introduced in two types: courtyards and corner cuts, and expansion uses any of the four roof corner points of the massing to create an expansion operation.

The variables introduced in this experiment are the same as sections 4 and 5, and the movements of initial points lead to resulting solutions [13]. All variables are in (x,y,z) format, and each coordinate is in meters. The variables defined in the NSGA-II algorithm in this experiment include the following:

Overall defining points' movements:

- p1 (base point 1) movement (x,y,z)
- p2 (base point 2) movement
- p3 (base point 3) movement
- p4 (base point 4) movement
- p5 (roof point 5) movement
- p6 (roof point 6) movement
- p7 (roof point 7) movement
- p8 (roof point 8) movement

Courtyard defining points' movements:

- p9, p10, p11, p12, p13, p14, p15, p16, p17, p18, p19, p20, p21, p22, p23, p24

Upper courtyard rim points' movements:

- p9', p10', p11', p12', p13', p16', p17', p20', p21', p22', p23', p24'

The design operations and the resulting positions of the vertices will result in differences in fitness values, including user ratings when visualizing the massing options. As an example, the maximum and minimum values selected for the variables allow the massing to have deep floor plates versus shallow floor plates, which can directly lead to higher or lower NPZ values in the massing options. In addition, depending on the final position of the vertices in the upper surface of the massing, carving can be applied in the form of a courtyard which can help reduce the amount of NPZ value in a massing option. The courtyard portion of the building in the proposed massing is a glass-covered atrium space that can act as a public space for building occupants. In addition, the positioning of the vertices on the lower level can directly impact the UOS values. Corner cuts in the massing reduce the overall gross square footage of the building while increasing the UOS value. The corner cut is generated to connect the outdoor areas to the inner courtyard area and help draw pedestrians to explore the inner public space of the building.

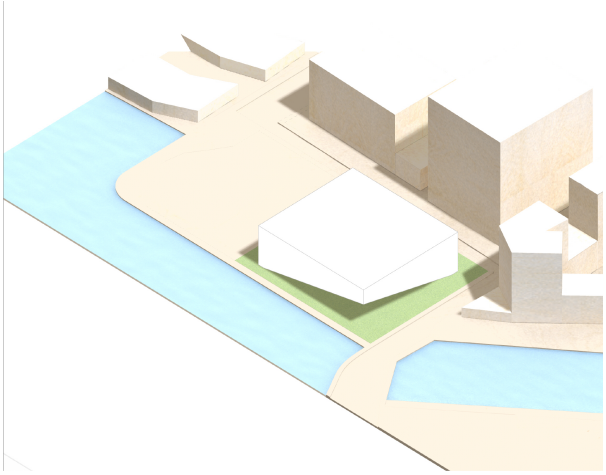


Figure 4.55. Carving as the corner cut concept.

Objectives

Experiment 1 uses the following objective functions, which will be discussed in further detail in this section: AV, NPZ, UOS, and User Rating. These experiments use a population size of 30, simulated binary crossover (SBX) with probability 0.9, and polynomial mutation (PM) with eta 10. The optimization is terminated after ten generations. The calculations' values are negated for maximization problems (UOS and User Rating).

1. **AV.** Experiment 1 uses AV as an objective function. The AV objective function is the difference between the goal area (40,000 m²) and the total floor area of each individual in each generation, divided by the goal area of each individual and multiplied by 100 to create a percentage [13]. The absolute value is used in the calculations. In this objective, 40,000 square meters is the goal area, and GFA is the Gross Floor Area of the building generated in square meters [13].

$$f1 = 100 * |40000 - GFA| / 40000 \quad (1)$$

2. **NPZ.** To calculate the NPZ, the floor slabs are defined on each floor along with the amount of carving on that floor, if it exists. Then the amount of non-passive area for each floor is calculated using subtraction [13]. The final result is turned into a percentage, dividing the resulting total non-passive area by the Gross Floor Area. The division results in the NPZ value [13].

   Similar to sections 4.1 and 4.2, H represents the height of one level within the massing. In the calculations, the first floor is considered as 4 meters, and the rest of the floors are to be 3 meters (floor to floor). In this objective function, Non-PassiveArea represents the total non-passive areas in the building in square meters, and GFA represents the Gross Floor Area in square meters. The final result is shown as a percentage.

$$f2 = 100 * (Non - PassiveArea)/(GFA) \quad (2)$$

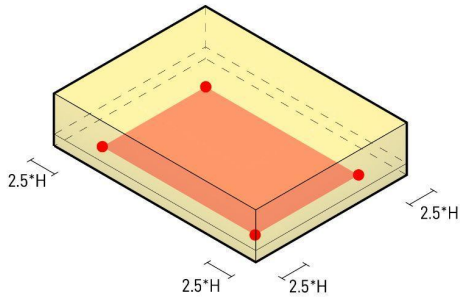


Figure 4.56. Non-passive area on the first-floor representation [13].

3. **UOS.** In these experiments, the percentage of the unbuilt area is calculated on the site as UOS, and the goal is to maximize UOS for improved user experience. The results are negated for this maximization objective function. AL is the area of the lot in square meters, and footprint is the first-floor area of the building in square meters [13]. In this calculation, if the courtyard roof reaches the first floor, the courtyard area is subtracted from the footprint.

$$f3 = -100 * (AL - footprint - courtyard)/AL \quad (3)$$

4. **User Rating.** User rating is a value provided by users to rate their aesthetic preference on each massing option. The values given by users are integers from -3 to +3, with -3 being the least desirable and +3 the most desirable massing. The users rate three options per generation, and the ratings are then predictively applied to all massing options in the generation based on their similarity to each rated massing option by users.

Once the first generation of massing options is generated based on the defined objective functions using the Operative NSGA-II algorithm, a k-means clustering algorithm is applied to the population to cluster the individuals into three clusters [12], [176]. The k-means clustering evaluates the clusters based on the shape similarity values. This value is calculated using the sum of Euclidean distances between each vertex in the mesh [16]. The shape similarity values are then used to rank each cluster's solutions with respect to distance from the cluster centroid. Then from each cluster, the most representative solution is selected to be viewed and rated by users. [12]. For easier calculations in the algorithm, the user rating number is mapped into a 0 to 6 scale and negated as a maximization problem. Once the rating is provided for all three cluster centroid options in a generation, the user's rating for the other 27 solutions is estimated based on their similarity to each of the three massing options. This process continues for ten generations.

Table 4.6. Experiments' definitions

Experiment	Objectives	Site
1 Intervention1	AV* (1), NPZ (2), UOS (3), User Rating (4)	100 m x 115 m
1 Intervention 2	AV (1), NPZ (2), UOS (3), User Rating (4)	
2 Intervention 1	AV (1), NPZ (2), UOS (3)	
2 Intervention 2	AV (1), NPZ (2), UOS (3)	

*Area Violation (AV) in these experiments is the absolute value of 40,000 m²- Gross Floor Area.

Experiment 1 Intervention1: Without Performance Metrics Provided To Users

In Experiment 1 Intervention 1, each expert and non-expert views and rates ten generations of massing options using InsightXR based on their subjective feedback on a scale of 7, with -3 as the least preferable (desirable) and +3 as the most preferable (desirable) design. This rating is based on their aesthetic preferences. The massing options are viewed from a bird' s-eye and a first-person view.

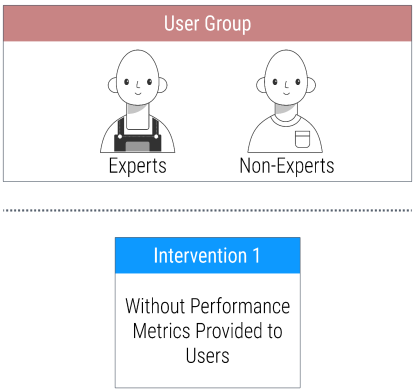


Figure 4.57. Experiment 1 Intervention 1.

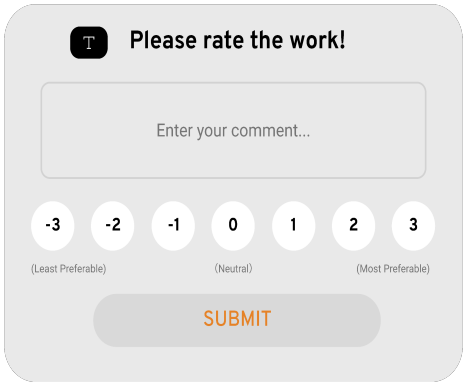


Figure 4.58. Rating Interface in InsightXR.

Experiment 1 Intervention 2: With Performance Metrics Provided To Users

In Experiment 1 Intervention 2, users consider the performance metrics (objective function values) while reviewing each design option and rate the designs while considering these conflicting performance metrics. The performance metrics for each design option are overlaid onto the augmented reality interface while the user views the massing.

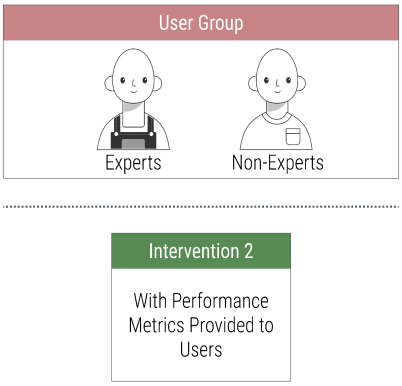


Figure 4.59. Experiment 1 Intervention 2.

The objective function values are color-coded to help users understand the metrics better. The three colors are green, yellow, and red. Green represents desirable values (low AV, low NPZ, and high UOS). Yellow represents mid-range values for AV, NPZ, and UOS. Red represents undesirable values (high AV, high NPZ, and low UOS).

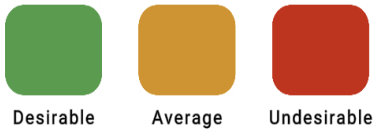


Figure 4.60. Color coding used for performance metrics.

4.3.3 Experiment 2: Algorithm Experiments Without Users (NSGA-II Alone)

In the next step, each user's initial population and seed values (Random seed values generated for a population used to randomize the generated initial generation massings) are used as seeds to run algorithms for ten generations with the same settings as experiment 1's algorithm, without human input. Overall, eighty-four optimization-only runs are executed for experiment 2. Forty-two algorithms are executed based on experts' Experiment 1 Intervention 1 and 2 results, and forty-two algorithms are executed based on non-experts' Experiment 1 Intervention 1 and 2 results.

The algorithms are executed using AV, NPZ, and UOS as objective functions. All algorithm parameters are kept the same. In these experiments, a population size of 30, simulated binary crossover (SBX) with probability 0.9, and polynomial mutation (PM) with eta 10 are used. In ten generations, the optimization is terminated. As before, for maximization problems (UOS), the values for the calculations are negated. The final results are saved as algorithm counterparts for each user type (experts and non-experts). Figure 4.61 shows the overall components of Experiment 2.

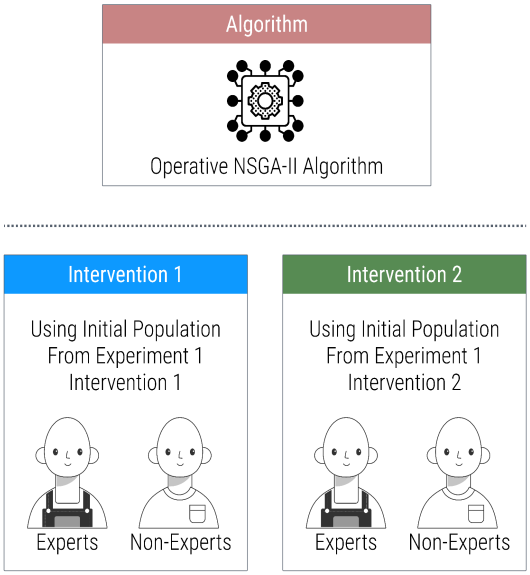


Figure 4.61. Experiment 2 Overview.

Experiment 2 can be defined as the following:

1. Experiment 2 Intervention 1: InsightXR is run without users, using the initial population (experts and non-experts) from Experiment 1 Intervention 1 as a seed. The objective functions are 1. AV, 2. NPZ, and 3. UOS.

2. Experiment 2 Intervention 2: InsightXR with no users, using the initial population from Experiment 1 Intervention 2 as a seed (experts and non-experts). The objective functions are 1. AV, 2. NPZ, and 3. UOS.

4.3.4 Results

This section will discuss the results of experiments one and two. These results will be quantitatively discussed and presented by comparing experts, non-experts, and the algorithm in the following areas:

1. Hypervolume indicator assessment
2. Shape analysis focusing on height analysis to evaluate the distribution of massings:
3. Objective Function Analysis (Performance Metrics):
    a. Area Violation (AV)
    b. Non-Passive Zone (NPZ)
    c. Usable Open Space (UOS)

In addition, the highly rated solutions by users in Experiment 1 Intervention 1 and 2 and the highest-performing solutions generated by the algorithm without user ratings (Experiment 2 Intervention 1 and 2) will be discussed.

Overview of Analysis Methods

For each analyzed data set, the distribution is first assessed using the Shapiro–Wilk test using the Scipy package in Python [197]. In all calculations, the alpha level is set to 0.05, and based on the distribution of the data, an independent t-test is conducted (when the data is normally distributed). When the data is not normally distributed, a Mann-Whitney U Test is performed [26]. The resulting p-value based on the appropriate test is reported, which is calculated with a threshold of 0.05 for reporting statistical significance (p-value $\leq$ 0.05). In addition, Cohen D's (effect size) will be reported, as well as statistical power, when the results are found to be statistically significant [46].

Highly Rated Solutions by Users

First, the highly rated solutions by users will be discussed to better understand the final solutions generated using InsightXR. Then the performance of algorithms will be assessed using the discussed methods. Two sample users' results from Experiment 1 and 2 are discussed using their final non-dominated solutions. However, all the other experiment output is also included in appendices.

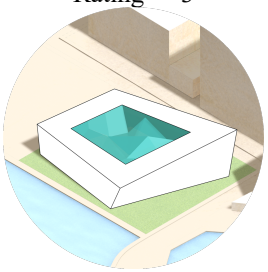
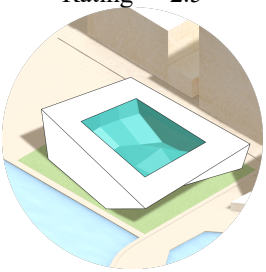
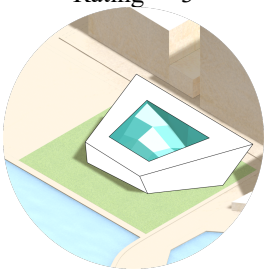
Three design options from each experiment are visualized for the sample users. For Experiment 1, the top three highest-rated solutions (by the user) are visualized. For Experiment 2, the top three design options based on the Pseudo-Weight Vector are visualized [59]. The Pseudo-Weight Vector method in this experiment calculates the distance to the worst solution for each objective function and uses Pymoo's PseudoWeights function [58], [19]. This method allows for finding the top-performing solutions based on their distance to the worst solution for each objective. The weight associated with all objectives is set equally in this experiment [19].

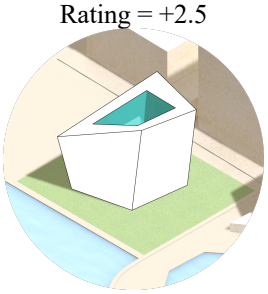
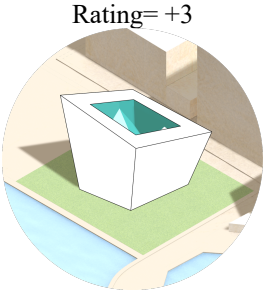
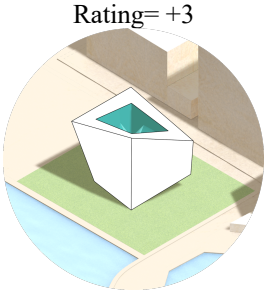
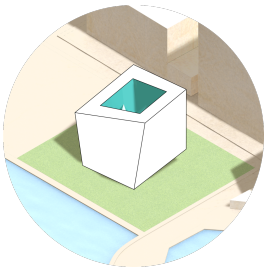
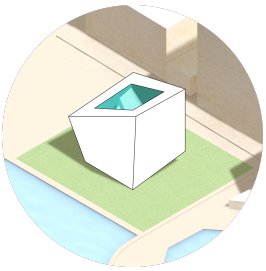
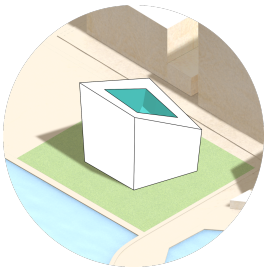
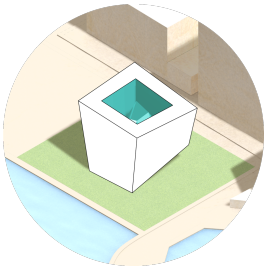
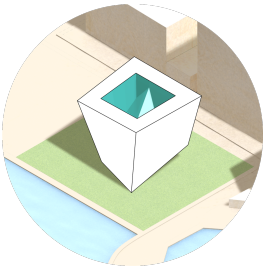
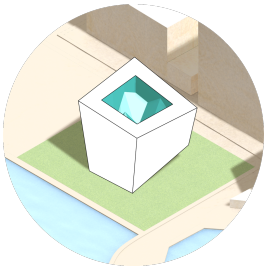
Assessing the highest-rated solutions will help understand the impacts of providing performance metrics to users and find the differences between what the algorithm generates with and without the user rating as an objective.

The example expert user (AA1) and non-expert user (2G2) are selected for further discussion in this section. In Experiment 1 Intervention 1, expert AA1 rated buildings with fewer stories and higher site coverage (lower UOS values) as their highest-ranked solutions. The highly rated solutions in Experiment 1 Intervention 1 have UOS values between 43% to 69% and an NPZ value between 20% to 30%.

However, when provided with performance metrics in Intervention 2, the user rating increased for buildings with more stories (taller buildings) and higher UOS values. In order to compare user and algorithm results, the algorithm-based results (Experiment 2) are selected for visualization using the pseudo-weight vector approach. For Experiment 2 Interventions 1 and 2, the visualized solutions consistently have low NPZ values and high UOS values, with their AV ranging from 7% to 31%, resulting in convergent massings but with different orientations. The algorithm's high-performing solutions in Experiment 2 resemble the highly rated options in Experiment 1 Intervention 2, which show the impacts of providing performance metrics to users. The tables below show expert user AA1's solutions. Other expert results can be found in Appendix A.

Table 4.7. Expert AA1 Experiment 1 and 2 Results.

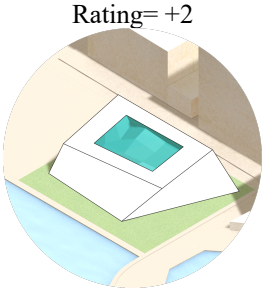
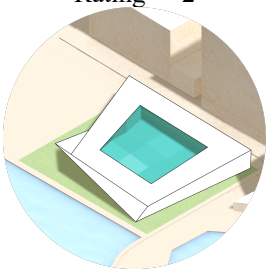
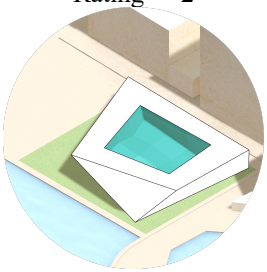
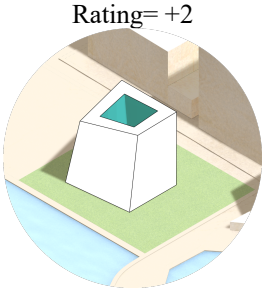
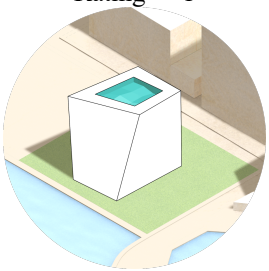
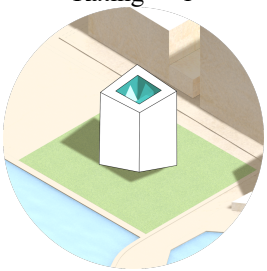
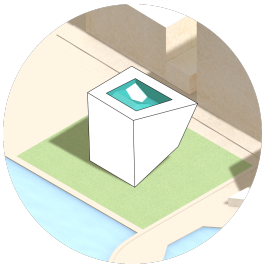
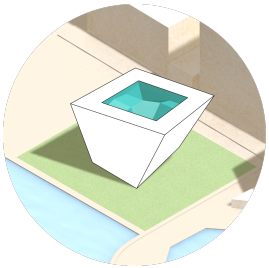
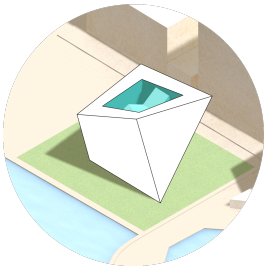
Participant	Experiment	Design Options		
AA1	Experiment 1 Intervention 1	Option 8 AV= 13% NPZ= 20% UOS= 51% Rating= +3 	Option 7 AV = 11% NPZ= 26% UOS= 43% Rating= +2.5 	Option 4 AV= 22% NPZ= 30% UOS= 69% Rating= +3 

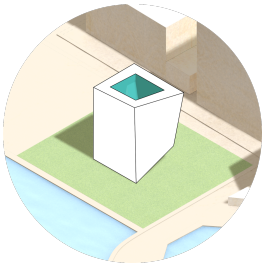
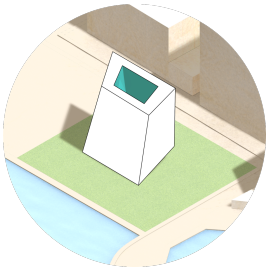
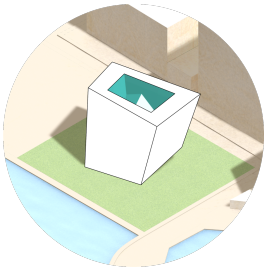
Participant	Experiment	Design Options		
	Experiment 1 Intervention 2	Option 9 AV= 2% NPZ= 24% UOS= 83% Rating = +2.5 	Option 13 AV= 6% NPZ= 25 UOS = 83% Rating= +3 	Option 2 AV= 23% NPZ= 1% UOS= 88% Rating= +3 
	Experiment 2 Intervention 1	Option 7 AV= 7% NPZ= 4% UOS= 87% 	Option 6 AV= 27% NPZ= 2% UOS= 88% 	Option 9 AV= 31% NPZ= 0.0 UOS= 80% 
	Experiment 2 Intervention 2	Option 7 AV= 15% NPZ= 1% UOS= 86% 	Option 9 AV=11% NPZ= 4% UOS= 88% 	Option 6 AV= 10% NPZ= 1% UOS= 86% 

In Experiment 1 Intervention 1, non-Expert 2G2 rated buildings with lower UOS values ranging between 32% to 50% and higher NPZ values ranging between 38% to 49% as their highest-rated solutions. However, when provided with performance metrics in Intervention 2, the user rating increased for buildings with more stories (taller buildings) and higher UOS values ranging between 81% to 92%. For Experiment 2 Interventions 1 and 2, the visualized solutions have NPZ values ranging between 0 to 38% and high UOS values ranging between 84% to 92%. The AV values

range between 0% to 38%. The algorithm's high-performing massings in Experiment 2 resemble the highly rated options in Experiment 1 Intervention 2, which again mirrors the expert's results after providing performance metrics to users. The tables below show non-expert user 2G2's solutions. Other non-expert results can be found in Appendix A.

Table 4.8. Non-Expert 2G2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2G2	Experiment 1 Intervention 1	Option 0 AV= 2% NPZ= 49% UOS= 50% Rating= +2 	Option 1 AV= 20% NPZ= 42% UOS= 32% Rating= +2 	Option 7 AV= 33% NPZ= 38% UOS= 45% Rating= +2 
	Experiment 1 Intervention 2	Option 1 AV= 13% NPZ= 19% UOS= 82% Rating= +2 	Option 8 AV= 11% NPZ= 42% UOS= 81% Rating= +1 	Option 5 AV= 46% NPZ= 13% UOS= 92% Rating= +1 
	Experiment 2 Intervention 1	Option 8 AV= 23% NPZ= 38% UOS= 91% 	Option 9 AV= 3% NPZ= 4% UOS= 90% 	Option 10 AV= 0 NPZ= 5% UOS= 84% 

Participant	Experiment	Design Options		
	Experiment 2 Intervention 2	Option 15 AV= 38% NPZ= 0 UOS= 92% 	Option 20 AV= 32% NPZ= 2% UOS= 91% 	Option 21 AV= 11% NPZ= 11% UOS= 90% 

Highly Rated Solutions Analysis

1. **Experts Highly Rated Solutions**

Based on Experiment 1's results, the experts' ratings for final solutions in Experiment 1 Intervention 1 are not normally distributed. In addition, the solutions' (building) heights are normally distributed in Experiment 1 Intervention 1. Figure 4.62 below shows the building height distribution for the expert group in Experiment 1 Intervention 1. The majority of solutions rated highest are below the median building height.

Experts' ratings for final solutions in Experiment 1 Intervention 2 are not normally distributed as well. In addition, the final solutions' (building) heights are not normally distributed once again. The figure below shows the building height distribution for the expert group in Experiment 1, Intervention 2. However, with the introduction of showing the metrics to the users, the majority of experts' highest-rated solutions are now above the median building height.

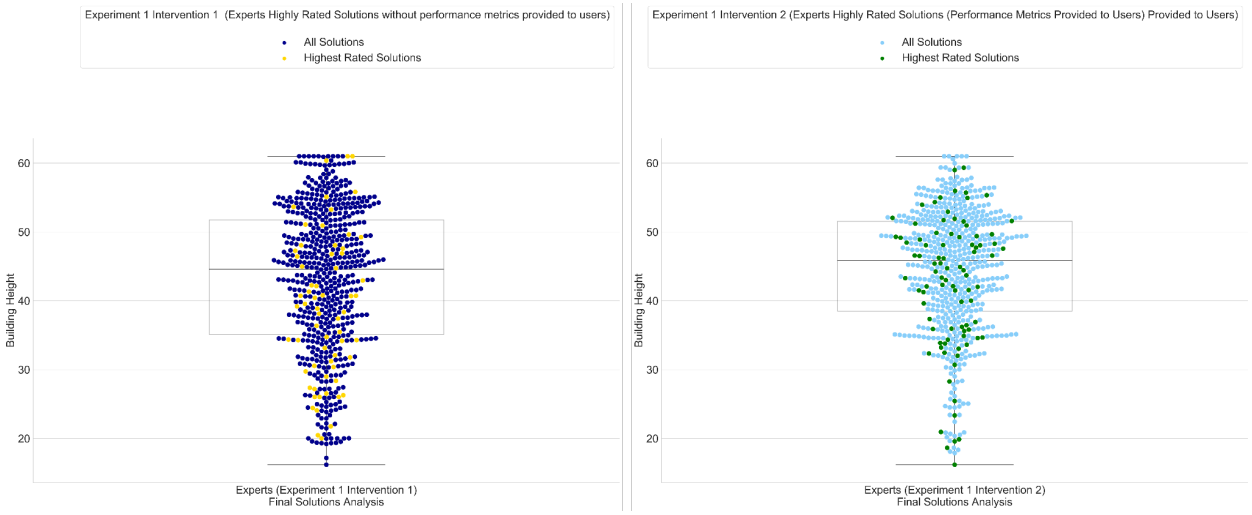


Figure 4.62. Expert's highly rated solutions in Experiment 1 Intervention 1 and 2.

2. Non-Experts Highly Rated Solutions

Figure 4.63 shows the building height distribution for the non-expert group in Experiment 1, Intervention 1. Non-experts' ratings for final solutions in Experiment 1 Intervention 1 are normally distributed. In addition, the solutions' (building) heights are normally distributed. As shown in the figure below, most solutions rated highest by the non-experts are below the median building height. Based on Intervention 2's results, the non-experts' ratings for final solutions are not normally distributed again, and the solutions' (building) heights are not normally distributed. The figure below shows the building height distribution for the non-expert group in Experiment 1, Intervention 2. In intervention 2, like with the experts, more of the high-rated results are now above the median height. For the non-experts, the median itself also increased substantially when presented with the metrics.

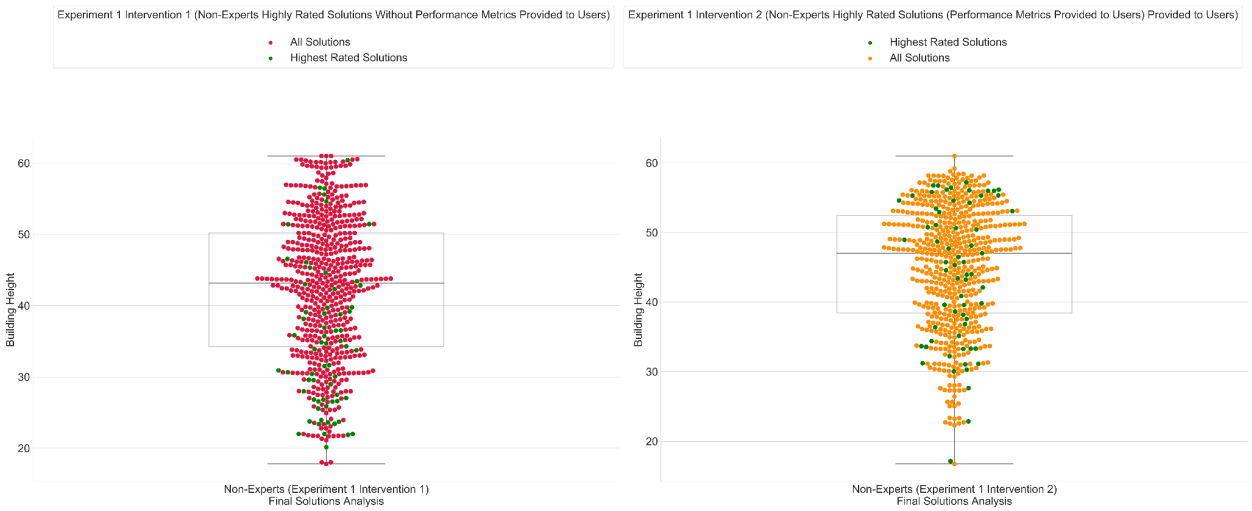


Figure 4.63. Non-expert's highly rated solutions in Experiment 1 Intervention 1 and 2.

The Hypervolume Indicator: Assessment of Non-Dominated Solutions' Performance

The Hypervolume (HV) Indicator is a method commonly used to assess the performance of multi-objective optimizations [19]. In this experiment, we used Pymoo's method of calculating Hypervolume based on DEAP [296]. The HV calculation in this experiment uses a dimension-sweep algorithm for finding the HV indicator values for a multi-objective optimization problem [84]. The HV is calculated using the area associated with the dominated solutions, where a higher HV value represents a higher quality of the non-dominated solutions [19]. An important benefit of the HV Indicator as a performance metric is the sole requirement to define a Reference Point for calculation purposes [19]. In these experiments, the hypervolume is calculated and normalized for each generation as well as the final solution space. The reference point is selected as a point 0.01 larger than 1 in each dimension. Figure 4.64 below shows the definition of the HV Indicator for a two-objective multi-objective optimization example [84]. For multi-objective optimizations, higher HV values are desirable.

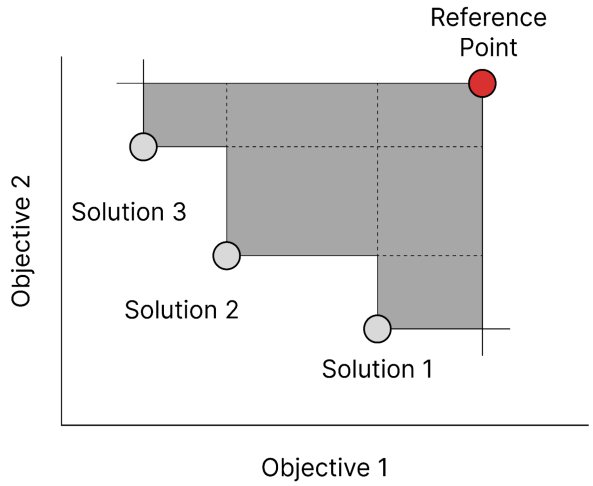


Figure 4.64. Hypervolume (HV) calculation example [84], figure recreated by author.

1. **Impacts of Providing Performance Metrics Among Experts**

To assess the HV values among experts with and without providing performance metrics (Experiment 1 Intervention 1 and 2), the normalized final solutions' HV value for experts for both interventions are calculated and visualized in figure 4.65.

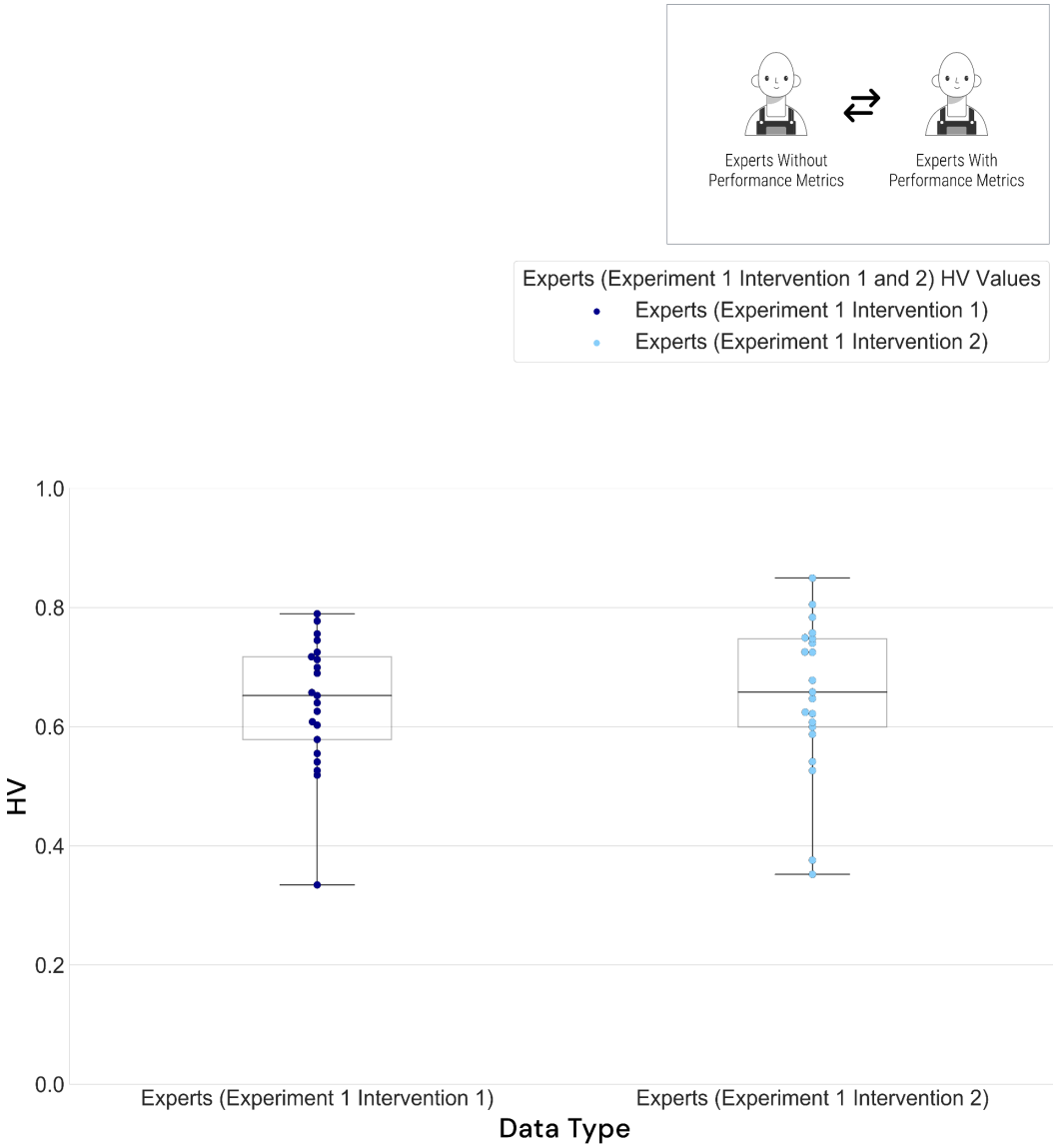


Figure 4.65. Experts' final HV values Experiment 1 Intervention 1 and 2.

The HV values for experts in Experiment 1, Intervention 1 and 2 are normally distributed. Using the t-test, the p-value resulted in 0.7, demonstrating that there is no significant difference between the performance of non-dominated solutions among experts with and without providing performance metrics. This result demonstrates the consistency of experts' preference impact on the optimization results, and how final solutions tend to perform similarly when provided with performance metrics. This result can be further investigated by visualizing all experts' HV values' changes over ten generations for all experiments. Figures 4.66 and 4.67 show the changes in HV values over time for all experts. As an example, for expert KK1, both interventions had similar behavior in HV values over time. For some experts, Intervention 2 resulted in lower HV values over time, such as experts DD1 and PP1. Some experts' HV values in Intervention 2 improved over time compared to their Intervention 1 results, such as expert QQ1.

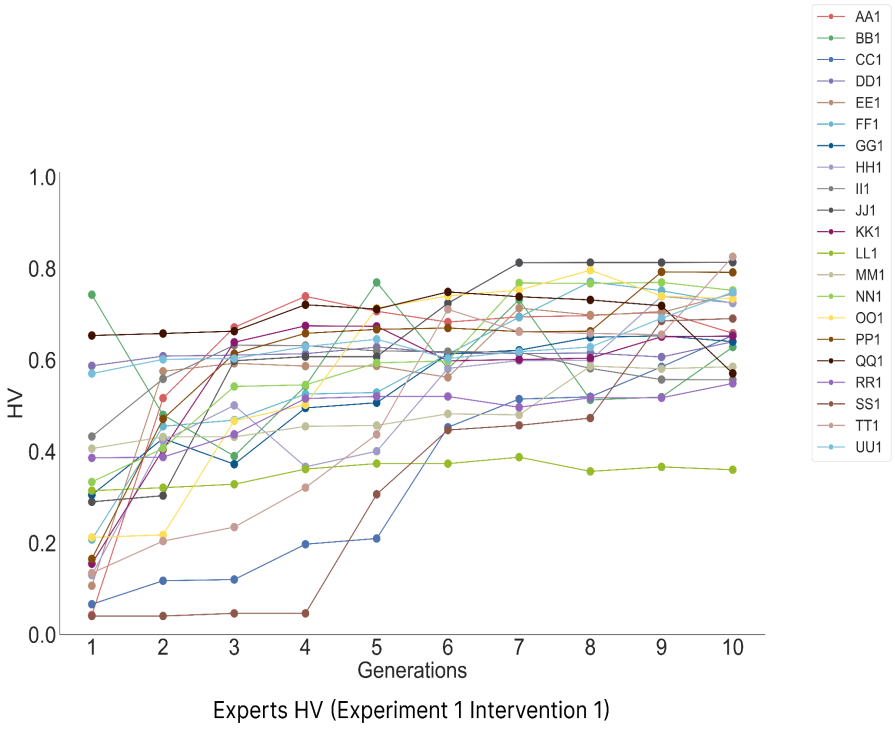


Figure 4.66. Experts (Experiment 1 Intervention 1).

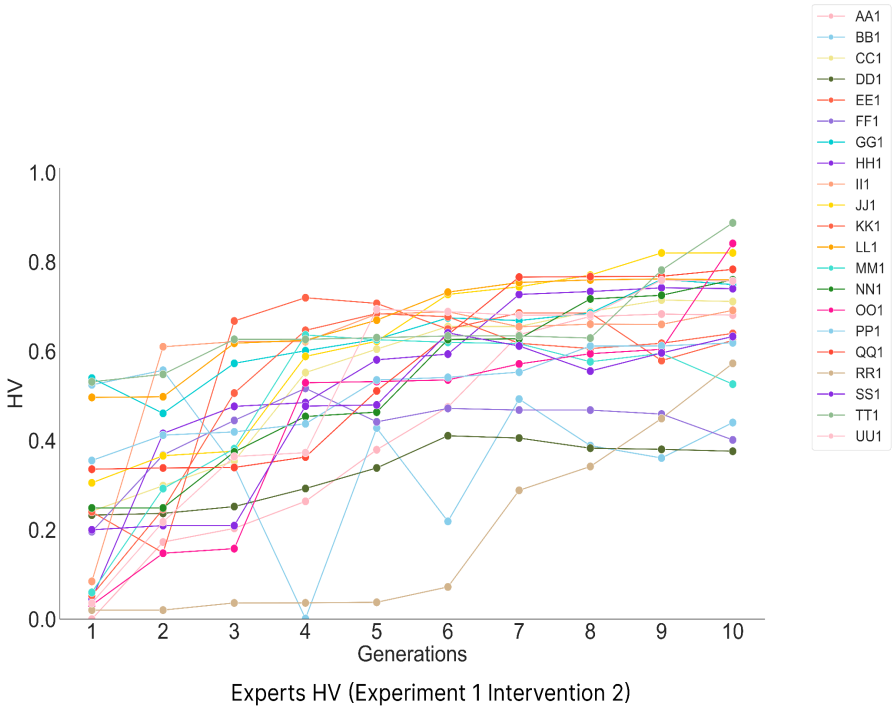


Figure 4.67. Experts (Experiment 1 Intervention 2).

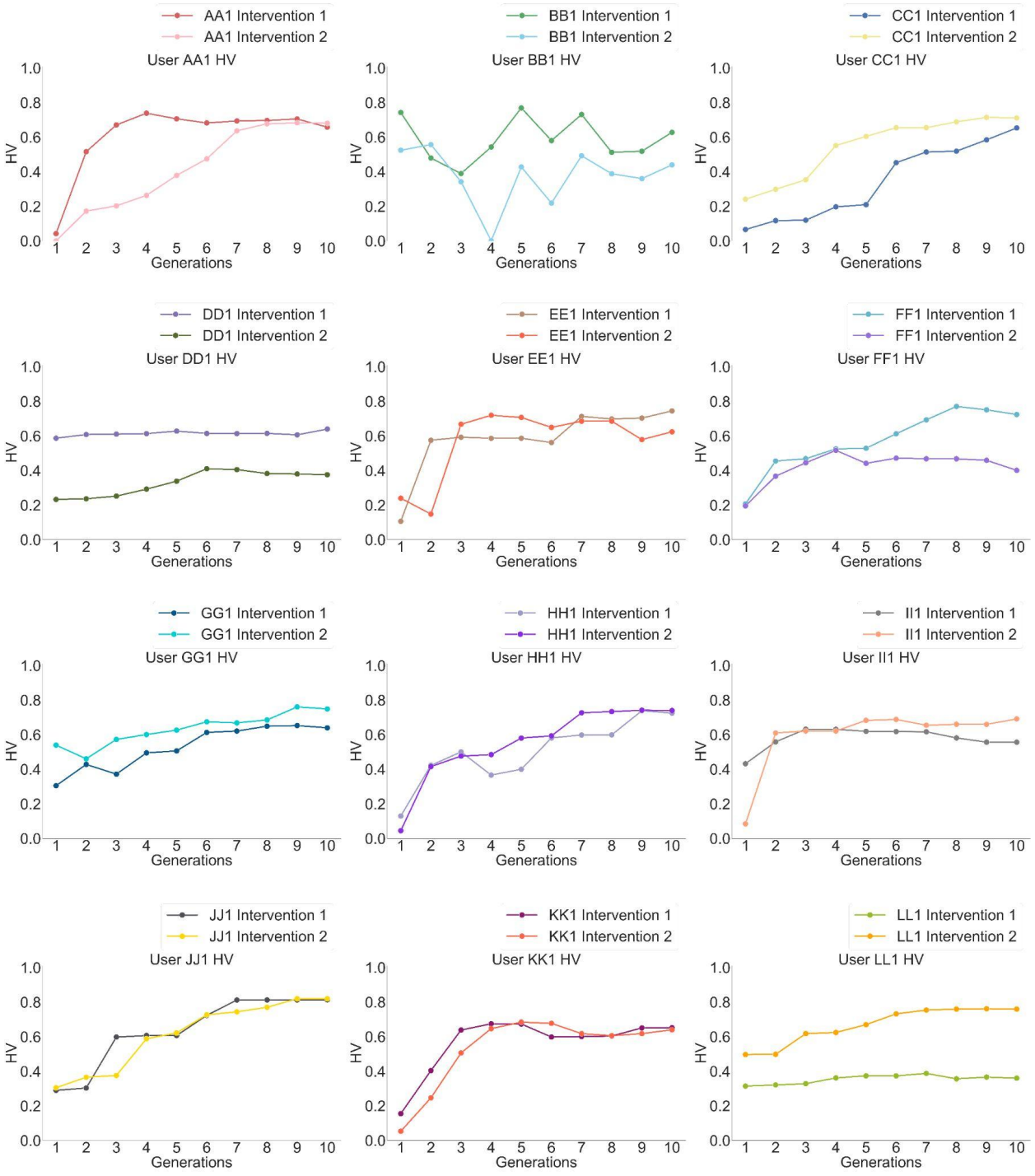


Figure 4.68. Experts AA1 to LL1 (Experiment 1 Intervention 1).

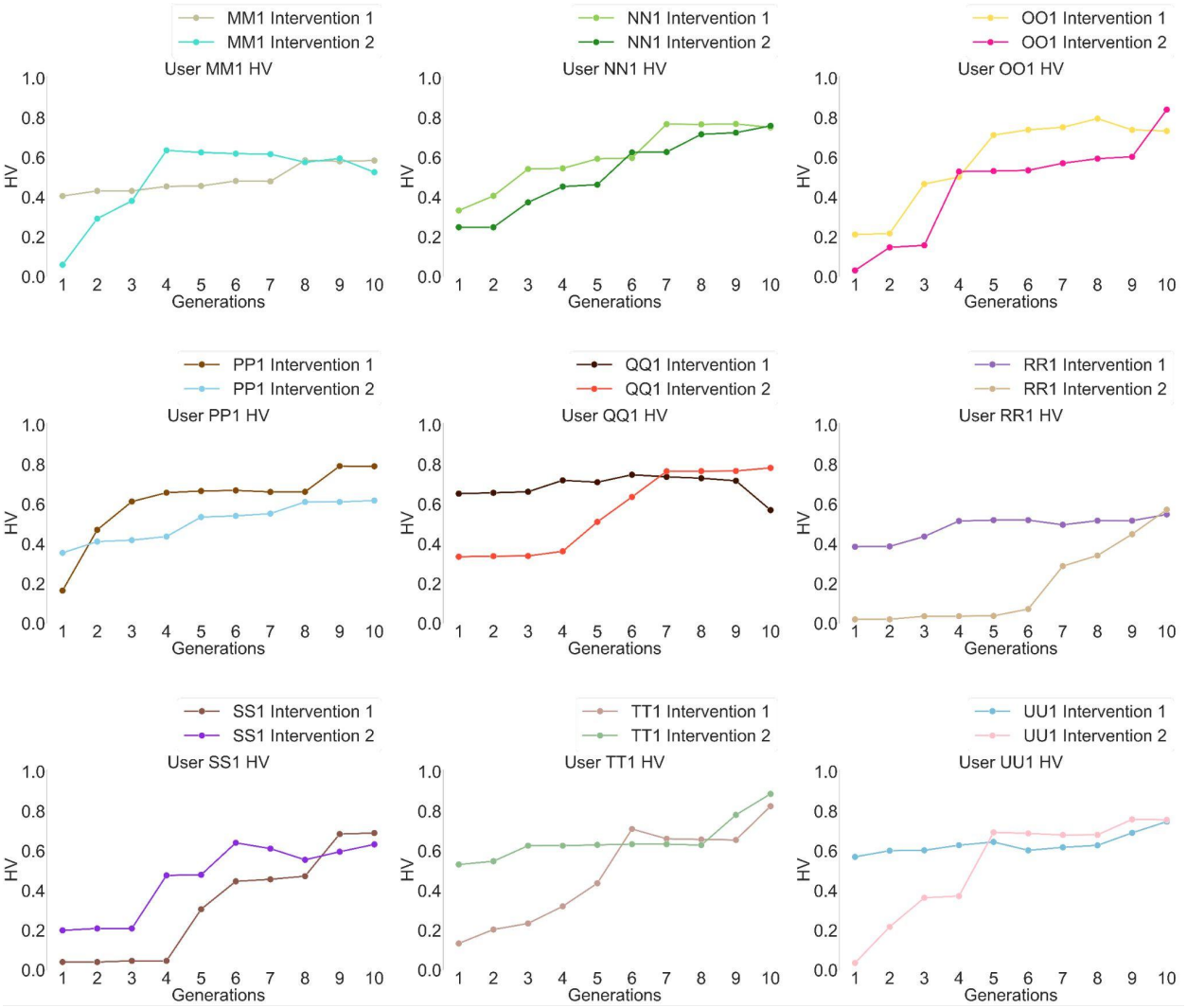


Figure 4.69. Experts MM1 to UU1 (Experiment 1 Intervention 1).

2. Impacts of Providing Performance Metrics Among Non-Experts

To assess the HV indicator values among non-experts with and without providing performance metrics (Experiment 1 Intervention 1 and 2), the HV values of the normalized final solutions for both interventions are once again analyzed. The HV values for non-experts in Experiment 1 Intervention 1 and 2 are normally distributed. The p-value calculated using the t-test resulted in 0.8, demonstrating that, again, there is no significant difference between the performance of non-dominated solutions among non-experts with and without providing performance metrics to users. This result demonstrates the consistency of non-experts' preference impact on the optimization results, and how final solutions tend to perform similarly when users are provided with performance metrics compared to when they are not provided with the metrics.

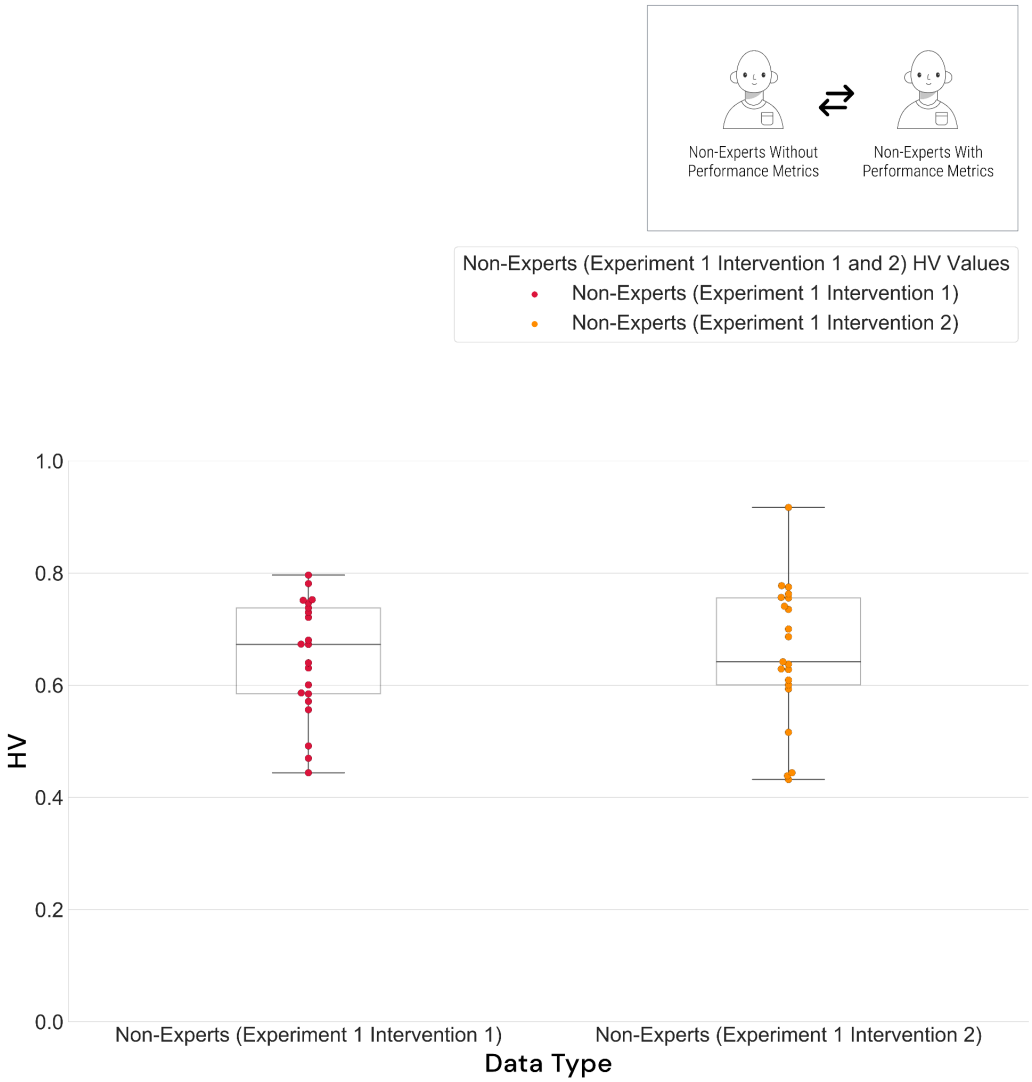


Figure 4.70. Non-experts' final HV Values Experiment 1 Intervention 1 and 2.

This result can be further investigated by visualizing all non-experts HV values changes over ten generations for all experiments. Figures 4.71 and 4.72 in the next pages show the changes in HV values over time for all non-experts. As an example, for non-expert 2I2, both interventions had similar behavior in HV values over time. For some non-experts, intervention 2 resulted in lower HV values over time, such as non-expert 2G2. Some non-experts' HV values in intervention 2 improved over time compared to their Intervention 1 results, such as non-expert 2D2.

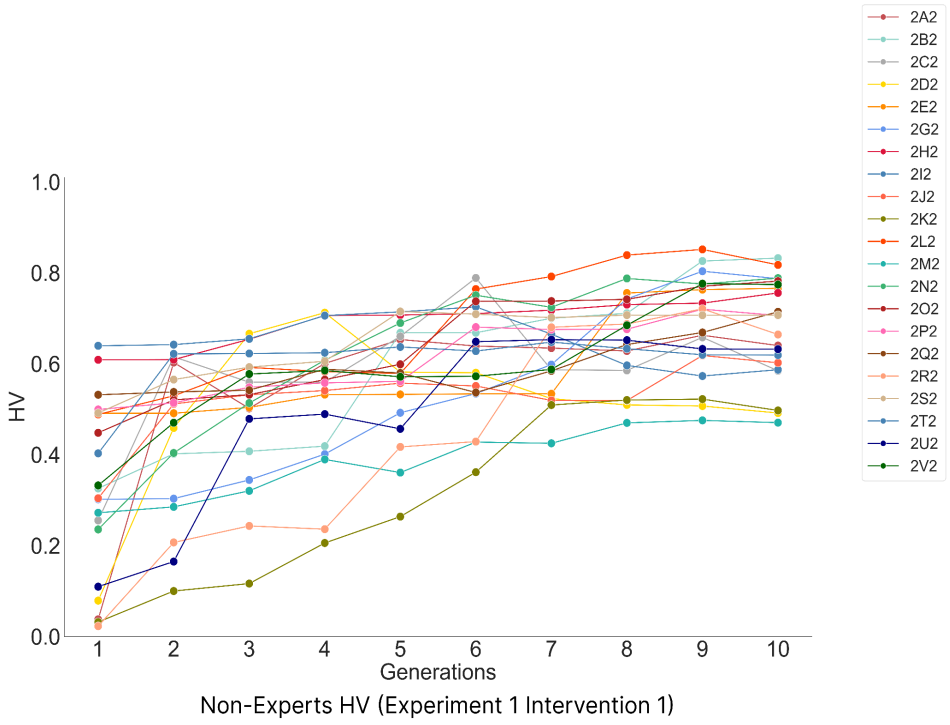


Figure 4.71. Non-Experts (Experiment 1 Intervention 1).

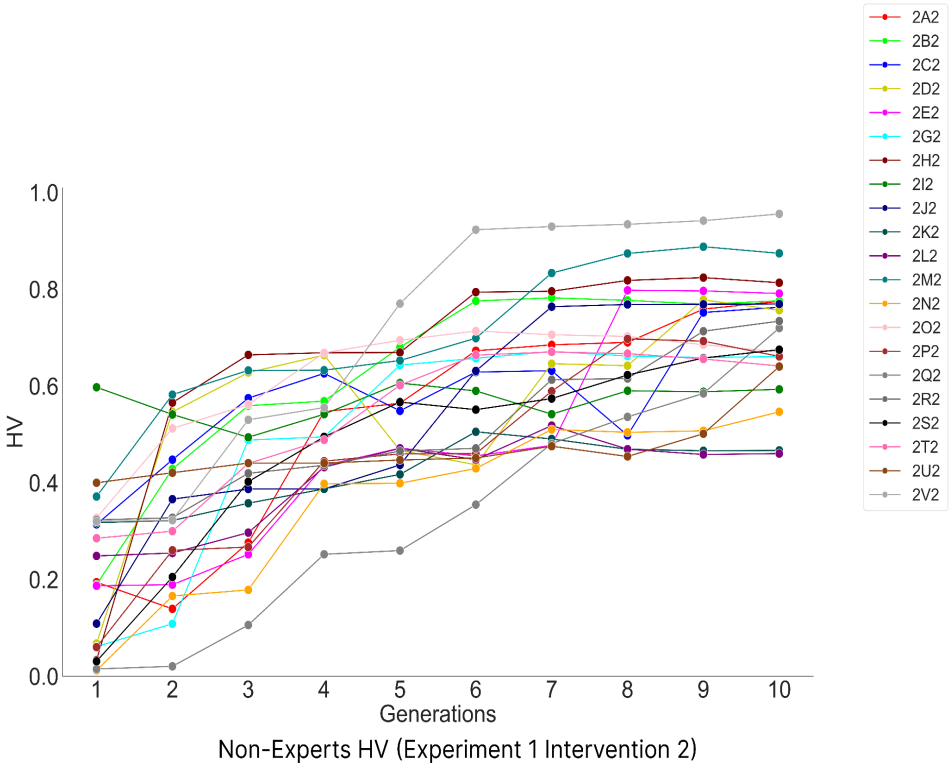


Figure 4.72. Non-Experts (Experiment 1 Intervention 2).

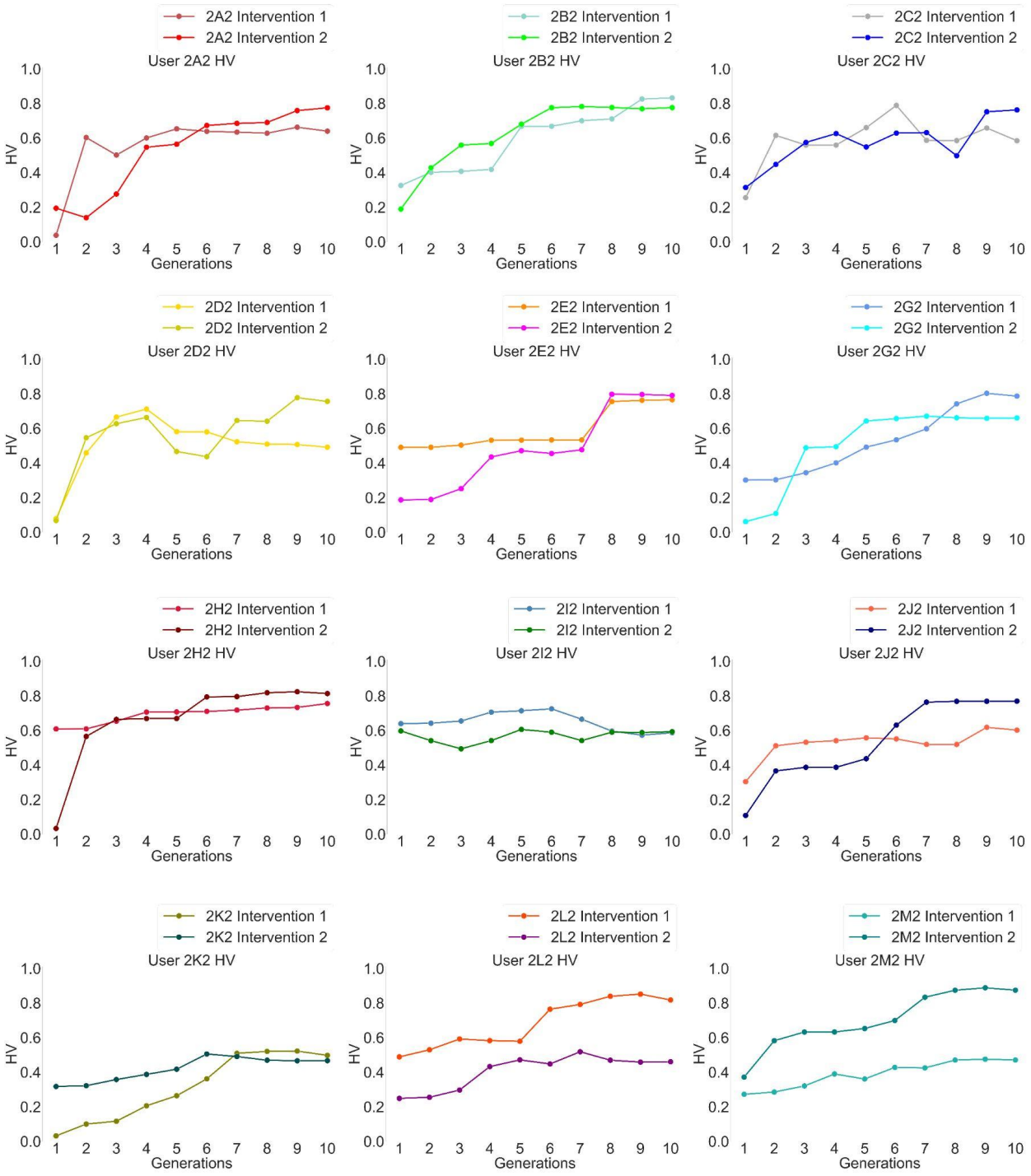


Figure 4.73. Non-Experts 2A2 to 2M2 (Experiment 1 Intervention 1).

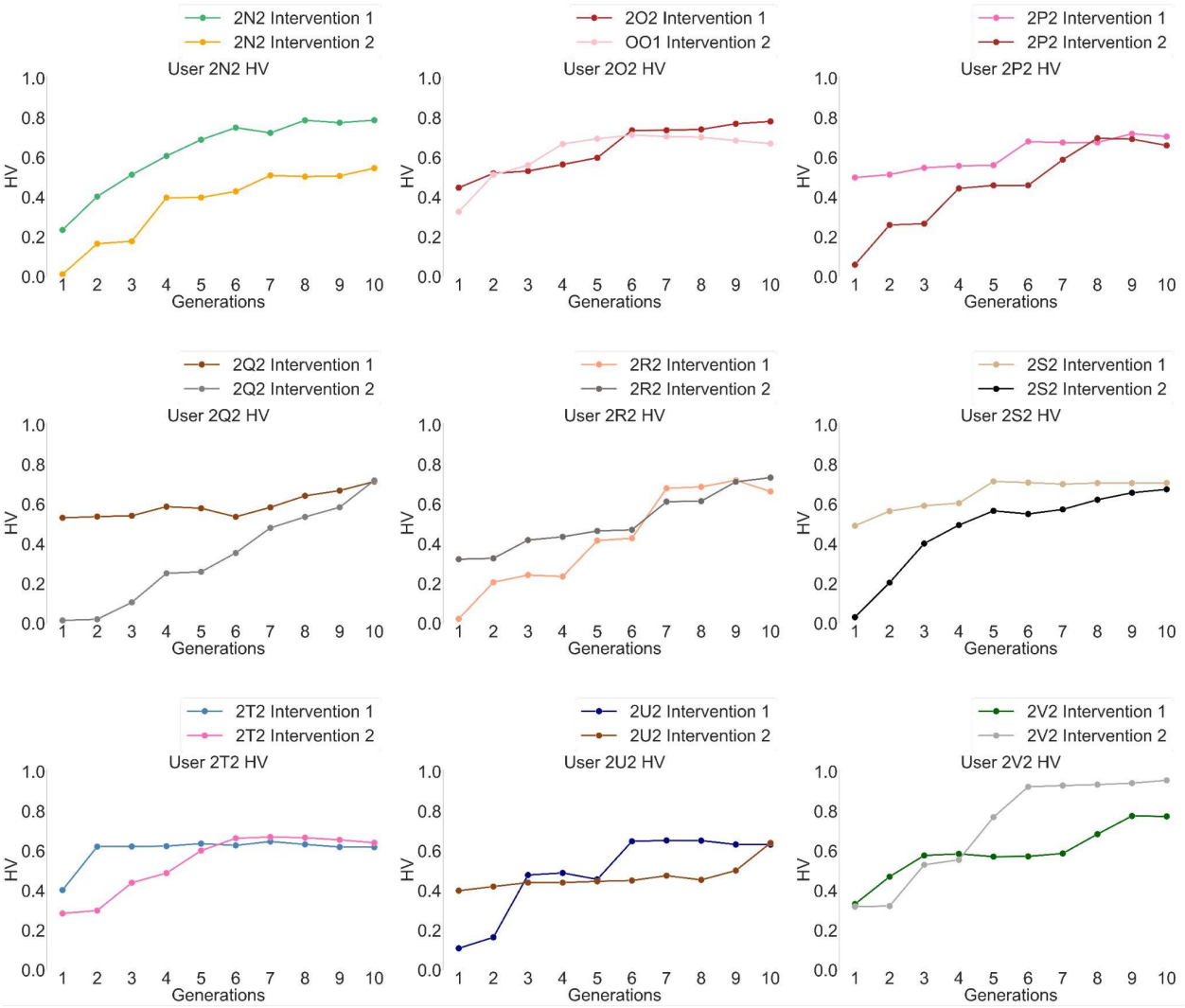


Figure 4.74. Non-Experts 2N2 to 2V2 (Experiment 1 Intervention 1).

3. Comparing Experts and Non-Experts (Experiment 1 Intervention 1)

There is no significant difference in the Hypervolume values among experts and non-experts when the performance metrics are not provided to users (Experiment 1 Intervention 1). As previously discussed, both these datasets are normally distributed. The calculated p-value based on a t-test is 0.8, higher than the 0.05 alpha value. This result demonstrates the similarity of final non-dominated solutions' performance among experts and non-experts, which can help advocate for including a combination of experts and non-experts in the early-stage generative design process. This result demonstrates that experts' preferences without access to performance metrics results in similar performance to non-experts' results. Based on the results of the NSGA-II Alone, the algorithm tends to achieve high performance in the final solutions and high performing solutions are building types with taller massing options (tower typology). Based on user interviews, the tower typology was less preferred among both experts and non-experts. Both

groups' similar preferences on massing options, specifically on high performing options, could have similarly impacted the performance of the final solutions.

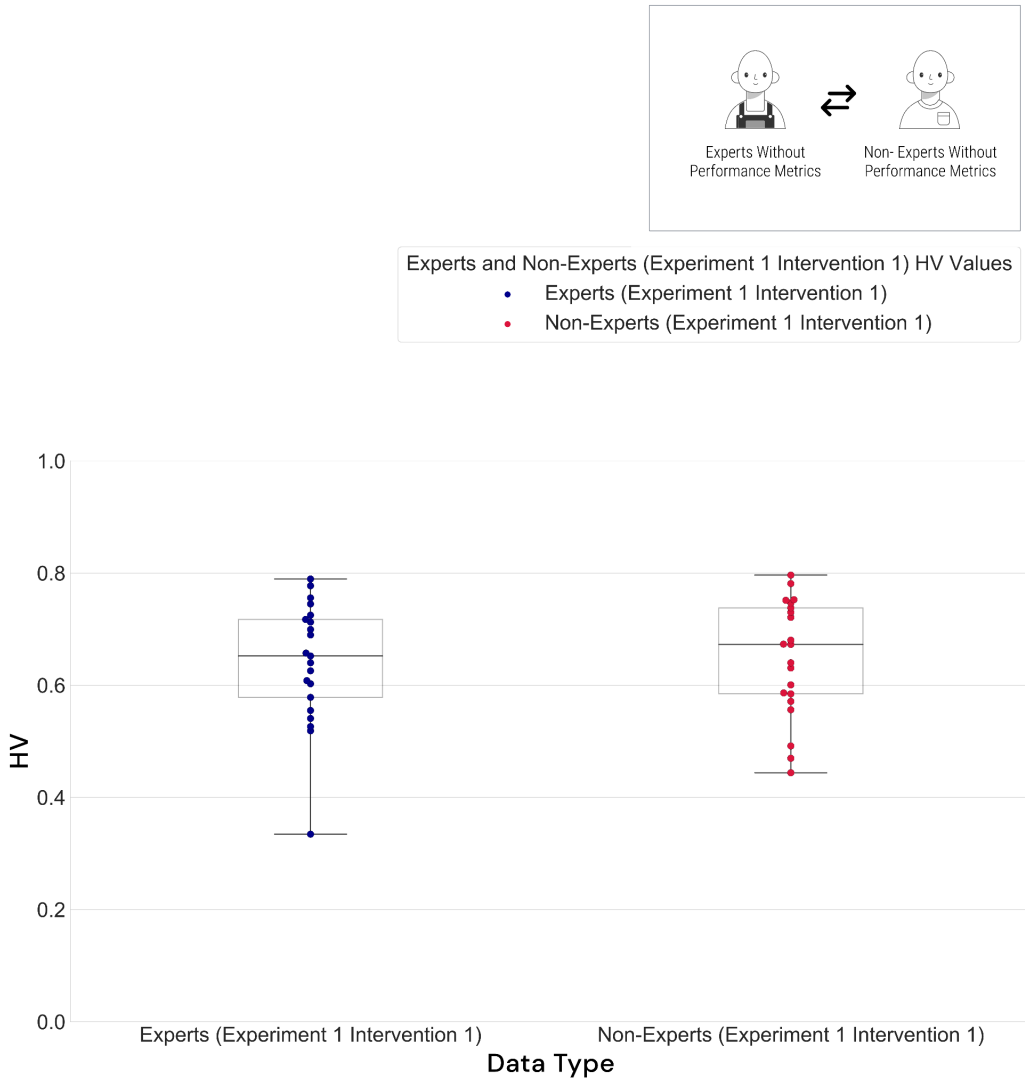


Figure 4.75. Experts and Non-Experts Final HV Values Experiment 1 Intervention 1.

4. Comparing Experts and Non-Experts (Experiment 1 Intervention 2)

When performance metrics are provided to both user groups, there is no significant difference between experts' and non-experts' HV values. Both data sets are normally distributed, and the calculated p-value is 0.9. Similar to previous results, this result demonstrates the similarity of final non-dominated solutions' performance among experts and non-experts. As discussed earlier, the high performing solutions without the inclusion of users are building types with taller massing options (tower typology), however this building type was less preferable among users. However, the majority of users discussed the impacts of providing performance metrics on their acceptance of the tower typology based on its high performance in other objective functions. Therefore, both

experts and non-experts ratings can result in similar performance when provided with performance metrics.

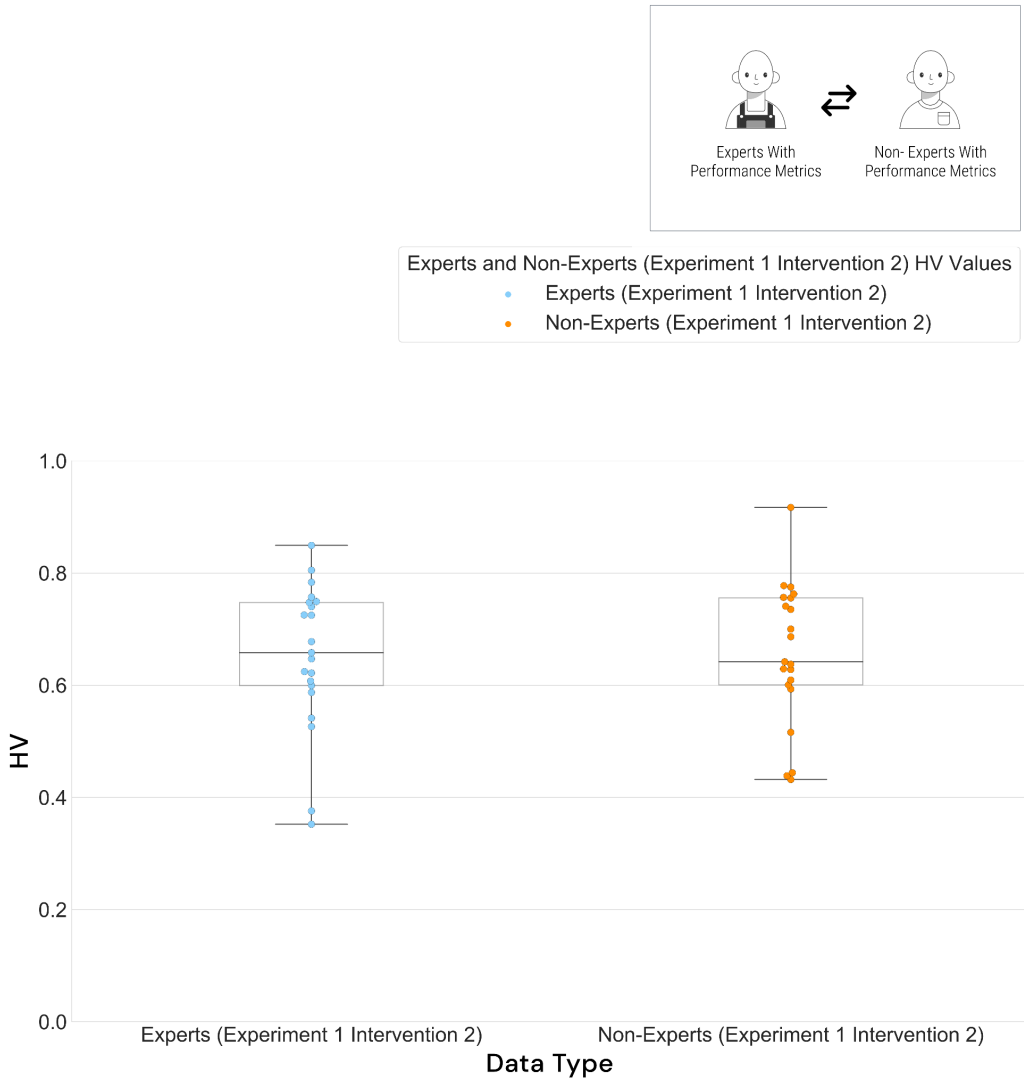


Figure 4.76. Experts and Non-Experts Final HV Values Experiment 1 Intervention 2.

5. Comparison of Experts and NSGA-II Alone (Experiments 1 and 2)

Next, the HV values between experts and the algorithm were compared for both interventions (Experiment 1 and 2 Intervention 1 and 2). Prior to conducting Experiment 2, it was expected for the algorithm results without the user's rating to have a higher performance (HV value) compared to Experiment 1 results. The HV values for Experiment 2 Interventions 1 and 2 are both normally distributed. There is a significant difference in HV values between experts and NSGA-II Alone Intervention 1 results (p-value=2e-7). There is also a significant difference in HV values between experts and NSGA-II Alone Intervention 2 results (p-value=5e-07). These results show that the

algorithm outperforms the experts with and without performance metrics, which was expected before conducting the experiments, as the addition of subjective ratings can depress convergence.

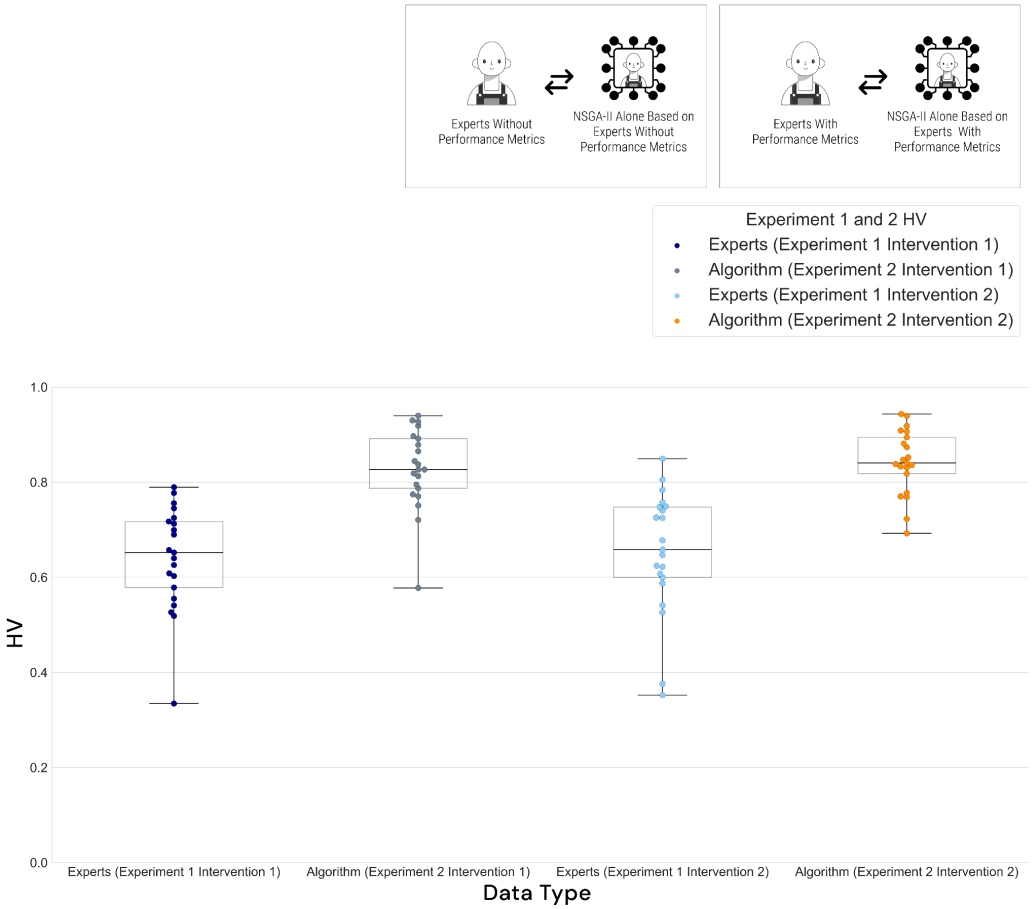


Figure 4.77. Experts and Algorithm Final HV Values Experiments 1 and 2.

The algorithm tends to achieve better performance in the final solutions but results in building types with taller massing options (tower typology) that were less preferred among experts. Despite the higher performance of the algorithm in Experiment 2, these results show the importance of including users in the generative design process. The user preference on massings can be in conflict with other performance metrics that are taken into account by the algorithm, which can lead to more desirable massing options.

6. Comparison of Non-Experts and NSGA-II Alone (Experiments 1 and 2)

The HV values between non-experts and the algorithm are compared for both interventions (Experiment 1 and 2 Intervention 1 and 2). The HV values for Experiment 2 Intervention 1 and 2 are both normally distributed. There is a significant difference in HV values between non-experts and NSGA-II Alone Intervention 1 results (p-value=9e-11). There is also a significant difference in HV values between non-experts and the NSGA-II Alone based on Intervention 2 results (p-value=9e-07). Similar to the experts, these results show that the algorithm alone outperforms the

non-experts both with performance metrics and without performance metrics, which, like for the experts, was expected prior to conducting the experiments.

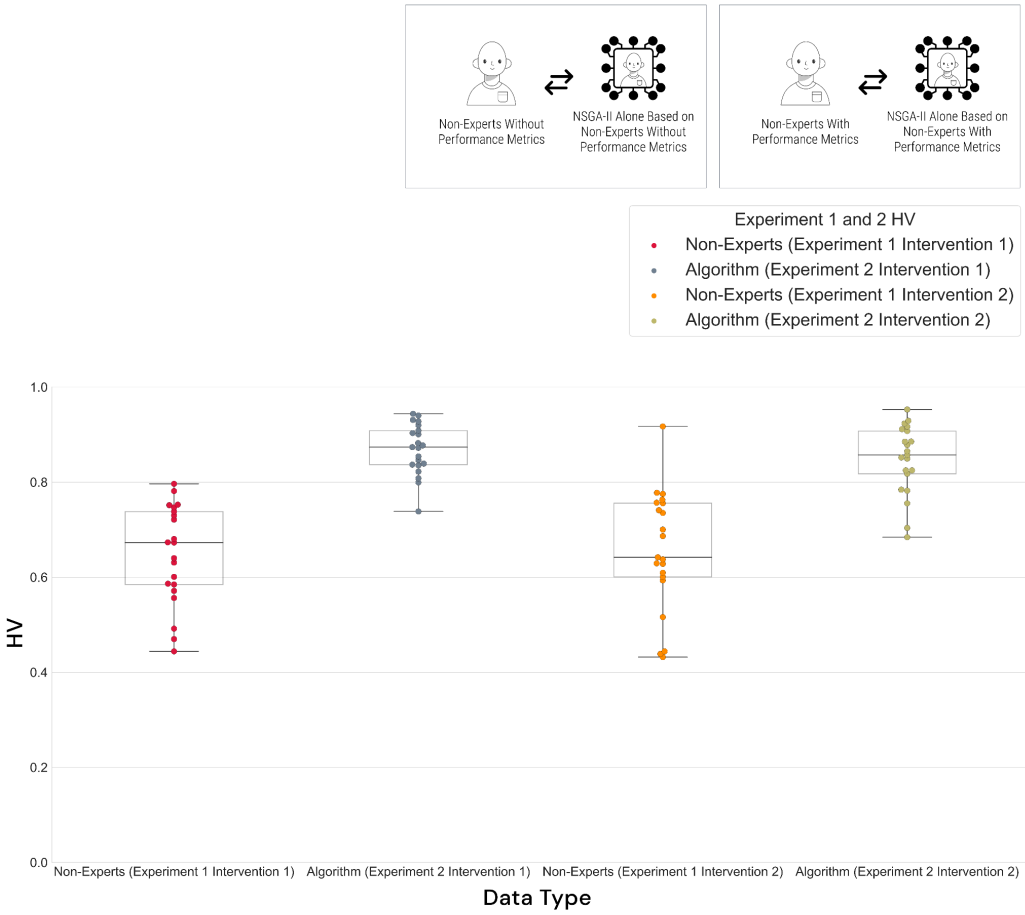


Figure 4.78. Non-Experts and Algorithm Final HV Values Experiments 1 and 2.

Building Height Analysis

A strategy that we employed to assess the variation in generated massing options by InsightXR was to analyze the final solutions for all users based on building heights. The building height is calculated by averaging the height of all four roof points of the roof surface. In this section, the height distribution (average of building heights calculated via the room rim vertices) among experts and non-experts, with and without providing performance metrics to users, are analyzed. In addition, all algorithms executed based on experts and non-experts, with and without performance metrics provided to users, are discussed. There is a significant difference in building heights for experts, with and without performance metrics. For non-experts, there is a significant difference in building heights with and without performance metrics provided to users.

1. Impacts of Providing Performance Metrics (Experiment 1 Intervention 1 and 2)

The experts' and non-experts' solutions' (building) heights are not normally distributed in Experiment 1, Intervention 1 and Intervention 2. Among experts, the difference in building heights between interventions 1 and 2 is statistically significant (p-value=0.02). The statistical power in this study is 0.6, showing a moderate likelihood of finding a true effect. The effect size is -0.1, which indicates a small magnitude for the difference between two datasets. Overall, the results of this study show a significant difference between the two datasets with a moderate statistical power, which shows a reasonable chance of detecting a difference between the two datasets. In addition, among non-experts, the difference in building heights between Interventions 1 and 2 is statistically significant (p-value=2e-07). For non-experts, the statistical power is 0.99, which indicates a very high chance of correctly finding a true effect. The effect size in this study is -0.29, showing a moderate value for the difference between the two datasets. Overall, there was a significant difference in the results of non-experts, with a moderate effect size and a very high statistical power.

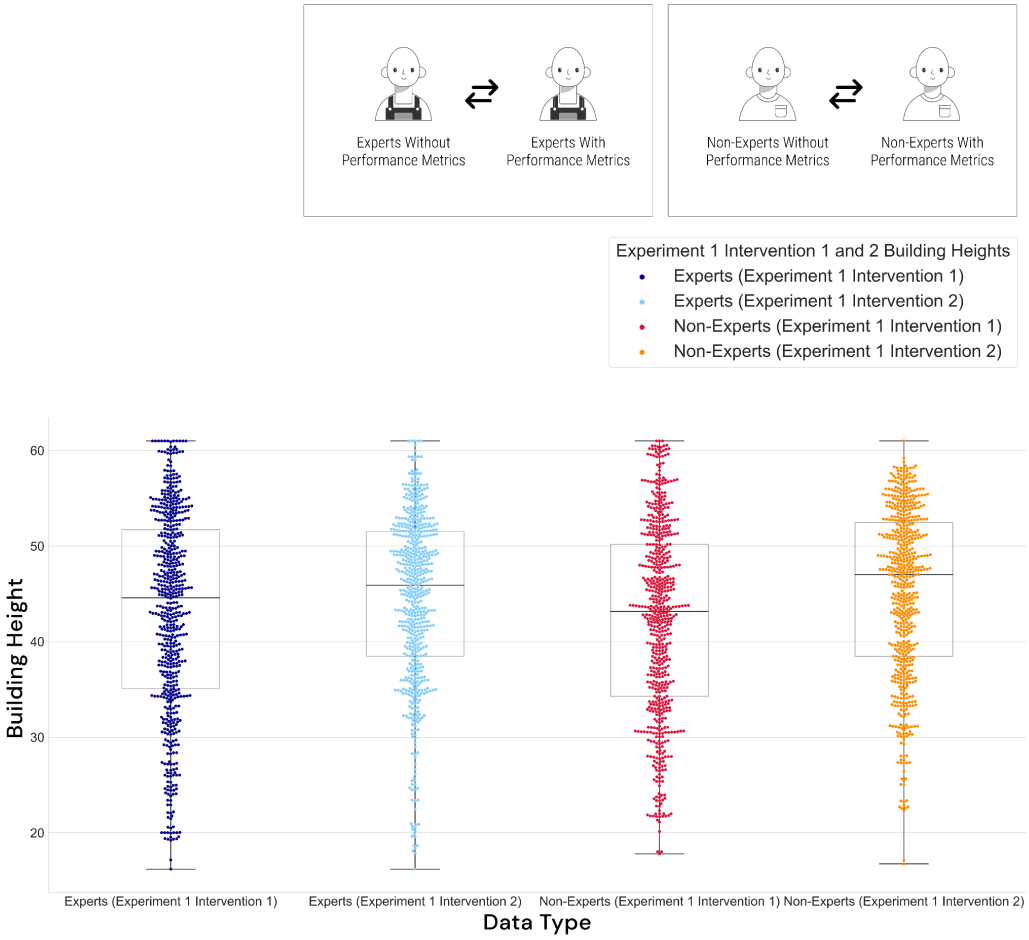


Figure 4.79. Experts and Non-Experts Experiment 1 Building Heights.

The figure above shows the building height distribution for the expert and non-expert groups in Experiment 1, Interventions 1 and 2. These results suggest that for both experts and non-experts, the final solutions tend to be taller when users are provided with performance metrics. As discussed previously, the NSGA-II Alone tended to create building types with taller massing options (tower typology) that were less preferred among experts. The increase in height in Intervention 2 shows the impact of providing performance metrics to users, where users show more flexibility with their less preferable massing options based on high performance in other objective functions.

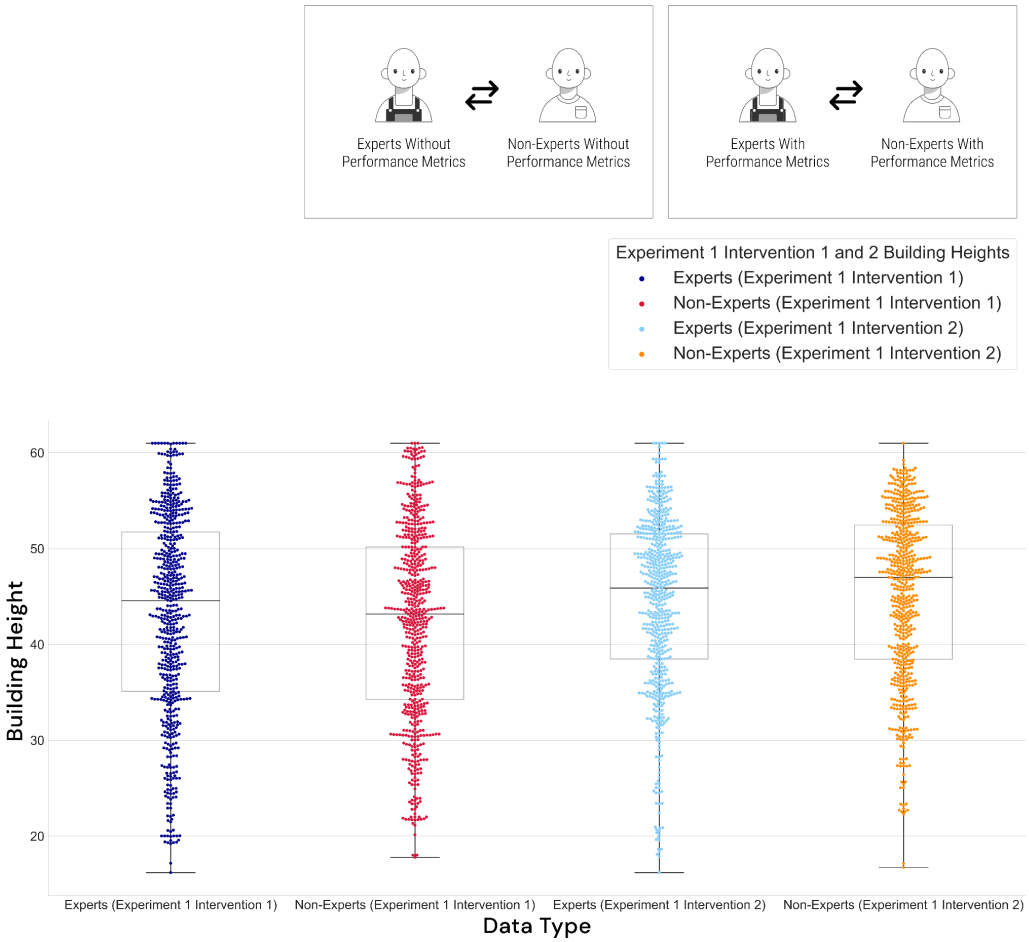


Figure 4.80. Experts Versus Non-Experts Experiment 1 Building Heights.

In Intervention 1, between experts and non-experts, the difference in building heights is not statistically significant (p-value=0.06). In Intervention 2, between experts and non-experts, the difference in building heights is not statistically significant (p-value=0.07). The figure shows the building height distribution for the expert and non-expert groups in Experiment 1, Interventions 1 and 2. These results suggest that the final solutions for experts and non-experts tend to be similar when users are provided with performance metrics and when they are not provided with performance metrics. In the user interviews, both user groups described the tower typology as their least preferred building type, and the result of this test shows how both user groups resulted in

similar building heights when user rating was purely based on user aesthetic preferences. However, when presented with performance metrics, both user groups resulted in similar building heights.

7. Comparison of Experts and NSGA-II Alone (Experiments 1 and 2, Intervention 1 and 2)

The algorithm's solutions' (building) heights based on experts' initial populations are not normally distributed in Experiment 2, Intervention 1 and Intervention 2. Between experts and NSGA-II Alone based on experts' initial populations, the difference in building heights between Experiment 1 and 2 Intervention 1 is statistically significant (p-value=4e-27). The statistical power in this study is 1, showing a very high likelihood of finding a true effect. The effect size is -0.7, which indicates a large magnitude for the difference between two datasets. Overall, the results of this study show a significant difference between the two datasets with a very high statistical power, which shows a high chance of detecting a difference between the two datasets.

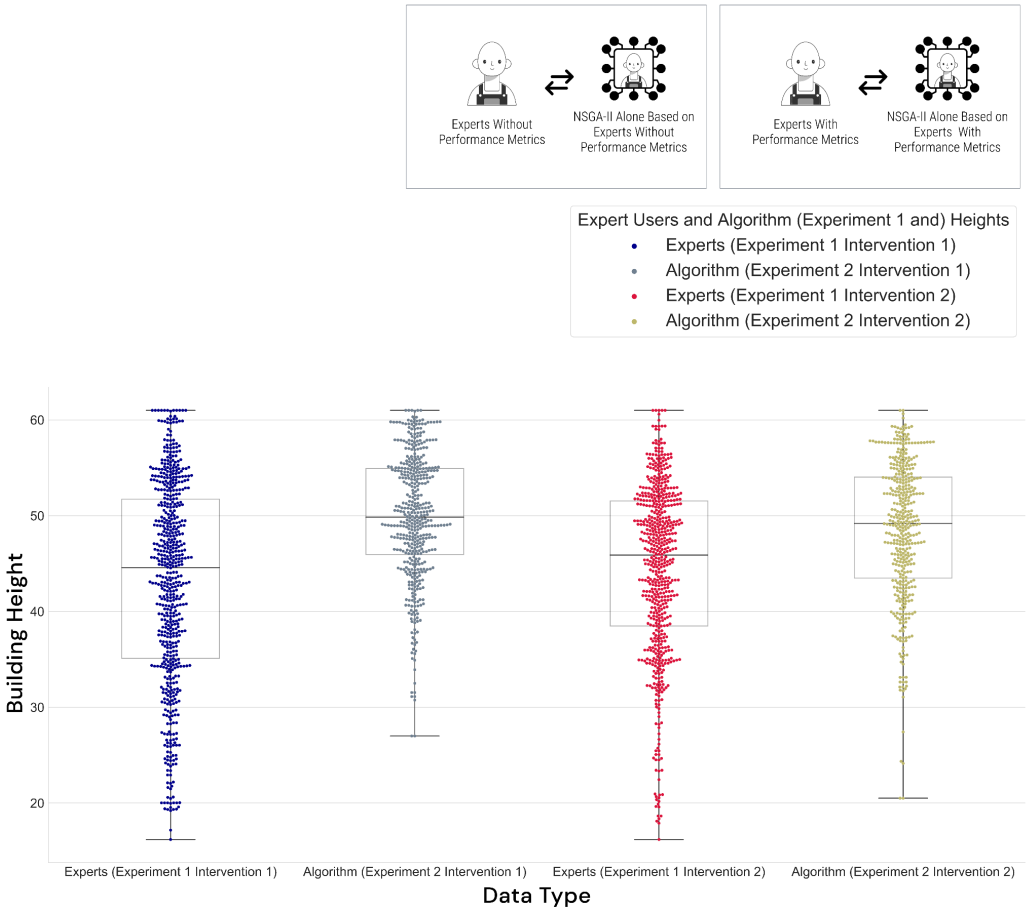


Figure 4.81. Experts Versus Algorithm Experiment 1 and 2 Building Heights.

In addition, between experts and NSGA-II Alone based on experts' initial populations, the difference in building heights between Experiment 1 and 2 Intervention 2 is statistically significant (p-value=4e-13). The figure above shows the building height distribution for the expert and

algorithm groups in Experiments 1 and 2, Interventions 1 and 2. For Intervention 2 comparisons, the statistical power is 1, showing a very high likelihood of finding a true effect. The effect size is -0.4, which indicates a moderate magnitude for the difference between two datasets. Overall, the results of this study show a significant difference between the two datasets with a very high statistical power, which shows a high chance of detecting a difference between the two datasets. These results suggest that for the algorithm performance without expert input, the final solutions tend to be taller compared to the expert's counterparts.

8. Comparison of Non-experts and NSGA-II Alone (Experiments 1 and 2, Intervention 1 and 2)

The algorithm's solutions' (building) heights (based on non-experts' initial populations Experiment 1, Intervention 1 and 2) are not normally distributed in Experiment 2, Intervention 1 and 2. Between non-experts and NSGA-II Alone based on non-experts, the difference in building heights between Experiment 1 and 2 Intervention 1 is statistically significant (p-value=1e-23). The statistical power in this study is 1, showing a very high chance of finding a true effect. The effect size is -0.6, which indicates a large magnitude for the difference between two datasets.

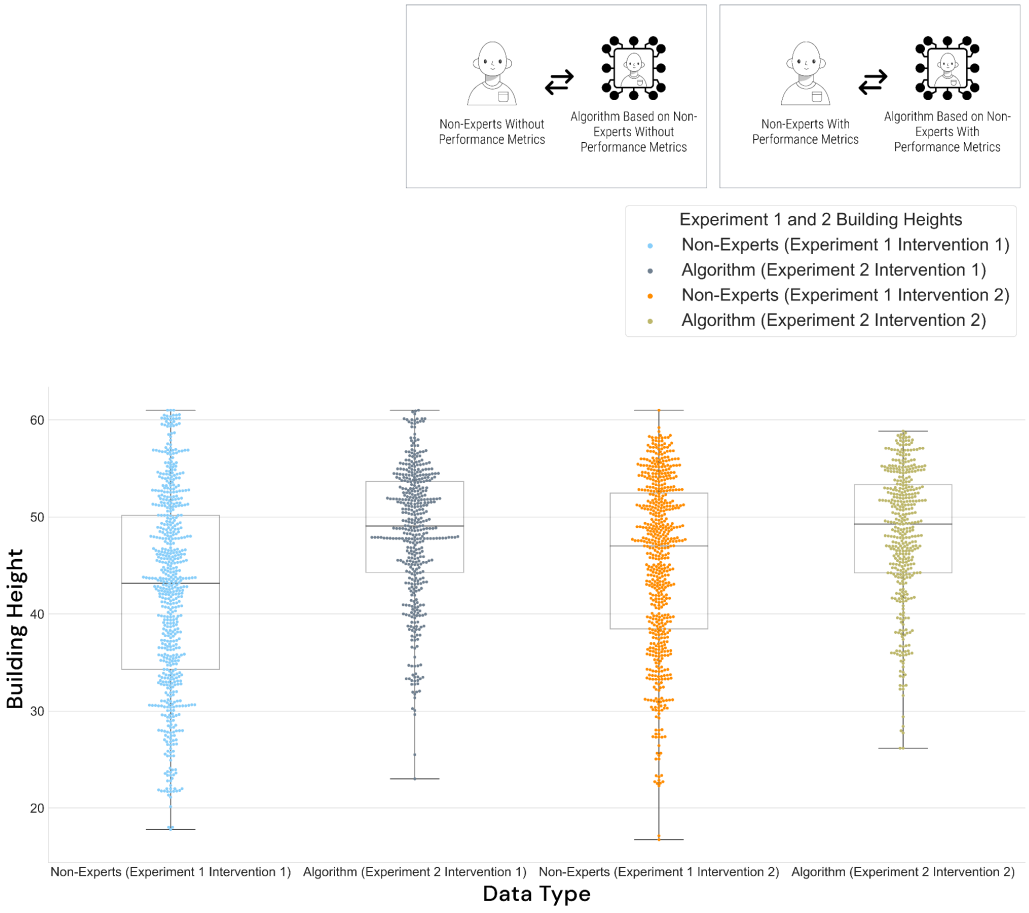


Figure 4.82. Non-Experts Versus Algorithm Experiments 1 and 2 Building Heights.

Overall, the results of this study show a significant difference between the two datasets with a very high power, which shows a high chance of detecting a difference between the two datasets.

In addition, between non-experts and NSGA-II Alone based on non-experts, the difference in building heights between Experiment 1 and 2 Intervention 2 is statistically significant (p-value=2e-07). For Intervention 2, the statistical power is 1, showing a very high likelihood of finding a true effect. The effect size is -0.3, which indicates a moderate magnitude for the difference between two datasets. Overall, the results of this study show a significant difference between the two datasets with a very high statistical power, which shows a high chance of detecting a difference between the two datasets. Figure 4.82 shows the building height distribution for the non-expert and algorithm groups in Experiments 1 and 2, Interventions 1 and 2. These results suggest that, like when comparing to the experts, the NSGA-II results based on non-experts have final solutions that tend to be taller compared to their users' counterparts.

Objective Functions Analysis: Impacts of Providing Performance Metrics

In this section the objective functions (performance metrics) values for Experiment 1 and 2 will be analyzed. The performance values will be discussed in three different categories: AV Values, NPZ Values, and UOS values.

AV Values (Experiment 1 Intervention 1 and 2)

1. **Experts Comparison (Experiment 1 Interventions 1 and 2)**

There is a significant difference in the AV values among experts when the performance metrics are not provided compared to when provided to users (Experiment 1 Interventions 1 and 2). Both these datasets are not normally distributed. The calculated p-value is 0.01, which is lower than the 0.05 alpha value.

This result shows that the final solutions in Intervention 2 had overall reduced in AV values. This result demonstrates the value in providing a strong rubric for both experts and non-experts, and the potentially valuable feedback it can generate.

While experts ranked AV as the least important performance metric in the user interviews, the AV values were lower in the final solutions when users were presented with performance metrics. Additionally, this can demonstrate the value in providing a strong rubric for non-experts, and the potentially valuable feedback it can generate. The calculated statistical power in this study is 0.7, which shows a high chance of detecting a true effect. The effect size, however, is low, with a value of 0.1. Overall, the statistical analysis results show a significant difference between the two data sets, with a small effect size while the achieved power is high.

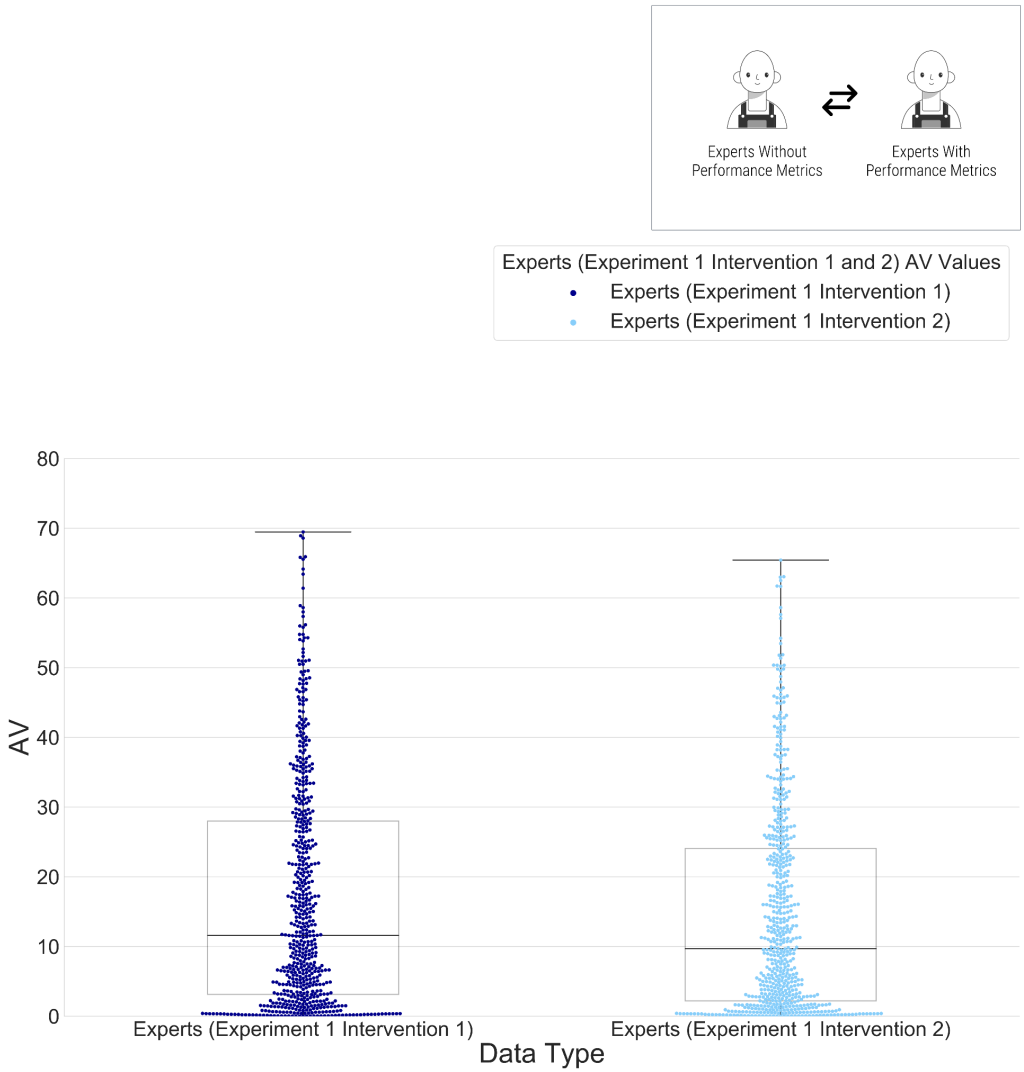


Figure 4.83. Comparison of Experts AV Values.

2. Non-Experts Comparison (Experiment 1 Intervention 1 and 2)

There is no significant difference in the AV values among non-experts when the performance metrics are not provided compared to when they are provided to users (Experiment 1 Intervention 1 and 2). Both these datasets are not normally distributed. The calculated p-value is 0.1, which is higher than the 0.05 alpha value. This result suggests that the AV values were increased in intervention 2, which is not a desired result.

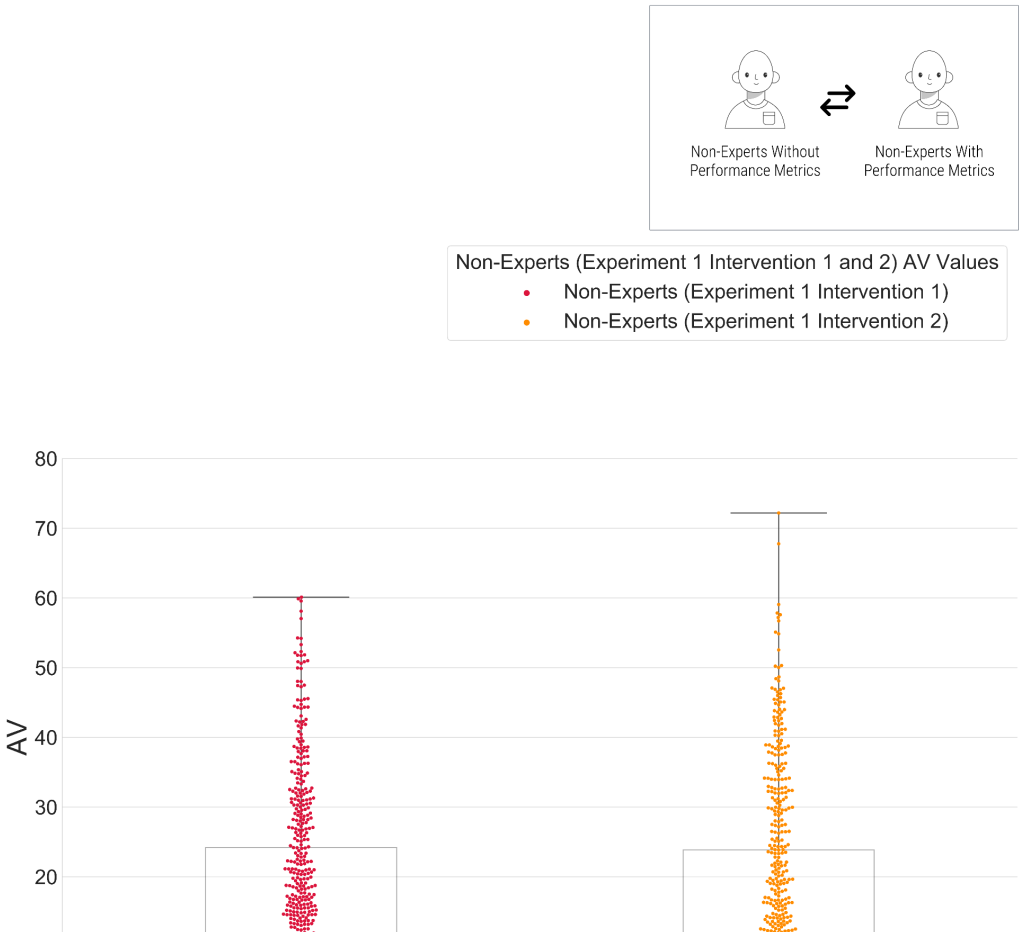


Figure 4.84. Comparison of Non-Experts AV values.

3. Experts and Non-Experts Comparison (Experiment 1 Intervention 1)

There is a significant difference in the AV values between experts and non-experts when the performance metrics are not provided to users (Experiment 1 Intervention 1). The calculated p-value is 0.03, which is lower than the 0.05 alpha value. This result demonstrates the difference of final non-dominated solutions AV values among experts and non-experts, where experts' experiments resulted in higher AV values without access to performance metrics. The statistical power in this study is 0.6, which shows a moderate to high chance of detecting a true effect. The effect size is 0.1, which is a low value. Overall, the results show a significant difference between the two data sets, with a small effect size while the achieved power is high.

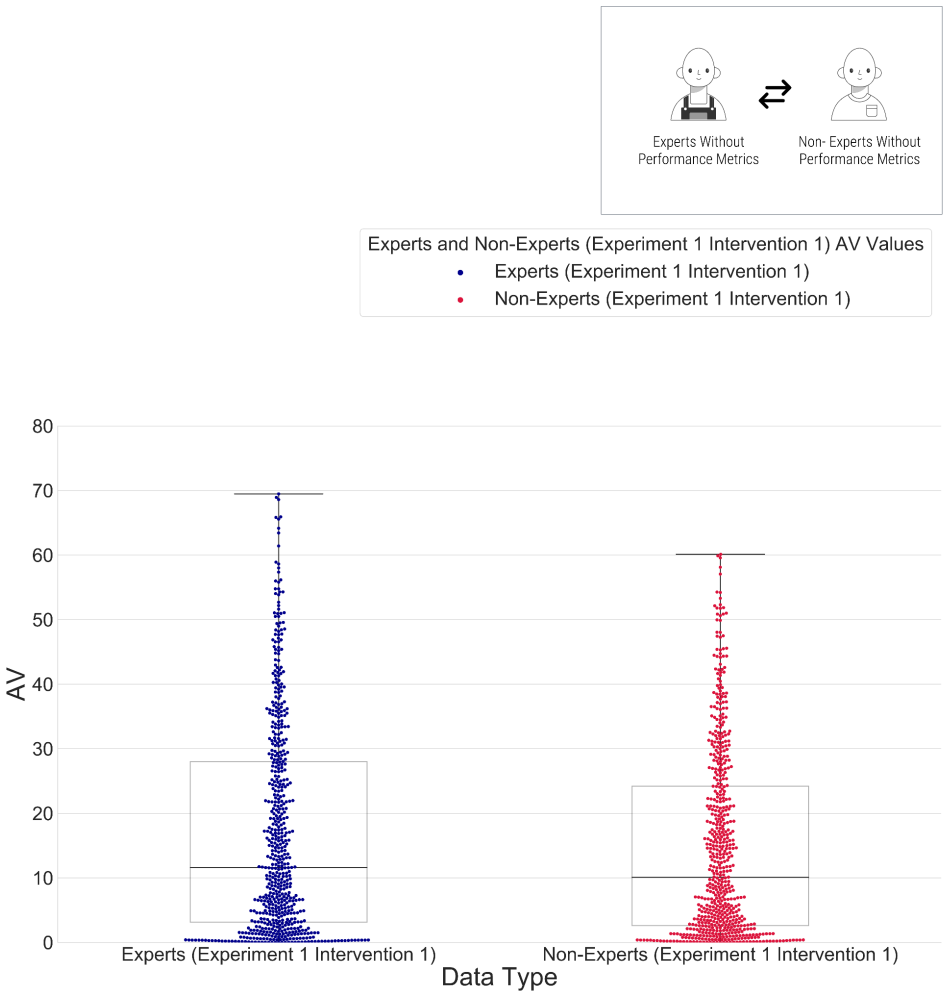


Figure 4.85. Comparison of Experts and Non-Experts AV Values (Intervention 1).

4. Experts and Non-Experts Comparison (Experiment 1 Intervention 2)

Both experts and non-experts ranked AV as the least important performance metric, and while for non-experts there was no significant difference between Intervention 1 and 2, experts' solutions improved in AV values in Intervention 2. However, there is no significant difference in the AV values between experts and non-experts when the performance metrics are provided to users (Experiment 1 Intervention 2). Both these datasets are not normally distributed. The calculated p-value is 0.3, which is higher than the 0.05 alpha value. This result demonstrates the similarity of final non-dominated solutions AV values among experts and non-experts.

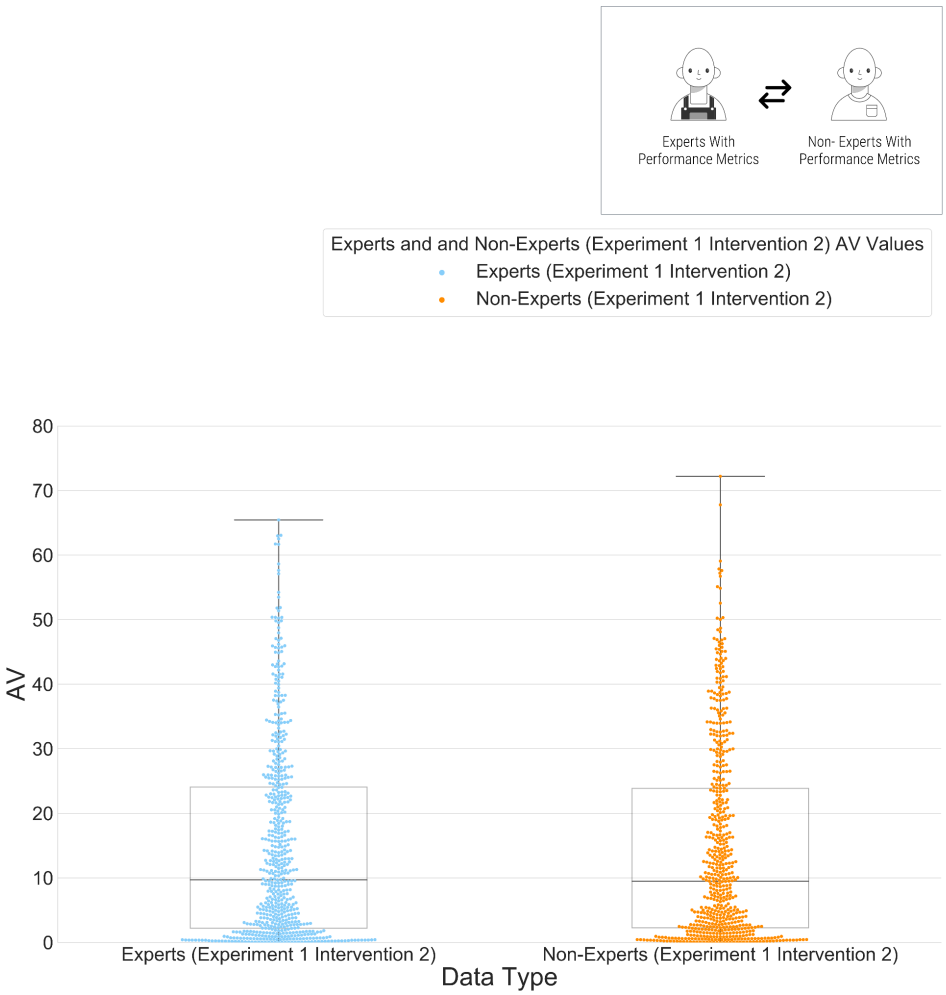


Figure 4.86. Comparison of Experts and Non-Experts AV Values (Intervention 2).

NPZ Values (Experiment 1 Intervention 1 and 2)

5. Experts Comparison (Experiment 1 Intervention 1 and 2)

In the user interviews conducted with experts, they ranked NPZ as the most important performance metric. The results of statistical analysis show that there is a significant difference in the NPZ values among experts when the performance metrics are not provided versus when they are provided to users (Experiment 1 Intervention 1 and 2). This result demonstrates the value in providing a strong rubric for both experts and non-experts, and the potentially valuable feedback it can generate.

This result demonstrates the improvement of final non-dominated solutions' NPZ values among experts with and without performance metrics provided to users, where experts with performance metrics provided, resulted in lower NPZ values. Both these datasets are not normally distributed. The calculated p-value is 0.001, which is lower than the 0.05 alpha value. The statistical power is

0.9, which shows a high likelihood of finding meaningful differences between the two datasets. The effect size is 0.2, showing a small to moderate value. Overall, the results of this study show that there is a statistically significant difference between the two datasets, with a small to moderate effect size and a high achieved power.

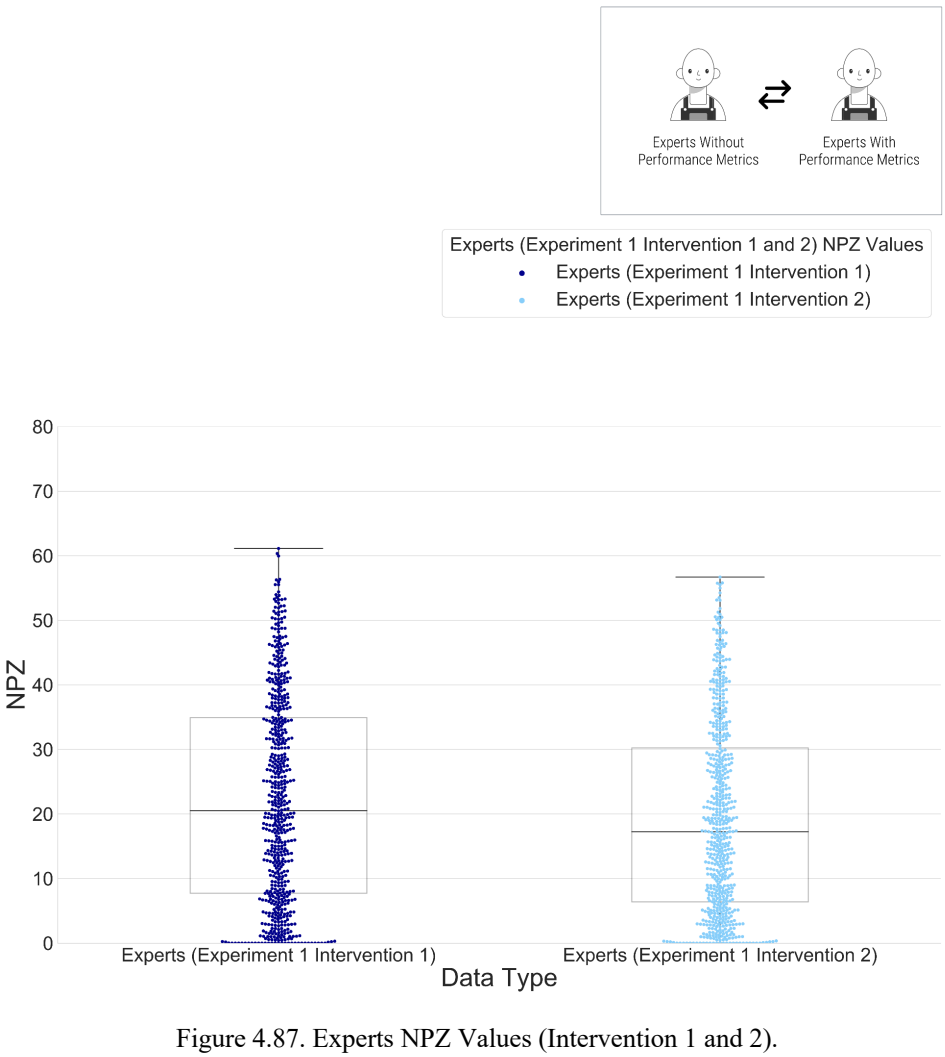


Figure 4.87. Experts NPZ Values (Intervention 1 and 2).

6. Non-Experts Comparison (Experiment 1 Intervention 1 and 2)

There is no significant difference in the NPZ values among non-experts when the performance metrics are not provided and when provided to users (Experiment 1 Intervention 1 and 2). Both these datasets are not normally distributed. The calculated p-value is 0.2, which is higher than the 0.05 alpha value. This result demonstrates the similarity of final non-dominated solutions NPZ values among non-experts with and without performance metrics provided to users.

The results within two user groups shows the importance of expertise in the generative design process and how providing complex performance information to experts can help improve their decision-making process. In contrast, it does not significantly improve the decision-making

process for non-experts. It is possible that the non-experts had difficulty conceptualizing NPZ due to its spatial nature and its impact on the design. Therefore, despite verbally expressing the importance of NPZ during the interviews, NPZ values did not significantly improve when non-experts were provided with performance metrics values.

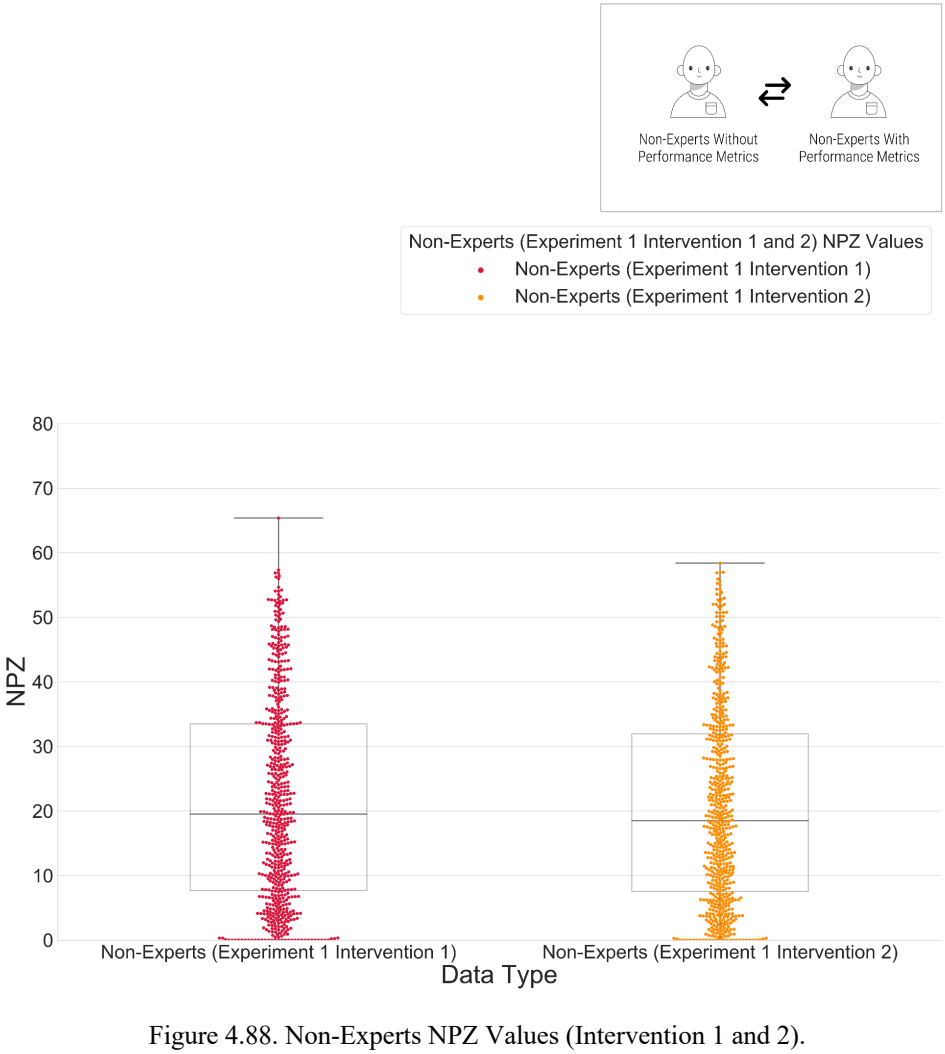


Figure 4.88. Non-Experts NPZ Values (Intervention 1 and 2).

7. Experts and Non-Experts Comparison (Experiment 1 Intervention 1)

There is no significant difference in the NPZ values between experts and non-experts when the performance metrics are not provided to users (Experiment 1 Intervention 1). Both these datasets are not normally distributed. The calculated p-value is 0.2, which is higher than the 0.05 alpha value. This result demonstrates the similarity of final non-dominated solutions' NPZ values between experts and non-experts. This may reflect a similarity in aesthetic decisions stemming from NPZ between experts and non-experts.

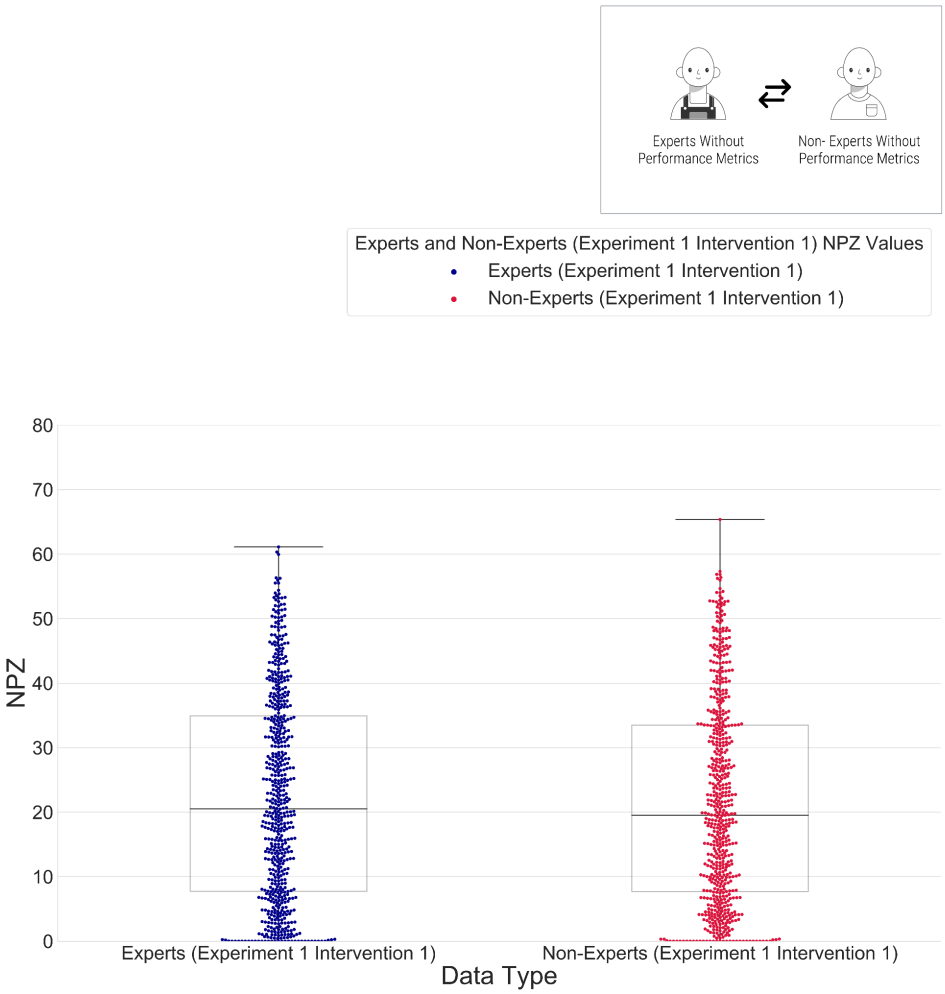


Figure 4.89. Comparison of Experts and Non-Experts NPZ Values (Intervention 1).

8. Experts and Non-Experts Comparison (Experiment 1 Intervention 2)

There is no significant difference in the NPZ values between experts and non-experts when the performance metrics are provided to users (Experiment 1 Intervention 2). Both these datasets are not normally distributed. The calculated p-value is 0.06, which is slightly higher than the 0.05 alpha value. While this result shows a trend towards statistical significance, it does not reach the statistical significance level defined (alpha level 0.05) in this study.

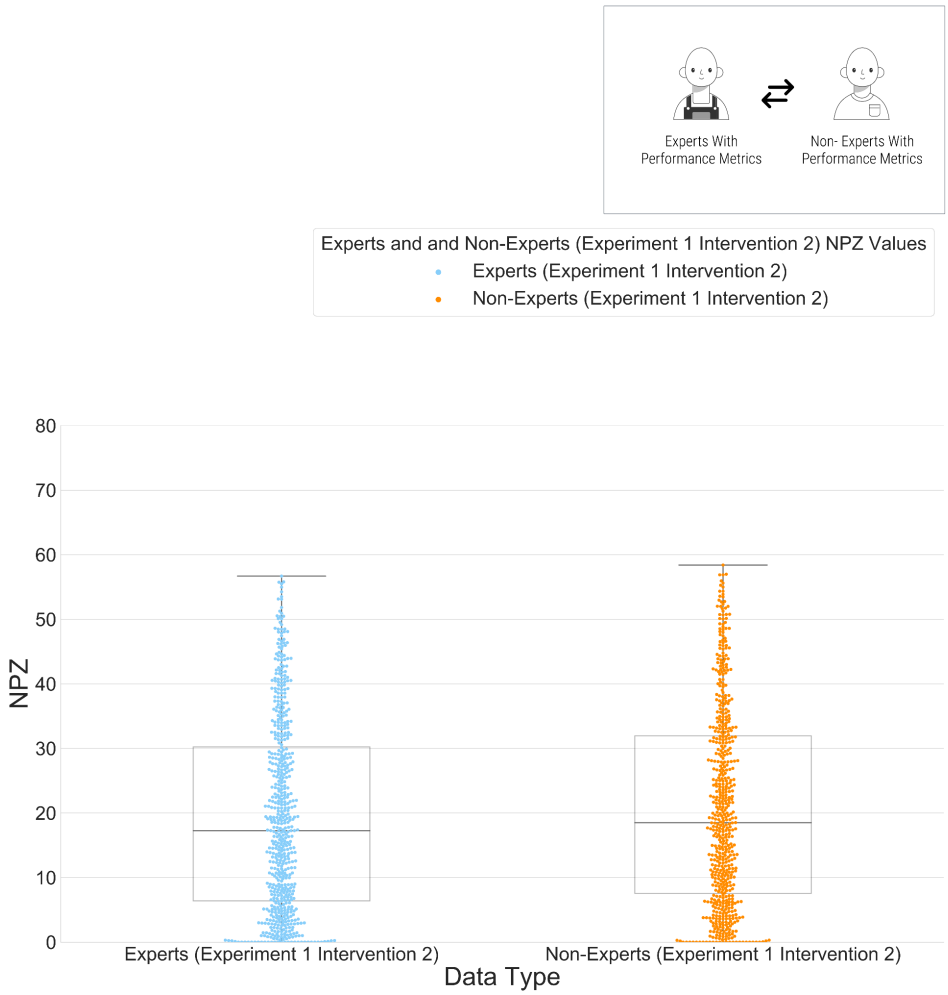


Figure 4.90. Comparison of Experts and Non-Experts NPZ Values (Intervention 2).

UOS Values (Experiment 1 Intervention 1 and 2)

9. Experts Comparison (Experiment 1 Intervention 1 and 2)

Based on user interviews, UOS has the second highest ranking with seven experts considering it as the highest-rated objective. The statistical analysis demonstrated that there is a significant difference in the UOS values among experts when the performance metrics are not provided and when provided to users (Experiment 1 Intervention 1 and 2). Both these datasets are not normally distributed. The calculated p-value is 0.01, which is lower than the 0.05 alpha value. This result demonstrates the difference of final non-dominated solutions UOS values among experts with and without performance metrics provided to users, where experts with performance metrics provided resulted in higher UOS values.

The statistical power in this experiment is 0.8, showing a high likelihood of finding a true effect. The effect size is -0.1, showing a small effect size. Overall, the results show a statistically

significant difference between the two groups, with a small effect size and a high statistical power value. This result shows that both experts and non-experts can benefit from receiving performance metrics, however, the usefulness of this feedback also relates to the level of difficulty of the provided performance metric.

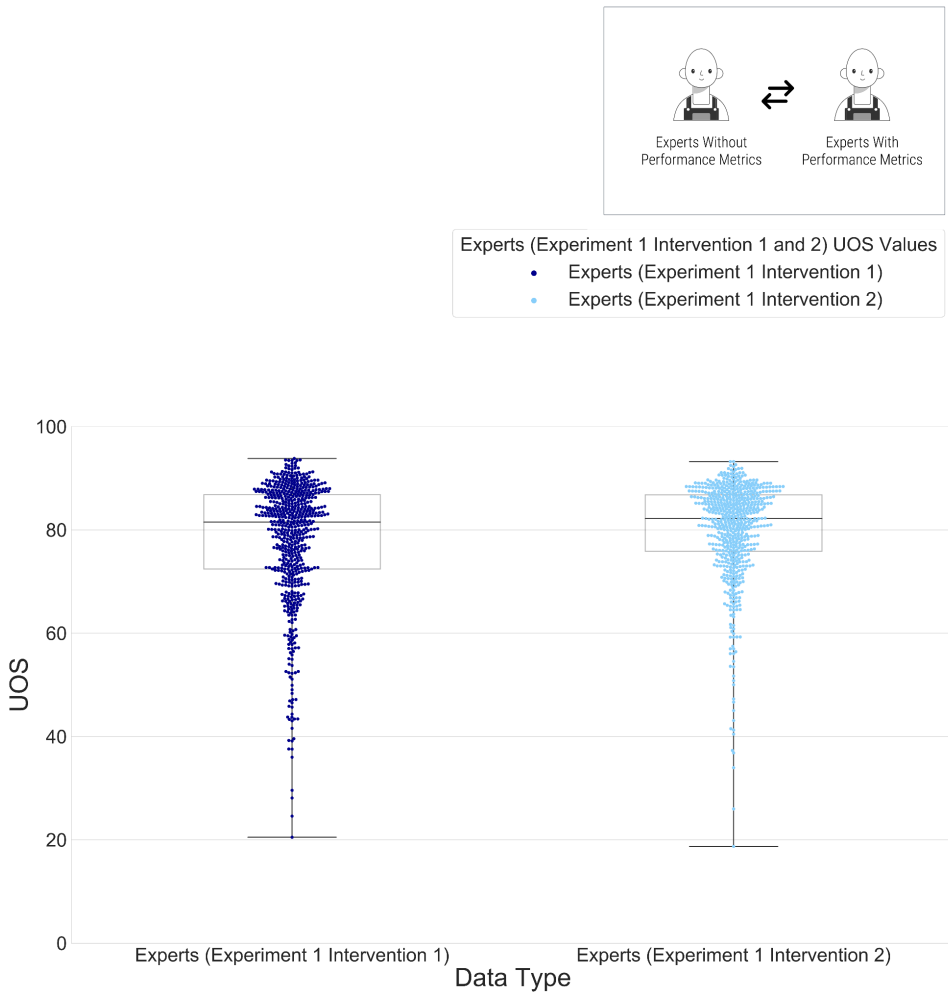


Figure 4.91. Comparison of Experts UOS Values (Intervention 1 and 2).

10. Non-Experts Comparison (Experiment 1 Intervention 1 and 2)

UOS has the second highest ranking with seven non-experts considering it as the highest-rated objective (n=7). Similar to experts, there is a significant difference in the UOS values among non-experts when the performance metrics are not provided and when provided to users (Experiment 1 Intervention 1 and 2). Both these datasets are not normally distributed. The calculated p-value is 0.0008, which is lower than the 0.05 alpha value. This result demonstrates the difference of final non-dominated solutions' UOS values among non-experts with and without performance metrics provided to users, where non-experts with performance metrics provided, resulted in higher UOS values. The statistical power calculated in this study is 0.96, which shows a very high likelihood

of finding a true effect. The effect size is -0.2, showing a small to moderate effect size. Overall, the results of this study show a significant difference between the two datasets, with a small to moderate effect size and a very high statistical power. The results of UOS shows that non-experts can benefit from receiving performance metrics, where simply providing a well defined rubric to non-expert users can demonstrably improve the quality of their feedback.

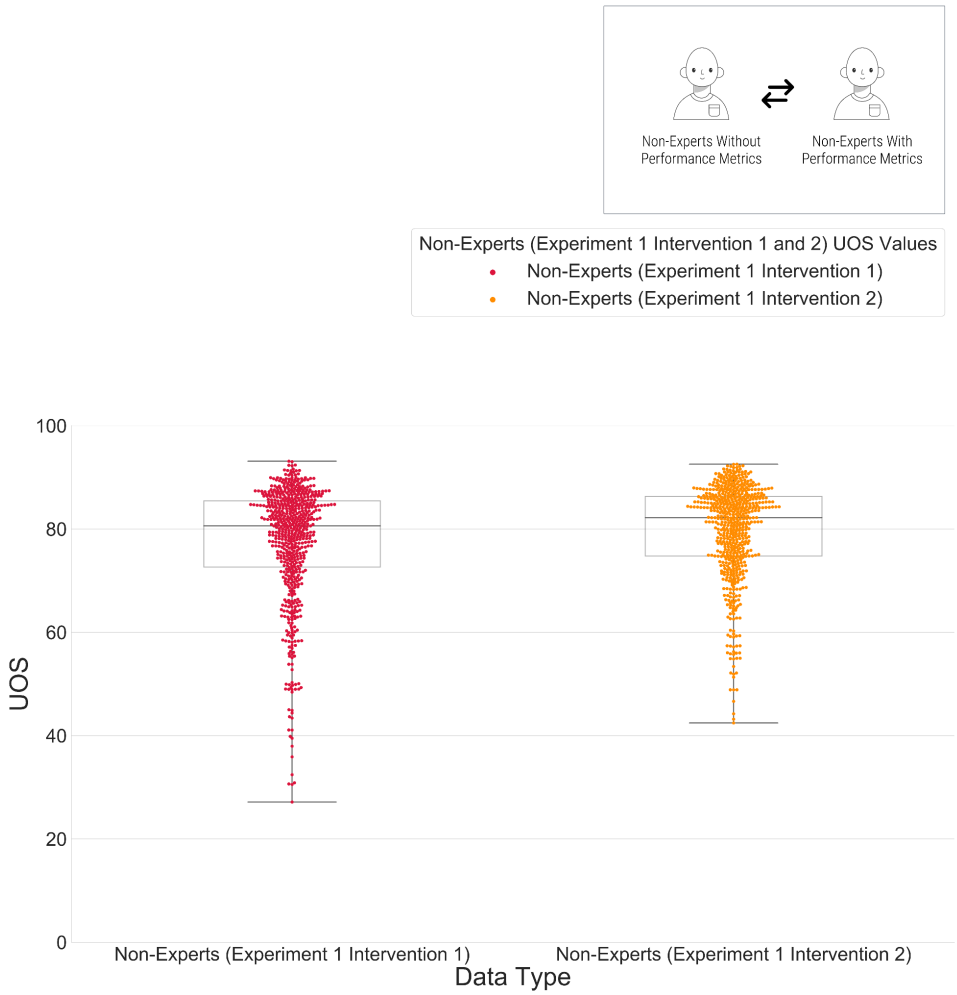


Figure 4.92. Comparison of Non-Experts UOS Values (Intervention 1 and 2).

11. Experts and Non-Experts Comparison (Experiment 1 Intervention 1)

There is no significant difference in the UOS values between experts and non-experts when the performance metrics are not provided to users (Experiment 1 Intervention 1). Both these datasets are not normally distributed. The calculated p-value is 0.05, which is slightly higher than the 0.05 alpha value. This result demonstrates the similarity of final non-dominated solutions' UOS values between experts and non-experts.

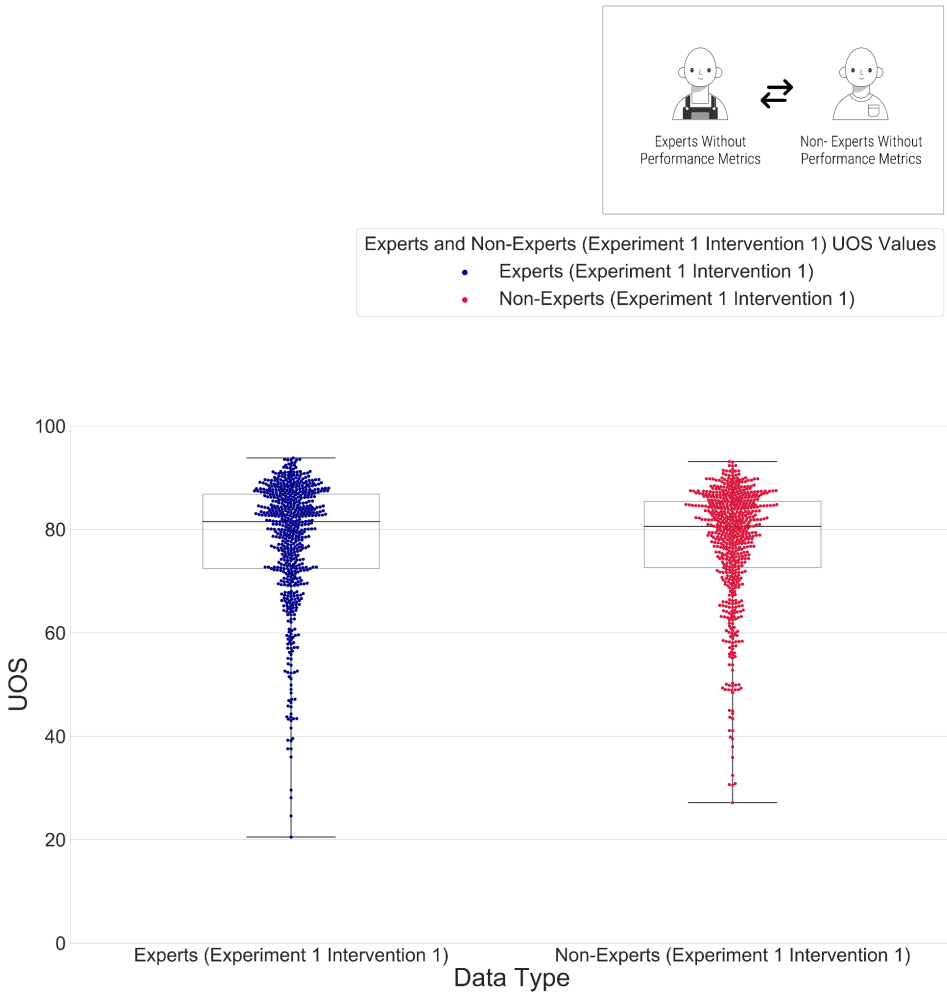


Figure 4.93. Comparison of Experts and Non-Experts UOS Values (Intervention 1).

12. Experts and Non-Experts Comparison (Experiment 1 Intervention 2)

There is no significant difference in the UOS values between experts and non-experts when the performance metrics are provided to users (Experiment 1 Intervention 2). Both these datasets are not normally distributed. The calculated p-value is 0.1, which is higher than the 0.05 alpha value. This result demonstrates the similarity of final non-dominated solutions' UOS values between experts and non-experts. As discussed previously, for both experts and non-experts, UOS values were significantly improved when users were provided with performance metrics. However, the result of this analysis shows that both groups' final solutions have similar UOS values when provided with performance metrics.

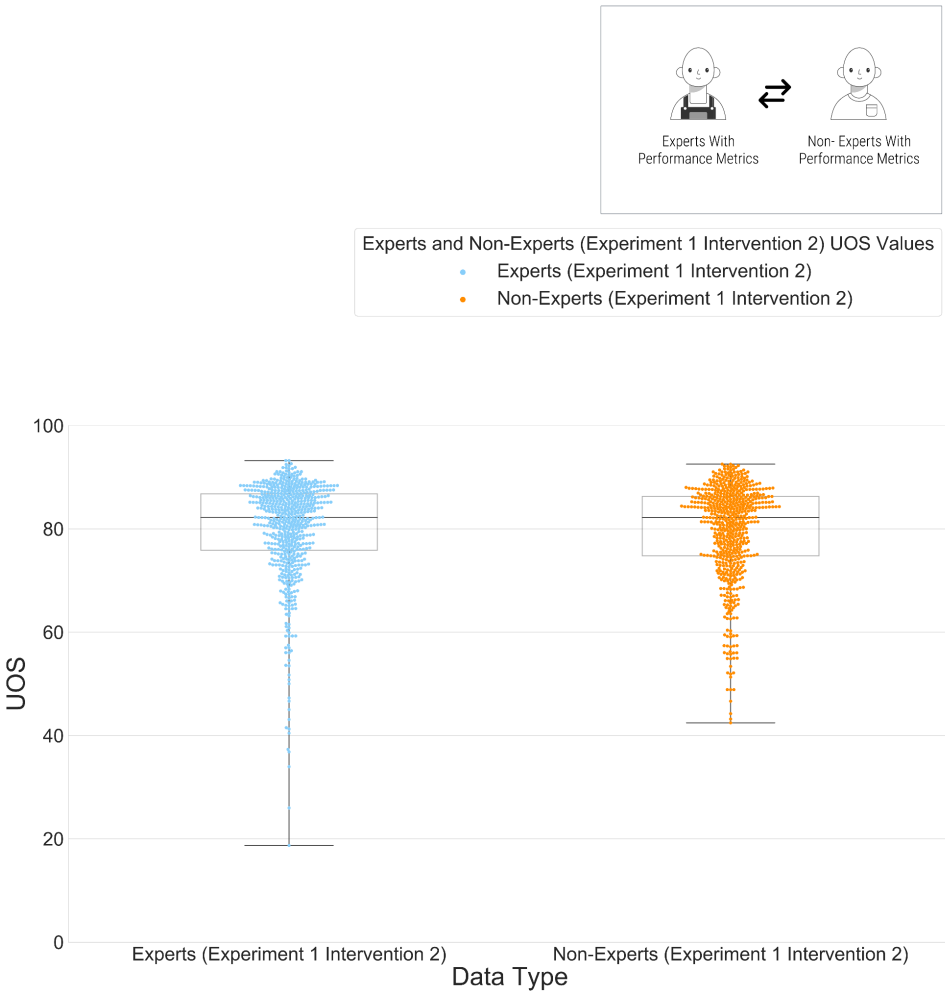


Figure 4.94. Comparison of Experts and Non-Experts UOS Values (Intervention 2).

4.3.5 Discussion

4.3.5.1 Objective Functions Results

The importance of providing rubrics to non-experts participating in design evaluations was discussed in the literature review in chapter 2. Providing rubrics to non-design experts was discussed to tackle the issues of unproductive [245]. In order to receive more useful feedback from non-expert users performance metrics were introduced in Experiment 1 [245]. To create consistency among experts and non-experts' evaluations of the designs, performance metrics were provided to both experts and non-experts in Experiment 1.

Among experts, the majority of users rated NPZ as the most important objective function (n=10). Some users rated UOS as the most important objective function (n=7), and some users rated AV as the most important objective function (n=4). Overall, the majority of users discussed the importance of sustainable design criteria in their design preferences and a large influence in their

ratings, and they viewed NPZ as an important sustainable design metric. The impact of providing this metric was evident in the solutions generated in Intervention 2, when users were provided with NPZ values. There is a significant difference in the NPZ values among experts when the performance metrics are not provided compared to when provided to users (Experiment 1 Intervention 1 and 2). When provided with performance metrics, experts' solutions resulted in lower NPZ values.

Among non-experts, the majority of users rated NPZ as the most important objective function (n=11). UOS has the next highest ranking with seven non-experts considering it as the highest-rated objective (n=7). Some non-experts rated AV as the most important objective function (n=3). However, there was no significant difference in the NPZ values among non-experts when the performance metrics were not provided compared to when provided to users (Experiment 1 Intervention 1 and 2).

The comparison of NPZ results within two user groups show the importance of expertise in the generative design process and how providing complex performance information to experts can help improve their decision-making process while it doesn't significantly improve the decision-making process for non-experts. One possibility is that the non-experts had difficulty conceptualizing NPZ due to its spatial nature and its impact on the 3D design. Therefore, despite verbally expressing the importance of NPZ during the interviews, NPZ values did not significantly improve when non-experts were provided with performance metrics values.

UOS results however showed a different impact on final solutions. Even though the majority of both user groups rated NPZ as the most important performance metric and UOS as the second most important performance metric, both groups' solutions improved in UOS values when provided with UOS values. This can be due to the clarity of UOS's definition visually on the massing, where users can see the open space easily in each massing and can prioritize that in their ratings. This finding was discussed in the user interviews, such as the discussion with non-expert user 2B2:

> "*UOS was the most important and easiest to see in the massing options. Then the NPZ was pretty important, although harder for me to relate to the design.*"

Even though only a small portion of experts rated AV as the highest priority (n=4), AV values were reduced significantly among experts in Intervention 2. For non-experts, there is no significant difference in the AV values when the performance metrics are not provided and when provided to users (Experiment 1 Intervention 1 and 2). This shows the importance of expertise in the generative design process and how experts acknowledged the importance of Area Violation in their decision-making. Area violation is directly tied to program requirements in building design and violations from required gross floor area makes massings not feasible.

The results of the three objective functions studies align with related studies in the literature, where providing rubrics to non-experts resulted in no substantial difference in their feedback compared to experts [245]. Experiment 1 results showed no significant difference in AV, NPZ, and UOS values between experts and non-experts without access to performance metrics. However, results of Experiment 1 also showed that there was no significant difference in NPZ and UOS values

between experts and non-experts when users were not provided with performance metrics. Based on Experiment 1 results, there was a significant difference in AV values between experts and non-experts.

Prior to this experiment, the assumption was that the final solutions in Intervention 2 for both experts and non-experts would be improved (higher) when provided with performance metrics, but the results showed otherwise. For both experts and non-experts, there was no significant difference in the HV values between Experiment 1 Intervention 1 and 2. This result demonstrates the consistency of user preference impact on the optimization results, and how final solutions tend to perform similarly when provided with performance metrics.

For both experts and non-experts, when comparing users with the NSGA-II Alone (Experiments 1 and 2), there was a significant difference in HV values. This result aligns with expectations, as adding user ratings in this experiment acts as a conflicting objective among the proposed objectives which can depress the converging performance of final solutions.

4.3.5.2 Building Heights Results

Based on user interviews, there was an agreement among both experts and non-experts in their building height preferences. For example, expert EE1 discussed their massing preferences:

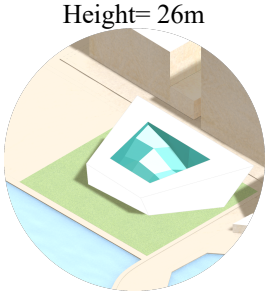
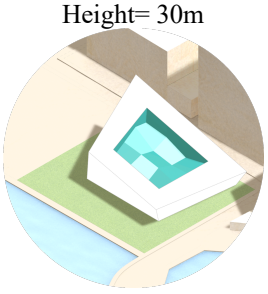
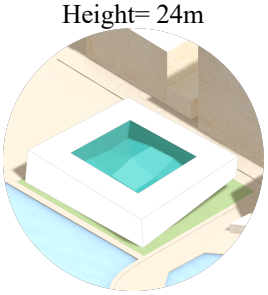
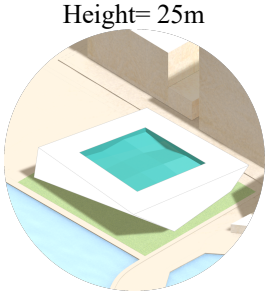
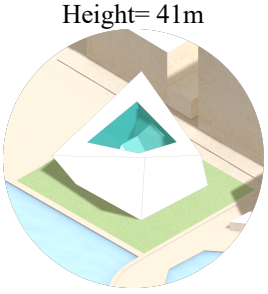
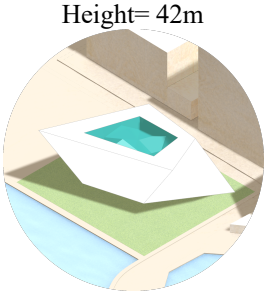
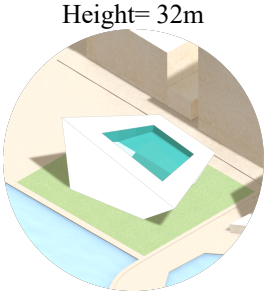
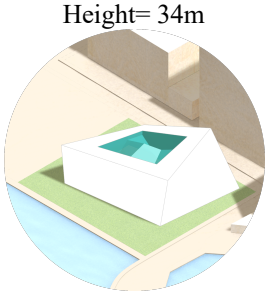
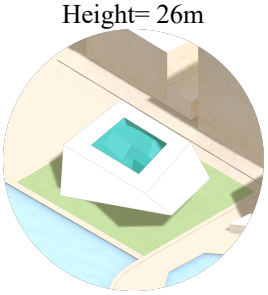
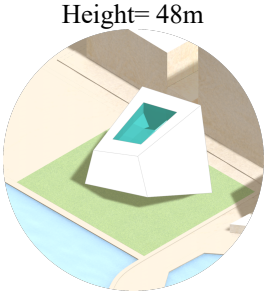
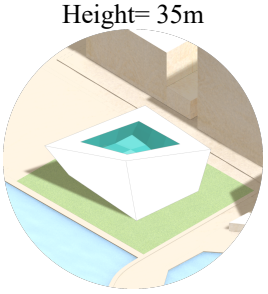
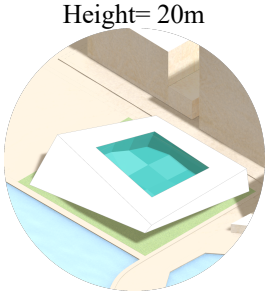
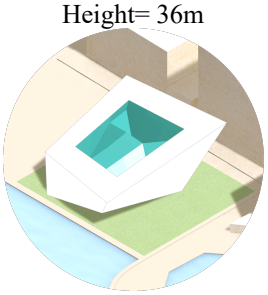
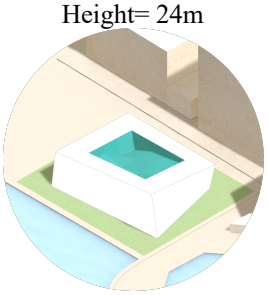
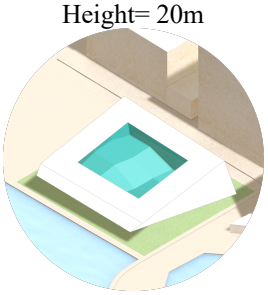
> "*I shied away from taller buildings with low footprints. I preferred the buildings with openings where the pedestrians were walking by, and when the openings on the roof were visible on the street view. That was a more preferable option for me.*"

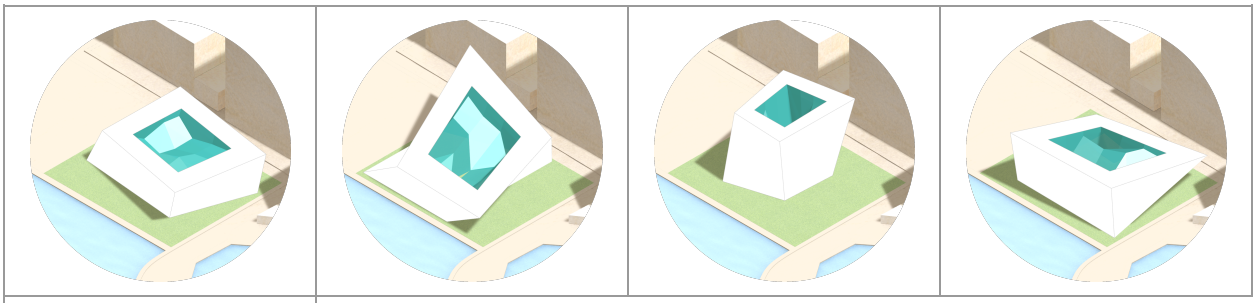
In addition, non-expert user 2N2 discussed:

> "*I was favoring a form that was more cantilevered that could maximize the floor space around the building. It was shorter than the average massing you saw.*"

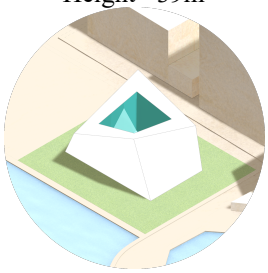
When analyzing building heights, the difference between Interventions 1 and 2 was statistically significant among experts and among non-experts. For experts and non-experts, the final solutions were taller when users were provided with performance metrics (Intervention 2). However, between experts and non-experts, the difference in building heights for both Intervention 1 and 2 is not statistically significant. In addition to statistical analysis, the preference of users in building heights can be further investigated by visualizing the highly rated final solutions by experts and non-experts. As can be seen in the table below, the buildings follow a trend where the majority of forms have an east-west orientation, while their height is not maximized (non-tower options).

Table 4.9. Sample of Highly Rated Solutions by Experts.

AA1 Height= 26m 	BB1 Height= 30m 	CC1 Height= 32 m 	DD1 Height= 24m 
EE1 Height= 25m 	FF1 Height= 41m 	GG1 Height= 42m 	HH1 Height= 32m 
II1 Height= 34m 	JJ1 Height= 26m 	KK1 Height= 48m 	LL1 Height= 35m 
MM1 Height= 20m 	NN1 Height= 36m 	OO1 Height= 24m 	PP1 Height= 20m 
QQ1 Height= 30m	RR1 Height= 28m	SS1 Height= 54m	TT1 Height= 27m



UU1
Height= 39m



Overall, for all users, the final solutions were taller when users were provided with performance metrics compared to those without performance metrics. This result can be explained by the importance of providing performance metrics on their decision-making, as taller buildings tend to perform better due to their ability to offer lower AV and higher UOS values and depending on the building depth and width, lower NPZ values. The majority of users did not prefer this typology in Experiment 1 Intervention 1, but once presented with performance metrics, users' final solutions tended to include taller buildings with smaller footprints. Between experts and NSGA-II Alone based on experts, the difference in building heights between Experiment 1 and 2 Interventions 1 and 2 are statistically significant.

Additionally, between non-experts and algorithm results based on non-experts (Experiment 1 and 2), the difference in building heights in Interventions 1 and 2 results are statistically significant. These results suggest that the final solutions tend to be taller for NSGA-II Alone based on users' initial populations, across the board when not accounting for their preferences.

The difference in building heights is statistically significant between experts without performance metrics provided to users, and the algorithm results based on these experts (Experiment 1 and 2 Intervention 1). For experts, the final solutions tend to be shorter than the algorithm results. The difference in building heights is statistically significant between experts with performance metrics provided to users, and the algorithm results based on experts (Experiment 1 and 2 Intervention 2). For non-experts, the results found a similar difference between Experiments 1 and 2. These results suggest that the final solutions are shorter for experts than their algorithm-based counterparts. These results show the importance of including experts in the generative design process. The NSGA-II Alone tended to achieve better performance in the final solutions but resulted in building types with taller massing options (tower typology) that were less preferred among experts. However, experts' final solutions reached a lower HV value with shorter massing options, which

were more preferred by experts and are more likely to be selected as the selected option during the later stages of the architectural design process.

4.3.5.3 Future Research

During an exploratory follow-up experiment, an expert user was presented with the final solutions based on Experiment 1 Intervention 1 and 2 and final solutions for Experiment 2 Intervention 1 and 2 and was asked to rate the final generated massing options. The expert was not provided any performance metrics to perform this task and was asked to rate the solutions purely from an aesthetic preference perspective.

The objective of this experiment was to understand the user rating prediction bias in the designed NSGA-II algorithm and which Experiments and Interventions generated the highest-rated solutions. Based on the results, the expert rated their final solutions higher when they directed the generative design process purely based on their aesthetic preferences (Experiment 1 Intervention 1) than when the algorithm generated the massing options (Experiment 2 Intervention 1). On the other hand, the expert rated solutions generated by the algorithm (Experiment 2 Intervention 2) were higher than when they directed the generative design process using performance metrics themselves (Experiment 1 Intervention 2). This experiment can be further explored by studying all experts' and non-experts' final solutions to investigate further the impacts of providing performance metrics to users in the generative design process.

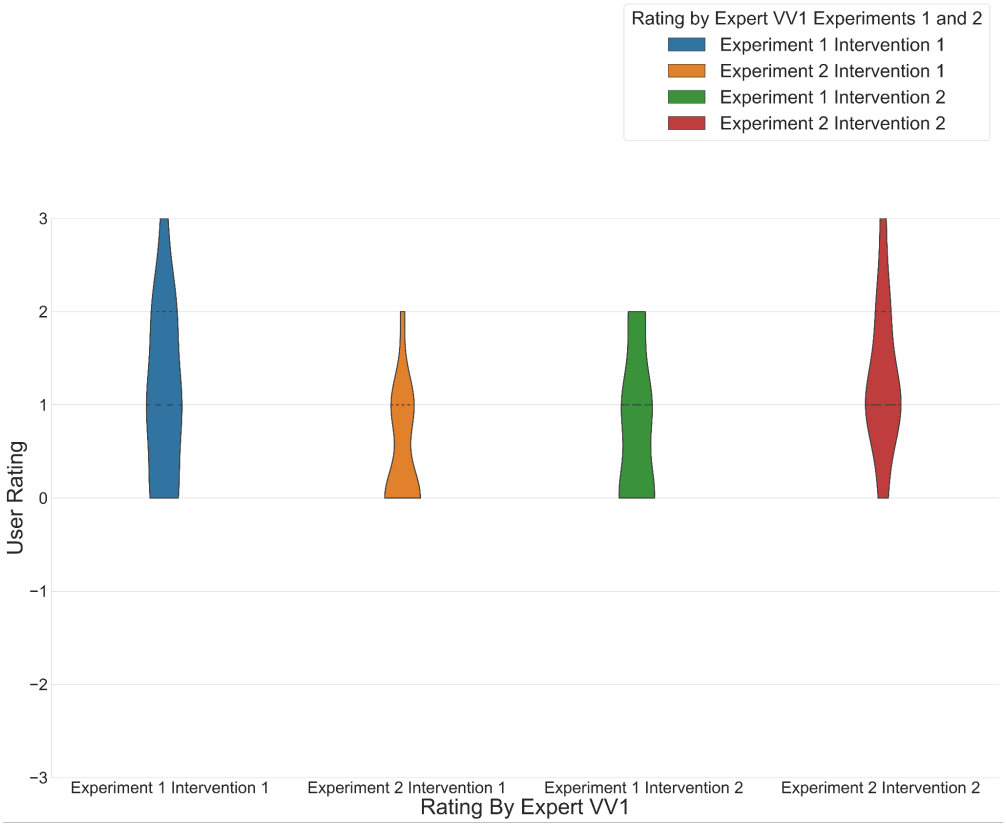


Figure 4.95. Rating Results in the Follow-up Experiment.

Table 4.10. Expert VV1 Ratings of Final Solutions in Experiments 1 and 2.

Experiment	User Rating Mean	User Rating Median
Experiment 1 Intervention 1	1.1	1.0
Experiment 2 Intervention 1	0.5	0.0
Experiment 1 Intervention 2	0.8	1.0
Experiment 2 Intervention 2	1.3	1

When comparing user ratings with the estimated user ratings by the algorithm in Experiment 1, the algorithm was able to successfully estimate the user's ratings in both interventions. For Intervention 1, the user rating's prediction bias has a mean value of 0.8. For Intervention 2, the user rating's prediction bias has a mean value of 0.7. These results show the proposed algorithm's reliability in predicting user's rating based on shape similarities during generative design, allowing for the small subset to only be provided to the user to manually rate. Future research can investigate this finding further. Figure 4.96 shows the user prediction bias, which is calculated by subtracting the algorithm's predicted rating by the user's actual rating provided during the follow-up study. These studies can be further explored as a new set of experiments in the future, where all experts and non-experts view and rate their entire population of generated massing options at the end of a run to calculate higher accuracy of the predicted rating by the algorithm.

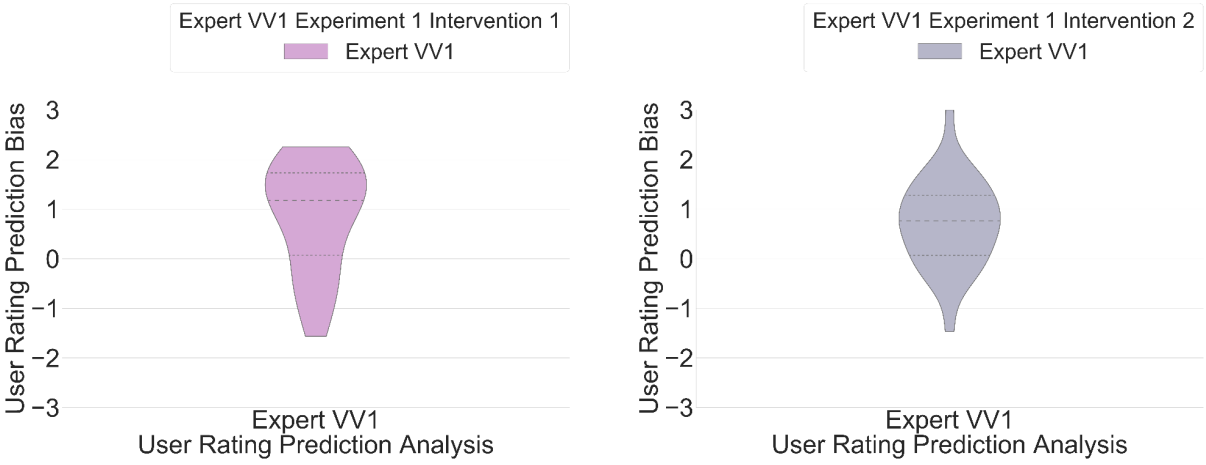


Figure 4.96. User Rating Prediction Bias.

In the future, larger user groups can be used to train machine learning models to understand user massing preferences for specific building programs and site typologies. In order to accomplish this, a dataset will be created by including the highest and lowest rated solutions based on Experiment 1 Intervention 1 results from all experts and non-experts can be selected to conduct a subjective test inspired by Corsini et al.'s research in Perceptual Metrics for Static and Dynamic Triangle Meshes [48]. The lowest-rated solutions based on Experiment 1 Intervention 1 will be

used to calibrate user ratings [48]. In a survey study, a large user group (mix of experts and non-experts) will be asked to rate each massing option a number from -3 to +3, similar to Experiment 1. In the next step, a mean opinion score (MOS) will be calculated for each massing option by averaging the score given by the user group [48]. Using the subjective MOS, the correlation between objective functions associated with massing options can be determined to understand the performance of the objective functions in relation to user feedback [48].

Next, the user preference can be simulated using a supervised learning method suggested by Bhat et al. [18]. Using this method, the 3D massing options will be used as inputs in a 3D convolutional neural network where they will turn into 1D vectors (feature vectors) for performance purposes [18]. A supervised learning problem will then be defined by using the feature vectors as input and user preferences as outputs [18]. Some recommended algorithms to try for simulating user preferences include Linear Support Vector Machines (SVM), Decision Trees, and Gradient Boosting Classifier or more advanced algorithms such as Deep Neural Networks [18]. Such a model can be created for both experts and non-experts for different building programs and sites for different climates. Figure 4.97 shows the proposed workflow for training a machine learning model to simulate user preference in a future step.

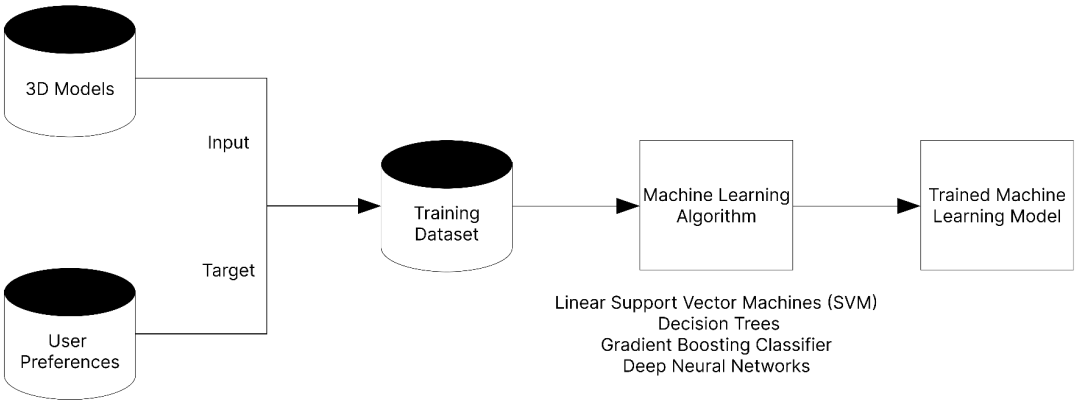


Figure 4.97. Machine Learning process to simulate user preference in 3D models [18].

4.3.6 Experiment Conclusions

Prior to conducting Experiment 1, it was expected that the performance of the experts' solutions (HV value) would be higher than that of the non-experts. However, when comparing experts' and non-experts' Experiment 1 results, there was no significant difference in the final solution's performance (HV value) between experts and non-experts. The visualization of experts' and non-experts' HV values changes over ten generations for all experiments showed the changes in user behaviors with and without providing performance metrics. For some experts and non-experts, both interventions had similar behavior in HV values over time. For some users, Intervention 2 resulted in lower HV values over time. Some users' HV values in Intervention 2 improved over time compared to their Intervention 1 results.

Prior to conducting Experiment 2, it was expected for the algorithm results without the user's rating to have a higher performance (HV value) compared to Experiment 1 results. When comparing the experts' results and the algorithm results based on experts (Intervention 1), there is a significant difference in HV values where the algorithm reaches higher HV values. In addition, there is a significant difference in HV values between experts and algorithms based on Intervention 2 results. These results are similar among non-experts, where the algorithm results have a higher performance compared to the non-experts' solutions in Experiment 1 Interventions 1 and 2. Despite the higher performance of the algorithm in experiment 2, these results show the importance of including users in the generative design process. The user preference on massings can be in conflict with other performance metrics that are taken into account by the algorithm, which can lead to more desirable massing options.

While most experts and non-experts rated NPZ as the most important objective function, the NPZ values had a higher impact on experts compared to non-experts. These results within two user groups show the importance of expertise in the generative design process and how providing complex spatial information to experts can help improve their decision-making process. In contrast, it doesn't significantly improve the decision-making process for non-experts. It is possible that the non-experts had difficulty conceptualizing NPZ and its impact on the design. While non-experts verbally expressed the importance of NPZ during the interviews, NPZ did not significantly improve when they were provided with the values.

UOS was rated as the second highest-rated objective function, and it showed a high impact on experts and non-experts when provided with performance metrics (Experiment 1 Intervention 2). While the majority of both user groups rated NPZ as the most important metric and UOS as the second objective, both groups' solutions improved in UOS values. This can be due to the clarity of UOS's definition visually on the massing, where users can see the open space easily in each option and can prioritize that in their ratings. This finding was also discussed by non-expert user 2B2:

> "*UOS was the most important and easiest to see in the massing options, then the NPZ was pretty important, although harder for me to relate to the design.*"

For experts and non-experts, the final solutions were overall taller when users were provided with performance metrics (Experiment 1 Intervention 2). However, between experts and non-experts, the difference in building heights in Intervention 1 is not statistically significant. This result can be explained by the importance of providing performance metrics on their decision metric, as taller buildings tend to perform better due to their ability to provide lower AV and higher UOS values, and depending on the building depth and width, lower NPZ values. Most users did not prefer this typology (tower typology) in Experiment 1 Intervention 1, but once presented with performance metrics, users' final solutions included taller buildings with smaller footprints.

Overall, the difference in building heights is statistically significant between all users without performance metrics provided to users and the NSGA-II Alone based on these users. In addition, the difference in building heights is statistically significant between all users with performance metrics provided to users and the NSGA-II Alone based on these users. These results suggest that the final solutions included shorter buildings for all users than their algorithm-based counterparts.

In the next step, during an exploratory follow-up experiment, an expert user was presented with the final solutions based on Experiment 1 Intervention 1 and 2 and final solutions for Experiment 2 Intervention 1 and 2 and was asked to rate the final generated massing options. When comparing user ratings with the estimated user ratings by the algorithm in Experiment 1, the algorithm was able to successfully estimate the user's ratings in both interventions. For Intervention 1, the user rating's prediction bias has a mean value of 0.8. For Intervention 2, the user rating's prediction bias has a mean value of 0.7. These results show the proposed algorithm's reliability in predicting user's rating based on shape similarities during generative design, and future research can investigate this finding further. In addition, the expert rated their final solutions higher when they directed the generative design process purely based on their aesthetic preferences (Experiment 1 Intervention 1) than when the algorithm generated the massing options (Experiment 2 Intervention 1). This result can be further investigated by conducting more experiments and shows the importance of including participants in the generative design process without performance metrics. These studies can be further explored as a new set of experiments in the future, where all experts and non-experts view and rate their generated massing options for higher accuracy of the predicted rating by the algorithm.

In addition, larger user groups can be used to train machine learning models to understand user massing preferences for specific building programs and site typologies. In order to accomplish this, a dataset will be created by including the highest and lowest rated solutions based on Experiment 1 Intervention 1 results from all experts and non-experts can be selected to conduct a subjective test inspired by Corsini et al.'s research in Perceptual Metrics for Static and Dynamic Triangle Meshes [48]. A supervised learning problem will then be defined by using the feature vectors as input and user preferences as outputs [18]. Some recommended algorithms to try for simulating user preferences include Linear Support Vector Machines (SVM), Decision Trees, and Gradient Boosting Classifier or more advanced algorithms such as Deep Neural Networks [18]. Such a model can be created for both experts and non-experts for different building programs and sites for different climates.

Chapter 5: Conclusion and Future Work

5.1 Discussion

This dissertation introduced a novel Operative Generative NSGA-II algorithm and its integration into a participatory design process for architectural massing. The methods and experiments proposed in this research try to answer the following question: How can we propose a novel interactive evolutionary generative design process leveraging users' feedback to improve early-stage massing design in the field of architecture?

The literature was reviewed in three main subject categories related to this dissertation: generative design, participatory design in architecture, and user experience assessment in AR.

We initially introduced the first prototypes of InsightXR in AR and VR, followed by presenting the final design prototype of InsightXR. The final design of InsightXR uses a Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to generate design alternatives inspired by how architects start early massing studies using Operative Design [13]. The Operative Design process in this research is based on the volumetric operations discussed by Mari and Yoo and uses an operative NSGA-II algorithm [148], [13]. The design operations explored in this research focus on displacement and subtraction by using expansion and carving operations as the main verbs applied to the proposed operative generative design [148], [13]. In InsightXR, users are divided into two categories: experts and non-experts. The application prompts users to choose between designer (expert) and reviewer (non-expert) roles. As a designer, a user can define new generative design projects, participate in a generative design project as a reviewer, or provide feedback on single massing projects.

The Experiments of this dissertation include three sections that aim to understand the benefits and limitations of the proposed generative design process.

In Experiment 1, titled Operative Generative Design Using Non-Dominated Sorting Genetic Algorithm II (NSGA-II), the goal was to introduce a new application of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) using the Pymoo framework for the field of Operative Design. Experiment 1 included nine experiments to test the algorithm's geometry optimization capabilities based on objective functions reflecting common architectural design goals, including FAR, NPZ, RBOS, and UOS. In experiments 4-9, a fifth objective was introduced: the AV in relation to target square footage. Due to its importance in architectural design practice, FAR was used as a consistent objective function in all experiments. Among the non-dominated solutions visualized in all experiments, there were hybrids (combinations of carving and expansion) and solutions with only either carving or expansion as the operation. All experiments were able to achieve their minimization/maximization goals. Based on the results of this experiment, the solutions that had the best tradeoff between the objective functions (Case 2) also resulted in meaningful massing options that can be realistically selected and continued to next stages of architectural design.

Experiment 2 introduced InsightXR, and the application of a novel operative generative design algorithm using Non-Dominated Sorting Genetic Algorithm II (NSGA-II) in the Pymoo framework that was introduced in Experiment 1. To test InsightXR's capabilities, a case study was presented to generate 3D design alternatives in real time for a mixed-use project in San Francisco, USA. The design objectives explored in this research were RBOS, UOS, and AV. Design alternatives were generated for five generations and visualized in InsightXR. In this experiment, design solutions in each generation were successfully clustered into three groups, where each cluster had distinct behaviors in terms of all three objective functions. In all five generations, cluster 1 included design solutions with lower AV values and higher RBOS values due to design options that provide larger floor plates. As the generations progressed, design options in cluster 1 provided an improved RBOS value compared to previous generations. In this case study, even though the goal was to maximize RBOS and minimize AV, it is essential to note that creating design options with smaller courtyards while having a deep floor plate (e.g., case 1 in generation 4) can reduce the access to daylighting and natural ventilation, which can result in non-desirable architectural design. It is ultimately the designer's role to evaluate the proposed non-dominated solutions provided by InsightXR to choose the most applicable design for the site at hand. This experiment showed a successfully diverse solution space in each generation, with clusters dividing the solution space efficiently into three main groups. However, Experiment 2 showed an important limitation of the selection criteria for massing options per cluster, since the most representative massing per cluster was selected for visualization. However, in order to receive and apply user ratings across massing options in each cluster, the representative massings should be sufficiently representative of the clusters while being as different from each other. In order to implement this, the selected solution was updated to be located the closest to the average shape similarity of each of the clusters.

Experiment 3 in this dissertation assessed the impact of including users in the generative design process using InsightXR for human-subject experiments and conducting post-study experiments. The human-subject experiments were titled Experiment 1, and the post-study experiments were titled Experiment 2. Experiment 1 was conducted in person at the XR Lab at UC Berkeley, lasting one hour per study. This experiment was conducted among experts and non-experts using two interventions. Experiment 2 was conducted after Experiment 1 using only the NSGA-II algorithm without including user feedback, using the initial population of Experiment 1 as seeds. For both experiments, the following analyses were performed:

1. Hypervolume indicator assessment
2. Shape analysis focusing on height analysis to evaluate the distribution of massings
3. Objective Function Analysis (Performance Metrics):
    a. Area Violation (AV)
    b. Non-Passive Zone (NPZ)
    c. Usable Open Space (UOS)

Among experts, the majority of users rated NPZ as the most important objective function (n=10). Some users rated UOS as the most important objective function (n=7), and some users rated AV as the most important objective function (n=4). Overall, the majority of users discussed the importance of sustainable design criteria in their design preferences, and they viewed NPZ as a sustainable design objective. Among non-experts, the majority of users rated NPZ as the most important objective function (n=11). UOS has the next highest ranking with seven non-experts considering it as the highest-rated objective (n=7). Some non-experts rated AV as the most important objective function (n=3).

There was a significant difference in the NPZ values among experts when the performance metrics were not provided compared to when they were provided to users (Experiment 1 Intervention 1 and 2). When provided with performance metrics, experts' solutions resulted in lower NPZ values. However, there was no significant difference in the NPZ values among non-experts when the performance metrics were not provided compared to when provided to users (Experiment 1 Intervention 1 and 2). This result demonstrated the similarity of final non-dominated solutions NPZ values among non-experts with and without performance metrics provided to users. The results within two user groups showed the importance of expertise in the generative design process and how providing complex performance information to experts can help improve their decision-making process. In contrast, it did not significantly improve the decision-making process for non-experts. It is possible that the non-experts had difficulty conceptualizing NPZ due to its spatial nature and its impact on the design. Therefore, despite verbally expressing the importance of NPZ during the interviews, NPZ values did not significantly improve when non-experts were provided with performance metrics values.

There was no significant difference in the UOS values between experts and non-experts when the performance metrics were not provided to users (Experiment 1 Intervention 1). Moreover, there was no significant difference in the UOS values between experts and non-experts when the performance metrics were provided to users (Experiment 1 Intervention 2). These results demonstrated the similarity of experts and non-experts in their decision-making related to UOS values.

The results of UOS and NPZ showed that both experts and non-experts can benefit from receiving performance metrics, however, the usefulness of this feedback also relates to the level of difficulty of the provided performance metric. Furthermore, for certain metrics, simply providing a well-defined rubric to non-expert users can demonstrably improve the quality of their feedback.

Within experts with and without performance metrics provided to users (Experiment 1 Intervention 1 and 2), there was no significant difference in Hypervolume Indicator values of the final solutions. This showed that experts' final solutions had similar performance behavior with and without access to performance metrics.

A strategy that we employed to assess the variation in generated massing options by InsightXR was to analyze the final solutions for all users based on building heights. Based on user interviews, there was an agreement among both experts and non-experts in their building height preferences. For example, expert EE1 discussed their massing preferences:

> "*I shied away from taller buildings with low footprints. I preferred the buildings with openings where the pedestrians were walking by, and when the openings on the roof were visible on the street view. That was a more preferable option for me.*"

In addition, non-expert user 2N2 discussed:

> "*I was favoring a form that was more cantilevered that could maximize the floor space around the building. It was shorter than the average massing you saw.*"

In order to investigate the final solutions' heights further, an independent t-test was performed to assess the statistical significance of building height comparisons for experts and non-experts between Intervention 1 and Intervention 2. For both experts and non-experts, there was a significant difference between building heights when users were not provided with performance metrics (Intervention 1) compared to when they were provided with performance metrics (Intervention 2). Overall, for all users, the final solutions tend to be taller when users were provided with performance metrics compared to those without performance metrics. This result can be explained by the importance of providing performance metrics on their decision-making, as taller buildings tend to perform better due to their ability to offer lower AV and higher UOS values and depending on the building depth and width, lower NPZ values. The majority of users did not prefer this typology in Experiment 1 Intervention 1, but once presented with performance metrics, users' final solutions tended to include taller buildings with smaller footprints. Between experts and NSGA-II Alone, the difference in building heights between the Experiment 1 and 2 Intervention 1 was statistically significant. In addition, between experts and NSGA-II Alone based on experts, the difference in building heights between Experiment 1 and 2 Intervention 2 was statistically significant. These results suggest that the final solutions tend to be taller for NSGA-II Alone, and the human aesthetics were at odds with the metric's minimization problem (Experiment 1 Interventions 1 and 2).

5.2 Conclusions

This research aimed to contribute to the meaningful inclusion of participants in the generative design process during early stages of architectural design. This goal was accomplished by designing an AR application called InsightXR that enables designers (experts) and non-designers (non-experts) to participate in a generative design process in real-time, remotely, or in person [13]. InsightXR allows for user participation in design using different methods. A user can define new generative design projects, participate in a generative design project by providing feedback, provide feedback on single massing projects, and view the results of their generative design project.

An important contribution of this research was the introduction of a Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to generate design alternatives inspired by how architects start early massing studies using Operative Design [13]. This algorithm was applied to create an interactive system in AR that allows for user intervention throughout the generative design process. The design operations explored in this research focused on displacement and subtraction

by using expansion and carving operations as the main verbs applied to the proposed operative generative design [13], [148].

The Operative Generative NSGA-II algorithm was able to successfully generate massing solutions for a case study site, which was discussed in detail in chapter 4.1. The first experiment investigated Floor Area Ratio (FAR), Non-Passive Zone (NPZ), Roofs and Best Oriented Surfaces (RBOS), and Usable Open Space (UOS) as objective functions. Among the non-dominated solutions visualized in all experiments, there were hybrid solutions (combinations of carving and expansion) and solutions with only either carving or expansion as the operation. All experiments were able to achieve their minimization/maximization goals. Visualizing high trade-off solutions (Case 1 and 3) and the best trade-off solution (Case 2) in this experiment revealed a similar pattern in operative design across all experiments [19]. Case 1, in most experiments, provided the smallest carvings, leading to the least passive options based on the proposed rules-of-thumb. Case 2, on the other hand, in most experiments, created efficient building volumes closer to how architects would design an efficient courtyard building. Overall, Case 3 (with the lowest FAR value) was always a hybrid solution in all experiments. The study of cases across the solution space is critical in the architectural design process, as they provide valuable insights for decision-making when deciding on the desired final solution. Based on the results of this experiment, the solutions that have the best tradeoff between the objective functions (Case 2) also resulted in meaningful massing options that can be realistically selected and continued to next stages of architectural design. This result shows the value of the proposed algorithm in this research by creating meaningful massing options that can be integrated into the architectural design process.

The second experiment was a continuation of Experiment 1 and introduced the integration of the Operative Generative NSGA-II with InsightXR's AR application using RBOS, UOS, and AV as objective functions. An important contribution of this experiment was the successful integration of a clustering method based on k-means clustering using the Scikit-learn Python module in InsightXR, and the integration of Amazon Web Services (AWS) with InsightXR [176]. The clustering was combined with a shape similarity algorithm that allowed for selecting representative 3D massing solutions to be presented to the users for further investigation, reducing user fatigue that can be caused by the presentation of excessive marginally differing solutions per generation [217], [176]. To understand the capabilities of InsightXR in selecting and visualizing representative design solutions per generation, a case study experiment was conducted using three objective functions: RBOS, UOS, and AV. Each cluster per generation successfully created a set of unique options with respect to the constraints and metrics. At each generation, the three presented options spanned the spectrum of the different metrics, RBOS, AV, and UOS, presenting one at each extreme as well as a hybrid option. This allows a designer to properly choose a design massing that will meet both objective and subjective goals, exploring the space of options across generations of the genetic algorithm. The results of this experiment revealed a fundamental requirement for clustering to inform the massing selection criteria. In Experiment 2, the most representative massing per cluster was selected for visualization. However, in order to receive and apply user ratings across massing options in each cluster, the representative massing should not be the most similar shape to all other clustered solutions, instead the selected solution needs to be located the closest to the average shape similarity of the entire cluster. This finding was applied in the massing selection per cluster in the Experiment 3.

Experiment 3 aimed to assess the inclusion of participants in the generative design process using InsightXR. This experiment showcased InsightXR's ability to integrate user's subjective feedback in the form of user rating combined with quantitative objective functions in an interactive generative design process in AR. The hypothesis was that including non-experts in the early-stage design process could provide valuable feedback for the final design solutions, however, there would be significant differences in the final solutions' performance between expert and non-expert participants. The results of human subject experiments showed that there was no significant difference among experts and non-experts in the final solution's Hypervolume (HV) Index value, a measure of the success of NSGA-II's objective function minimization. For some experts and non-experts, both interventions had similar behavior in HV values over time. For some users, Intervention 2 resulted in lower HV values over time. Some users' HV values in Intervention 2 improved over time compared to their Intervention 1 results. The behavior of HV for both experts and non-experts, compared to the NSGA-II Alone, showed a fundamental conflict between users' feedback, both experts and non-experts, and other objective functions used in this experiment. This conflict shows an important benefit of user inclusion in the generative design process. This was due to the tendency of the algorithm to generate massing options with smaller footprints and taller (tower) typologies, which maximizes the UOS and minimizes the NPZ value. The typology preferred by the NSGA-II algorithm Alone is the typology that most users described as their least favorite building type. Such behavior can also be seen in building performance optimization using Energy simulation tools, where the tools can lead to the smallest possible building to reduce the energy consumption. While often the most efficient, such building typologies might not be in accordance with architect's preferences. The results of examining the HV values demonstrate the success and importance of including participants in the design process, for both experts and non-experts to ensure subjective opinions are properly factored into the design.

In order to receive more useful feedback from non-expert users, performance metrics were introduced in these experiments [245]. A high percentage of both experts and non-experts rated NPZ as the most important objective function, demonstrating a valorization of design aspects that contribute to climate change mitigation. In addition, the NPZ values had a higher impact on experts compared to non-experts, signaling the importance that architects in training currently attribute to these performance criteria. These results within two user groups also highlighted the importance of expertise in the generative design process. While providing complex spatial information to experts can help improve their decision-making process, in contrast, it did not significantly improve the decision-making process for non-experts. It is possible that the non-experts had difficulty conceptualizing NPZ and its impact on the design. Ultimately this was borne out when we saw that while non-experts verbally expressed the importance of NPZ during the interviews, NPZ did not significantly improve when they were provided with the values. UOS was rated as the second highest-rated objective function, and it showed a high impact on experts and non-experts when provided with performance metrics. While the majority of both user groups rated NPZ as the most important metric with UOS as second, both groups' solutions improved in UOS values. This can be due to the clarity of UOS's definition visually on the massing, where users can see the open space easily in each option and can prioritize that in their ratings. These results showed that both experts and non-experts can benefit from receiving performance metrics when providing feedback, however, the usefulness of this feedback also relates to the level of difficulty of the provided performance metric. For certain types of metrics providing the detailed rubric to non-

experts could actually improve their feedback to a level on par with experts, however as seen with NPZ that is not always the case with harder to conceptualize metrics.

As an overall conclusion, while the AI system alone could often generate better-performing solutions in a quantitative-only optimization space, it tended to create solutions that were qualitatively unsatisfactory to many users. The absence of user preference in generative design processes has arguably been a major obstacle for the adoption of generative design in real world situations. Introducing user preference ratings as an additional objective function often generated a relative reduction in some quantitative metrics, but it simultaneously increased the likelihood that solutions were qualitatively appreciated by users, and thus more likely to be adopted. Furthermore, when expert users of our AR-driven participatory design interface were asked to rate alternative design purely based on aesthetic preferences, results were typically inferior to when users were additionally given some simplified quantitative information to inform their rating preferences. This suggests that the successful integration of informed user preference with AI-based support for complex multi-dimensional design tasks may prove a successful and promising avenue for future design methods research.

With InsightXR, we successfully completed the design and implementation of a novel Operative Generative Design Using Non-Dominated Sorting Genetic Algorithm II (NSGA-II) in an AR application directed for collaborative and participatory design. The proposed interactive generative design process was further investigated using human subject experiments, which revealed the benefits of including both experts and non-experts in the design process, while showing the importance of expertise, well-defined rubrics, and access to quantitative metrics in assessing complex performance metrics during the evaluation of solutions in the early massing stages of design.

5.3 Future Work

The InsightXR platform has many levels of further improvement that can be implemented going forward. The proposed methods in this research are directly applicable to guiding computational design in other fields, such as product design, mechanical engineering, and material science, which can be explored further in the future. To support these other fields, this research on immersive environments can be continued by adding interfaces to digital fabrication for future iterations of InsightXR to create novel computational fabrication methods.

The primary challenge faced in the early stages of this research was the dependence on use of the HTC Vive Pro eye in the VR platform, which limited the accessibility as the headset would need to be physically brought to the users. This is a challenge that was mitigated through the focus on designing an AR platform and the use of smartphone apps, which can reach most people. In terms of user interface and user experience improvements of the app, the user experiments performed suggest several possible avenues to further the potential use of AR in interactive generative design. For example, the result visualization feature can be further developed to allow for visualizing multiple user results together. It can also be expanded to allow for exporting final solutions directly from the application to personal devices for more detailed visualization and analysis.

Moving forward, new experiments can be designed with a different initial population and different parameters generated by architects. These experiments can discover different relationships between other sets of objectives and can be used to allow non-experts to help with the generative design workflow of an actively ongoing massing study in InsightXR.

Future experiments may also focus on comparing interactively generated solutions with algorithm-only solutions, in terms of preference by users of the solutions generated by both methods. As an example, during an exploratory follow-up experiment, an expert user was presented with the final solutions based on Experiment 1 Intervention 1 and 2 and final solutions for Experiment 2 Intervention 1 and 2 (NSGA-II Alone generated options) and was asked to rate the final generated massing options. The expert was not presented with any performance metrics to perform this task and was asked to rate the solutions purely from an aesthetic preference perspective. The objective of this experiment was to understand which Experiments and Interventions can generate solutions that will be rated the highest by users once final solutions are generated using an interactive generative design process. When comparing user ratings with the estimated user ratings by the algorithm in Experiment 1, the algorithm was able to successfully estimate the user's ratings in both interventions. For Intervention 1, the user rating's prediction bias had a mean value of 0.8 (on a scale from -3 to 3). For Intervention 2, the user rating's prediction bias had a mean value of 0.7. These results show the proposed algorithm's reliability in predicting user's rating based on shape similarities during generative design, and future research can investigate this finding further. These studies can be further explored as a new set of experiments in the future, where all experts and non-experts view and rate their generated massing options to thoroughly analyze and even tune the performance of the predicted rating by the algorithm. In addition, larger user groups can be used to train machine learning models to understand user massing preferences for specific building programs and site typologies.

While the current research implemented Non-Dominated Sorting Genetic Algorithm II (NSGA-II) soliciting user feedback to drive the generation of design options, InsightXR can seamlessly swap in a wide array of machine learning algorithms or artificial intelligence to generate the 3D design solutions. An example of this future work is using neural networks for semi-supervised learning to improve the generated design solutions based on users' feedback. In addition, using machine learning, expert and non-expert users' preferences can be used to train predictive models to help generate massing options that align more with theoretical user preferences for different use cases and physical properties of the site desired by users.

Bibliography

1. Ilyass Abouelaziz, Aladine Chetouani, Mohammed El Hassouni, and Hocine Cherifi. 2018. A blind mesh visual quality assessment method based on convolutional neural network. *Electronic Imaging* 2018: 423–1. https://doi.org/10.2352/ISSN.2470-1173.2018.18.3DIPM-423
2. Ahmed M. Abuzuraiq and Halil Erhan. 2020. The Many Faces of Similarity - A Visual Analytics Approach for Design Space Simplification. In *D. Holzer, W. Nakapan, A. Globa, I. Koh (eds.), RE: Anthropocene, Design in the Age of Humans - Proceedings of the 25th CAADRIA Conference - Volume 1, Chulalongkorn University, Bangkok, Thailand, 5-6 August 2020, pp. 485-494*. Retrieved July 27, 2021 from http://papers.cumincad.org/cgi-bin/works/paper/caadria2020_443
3. Najood Al Ghamdi and Wadee Alhalabi. 2019. Fixation Detection with Ray-casting in Immersive Virtual Reality. *International Journal of Advanced Computer Science and Applications* 10. https://doi.org/10.14569/IJACSA.2019.0100710
4. American Society of Heating, Refrigerating and Air-Conditioning Engineers. 2022. *Standard 90.1-2022—Energy Standard for Sites and Buildings Except Low-Rise Residential Buildings*. Retrieved March 20, 2023 from https://www.techstreet.com/ashrae/standards/ashrae-90-1-2022-i-p?product_id=2522082
5. David Anderson, Emily Anderson, Neal Lesh, Joe Marks, Brian Mirtich, David Ratajczak, and Kathy Ryall. 2000. Human-Guided Simple Search. *AAAI/IAAI 2000*: 12.
6. John L. Andreassi. 2006. *Psychophysiology: Human Behavior and Physiological Response*. Psychology Press. Retrieved November 20, 2021 from https://www.routledge.com/Psychophysiology-Human-Behavior-and-Physiological-Response/Andreassi/p/book/9780805849516
7. Chairil Andri, Mohammed Hazim Alkawaz, and Safa Riyadh Waheed. 2019. Examining Effectiveness and User Experiences in 3D Mobile based Augmented Reality for MSU Virtual Tour. In *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, 161–167. https://doi.org/10.1109/I2CACIS.2019.8825054
8. Gordon C. Armour and Elwood S. Buffa. 1963. A Heuristic Algorithm and Simulation Approach to Relative Location of Facilities. *Management Science* 9, 2: 294–309.
9. Meghna Babbar-Sebens, Snehasis Mukhopadhyay, Vidya Bhushan Singh, and Adriana Debora Piemonti. 2015. A web-based software tool for participatory optimization of conservation practices in watersheds. *Environmental Modelling & Software* 69: 111–127. https://doi.org/10.1016/j.envsoft.2015.03.011
10. Johannes Bader and Eckart Zitzler. 2011. HypE: an algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation* 19, 1: 45–76. https://doi.org/10.1162/EVCO_a_00009
11. Nan Bai, Wenqia Ye, Jianan Li, Huichao Ding, Meram Pienaru, and Bunschoten Raoul. 2018. Customised Collaborative Urban Design - A Collective User-based Urban Information System through Gaming.
12. Elnaz Bailey and Luisa Caldas. 2023. InsightXR: Operative Generative Design in Augmented Reality. In *Generative Art 2023*.

13. Elnaz Tafrihi Bailey and Luisa Caldas. 2023. Operative generative design using non-dominated sorting genetic algorithm II (NSGA-II). *Automation in Construction* 155: 105026. https://doi.org/10.1016/j.autcon.2023.105026
14. Rohan O'Neil Bailey. 2005. The Digital Design Coach Enhancing Design Conversations in Architectural Education. Retrieved October 4, 2021 from http://researcharchive.vuw.ac.nz/handle/10063/392
15. Amit Banerjee, Juan Quiroz, and Sushil Louis. 2008. A Model of Creative Design Using Collaborative Interactive Genetic Algorithms. . 397–416. https://doi.org/10.1007/978-1-4020-8728-8_21
16. Michael Beham, Wolfgang Herzner, M. Eduard Groller, and Johannes Kehrer. 2014. Cupid: Cluster-Based Exploration of Geometry Generators with Parallel Coordinates and Radial Trees. *IEEE Transactions on Visualization and Computer Graphics* 20, 12: 1693–1702. https://doi.org/10.1109/TVCG.2014.2346626
17. Ammon Beyerle. 2018. Participation in architecture: agonism in practice. The University of Melbourne, Melbourne, Australia. Retrieved September 27, 2021 from https://minerva-access.unimelb.edu.au/handle/11343/225579
18. Prashant Bhat, Iaroslav Shcherbatyi, and Wolfgang Maass. 2019. Automated Learning of User Preferences for Selection of High Quality 3D Designs. *Procedia CIRP* 84: 814–819. https://doi.org/10.1016/j.procir.2019.04.249
19. Julian Blank and Kalyanmoy Deb. 2020. Pymoo: Multi-Objective Optimization in Python. *IEEE Access* 8: 89497–89509. https://doi.org/10.1109/ACCESS.2020.2990567
20. T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. 2017. Visualization of Eye Tracking Data: A Taxonomy and Survey. *Computer Graphics Forum* 36, 8: 260–284. https://doi.org/10.1111/cgf.13079
21. Peter Blundell Jones, Doina Petrescu, and Jeremy Till. 2013. *Architecture and Participation.* Taylor and Francis, Hoboken. Retrieved September 27, 2021 from http://grail.eblib.com.au/patron/FullRecord.aspx?p=214723
22. Samuel Brody, David Godschalk, and Raymond Burby. 2003. Mandating Citizen Participation in Plan Making: Six Strategic Planning Choices. *Journal of The American Planning Association - J AMER PLANN ASSN* 69: 245–264. https://doi.org/10.1080/01944360308978018
23. Emily Brown and Paul Cairns. 2004. A grounded investigation of game immersion. In *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*, 1297. https://doi.org/10.1145/985921.986048
24. Nathan C Brown and Caitlin T Mueller. 2017. Automated performance-based design space simplification for parametric structural design. 10.
25. Nathan; Mueller Brown. 2017. Designing With Data: Moving Beyond The Design Space Catalog. In *ACADIA 2017: DISCIPLINES & DISRUPTION [Proceedings of the 37th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA) ISBN 978-0-692-96506-1] Cambridge, MA 2-4 November, 2017), pp. 154-163*. Retrieved July 29, 2021 from http://papers.cumincad.org/cgi-bin/works/paper/acadia17_154

26. A. Di Bucchianico. 1999. Combinatorics, computer algebra and the Wilcoxon-Mann-Whitney test. *Journal of Statistical Planning and Inference* 79, 2: 349–364. https://doi.org/10.1016/S0378-3758(98)00261-4
27. Inês Caetano, Luís Santos, and António Leitão. 2020. Computational design in architecture: Defining parametric, generative, and algorithmic design. *Frontiers of Architectural Research* 9, 2: 287–300. https://doi.org/10.1016/j.foar.2019.12.008
28. L. G. Caldas and L. K. Norford. 1999. A Genetic Algorithm Tool for Design Optimization. In *Proceedings of ACADIA '99*, 260–271. Retrieved November 8, 2021 from https://www.semanticscholar.org/paper/A-Genetic-Algorithm-Tool-for-Design-Optimization-Caldas/b40aabb93d9d40337f925588c88486b98a450119
29. L. G. Caldas and L. K. Norford. 2003. Genetic Algorithms for Optimization of Building Envelopes and the Design and Control of HVAC Systems. *Journal of Solar Energy Engineering* 125, 3: 343–351. https://doi.org/10.1115/1.1591803
30. Luisa Caldas. 2008. Generation of energy-efficient architecture solutions applying GENE_ARCH: An evolution-based generative design system. *Advanced Engineering Informatics* 22, 1: 59–70. https://doi.org/10.1016/j.aei.2007.08.012
31. Luisa Gama Caldas and Leslie K Norford. 2002. A design optimization tool based on a genetic algorithm. *Automation in Construction* 11, 2: 173–184. https://doi.org/10.1016/S0926-5805(00)00096-0
32. Luisa Caldas and Luis Santos. 2016. Painting with light: An interactive evolutionary system for daylighting design. *Building and Environment* 109. https://doi.org/10.1016/j.buildenv.2016.07.023
33. Victor Calixto and Gabriela Celani. 2015. A literature review for space planning optimization using an evolutionary algorithm approach: 1992-2014. In *Anais do XIX Congresso da Sociedade Ibero-americana de Gráfica Digital 2015*, 662–671. https://doi.org/10.5151/despro-sigradi2015-110166
34. Patricia E. Camporeale and Pilar Mercader-Moyano. 2019. Towards nearly Zero Energy Buildings: Shape optimization of typical housing typologies in Ibero-American temperate climate cities from a holistic perspective. *Solar Energy* 193: 738–765. https://doi.org/10.1016/j.solener.2019.09.091
35. Berk Cebeci, Ufuk Celikcan, and Tolga K. Capin. 2019. A comprehensive study of the affective and physiological responses induced by dynamic virtual reality environments. *Computer Animation and Virtual Worlds* 30, 3–4: e1893. https://doi.org/10.1002/cav.1893
36. Tommy K. H. Chan, Christy M. K. Cheung, and Zach W. Y. Lee. 2017. The state of online impulse-buying research: A literature analysis. *Information & Management* 54, 2: 204–217. https://doi.org/10.1016/j.im.2016.06.001
37. Andre Chaszar, Peter Von Buelow, and Michela Turrin. 2016. Multivariate Interactive Visualization of Data in Generative Design.
38. Monchu Chen and Veranika Lim. 2013. Tracking Eyes in Service Prototyping. https://doi.org/10.1007/978-3-642-40498-6_19
39. Hung-Lin Chi, Shih-Chung Kang, and Xiangyu Wang. 2013. Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. *Automation in Construction* 33: 116–122. https://doi.org/10.1016/j.autcon.2012.12.017

40. James Childs-Maidment. Pyrebase: A simple python wrapper for the Firebase API. Retrieved September 29, 2021 from https://github.com/thisbejim/Pyrebase
41. Michael Chu, Jane Matthews, and Peter E. D. Love. 2018. Integrating mobile Building Information Modelling and Augmented Reality systems: An experimental study. *Automation in Construction* 85: 305–316. https://doi.org/10.1016/j.autcon.2017.10.032
42. Adriana Ciardiello, Federica Rosso, Jacopo Dell'Olmo, Virgilio Ciancio, Marco Ferrero, and Ferdinando Salata. 2020. Multi-objective approach to the optimization of shape and envelope in building energy design. *Applied Energy* 280: 115984. https://doi.org/10.1016/j.apenergy.2020.115984
43. Paul Coates, Terence Broughton, and Helen Jackson. 1999. Exploring Three-dimensional design worlds using Lindenmeyer systems and Genetic Programming. In P. Bentley (ed.). : San Francisco, 323–341. Retrieved October 9, 2021 from http://www.cs.ucl.ac.uk/staff/P.Bentley/evdes.html
44. Carlos A. Coello Coello. 2000. Treating constraints as objectives for single-objective evolutionary optimization. *Engineering Optimization* 32, 3: 275–308. https://doi.org/10.1080/03052150008941301
45. C.A. Coello Coello and G.T. Pulido. 2005. Multiobjective structural optimization using a microgenetic algorithm. *Structural and Multidisciplinary Optimization* 30, 5: 388–403. https://doi.org/10.1007/s00158-005-0527-z
46. Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, New York. https://doi.org/10.4324/9780203771587
47. Mary C. Comerio. 1987. Design and Empowerment: 20 Years of Community Architecture. *Built Environment (1978-)* 13, 1: 15–28.
48. M. Corsini, M. C. Larabi, G. Lavoué, O. Petřík, L. Váša, and K. Wang. 2013. Perceptual Metrics for Static and Dynamic Triangle Meshes. *Computer Graphics Forum* 32, 1: 101–125. https://doi.org/10.1111/cgf.12001
49. Nigel Cross (ed.). 1971. Design Participation. In *Proceedings of the Design Research Society Conference*, 131.
50. Nigel Cross. 2001. Can a Machine Design? *Design Issues* 17, 4: 44–50. https://doi.org/10.1162/07479360152681083
51. Elie Daher, Sylvain Kubicki, and Burak Pak. 2018. Participation-based Parametric Design in Early Stages - A participative design process for spatial planning in office building. In *The 36th eCAADe Conference*, 10.
52. Yuze Dan, Zhenjiang Shen, Yumin Lin, and Zhu Yiyun. 2021. The Development of Design Support System for Public Participation of Community Public Space Design Using Mixed Reality. In *Stojakovic, V and Tepavcevic, B (eds.), Towards a new, configurable architecture - Proceedings of the 39th eCAADe Conference - Volume 1, University of Novi Sad, Novi Sad, Serbia, 8-10 September 2021, pp. 205-212*. Retrieved September 28, 2021 from http://papers.cumincad.org/cgi-bin/works/2015%20+dave=2:/Show?ecaade2021_240
53. Renaud Danhaive and Caitlin Mueller. 2015. Combining parametric modeling and interactive optimization for high-performance and creative structural design.
54. Juan Manuel Davila Delgado, Lukumon Oyedele, Peter Demian, and Thomas Beach. 2020. A research agenda for augmented and virtual reality in architecture, engineering and

construction. *Advanced Engineering Informatics* 45: 101122. https://doi.org/10.1016/j.aei.2020.101122
55. Richard Dawkins. 2015. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design*. W. W. Norton & Company, New York.
56. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 2: 182–197. https://doi.org/10.1109/4235.996017
57. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 2: 182–197. https://doi.org/10.1109/4235.996017
58. Kalyanmoy Deb. 2011. Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction. In *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*, Lihui Wang, Amos H. C. Ng and Kalyanmoy Deb (eds.). Springer, London, 3–34. https://doi.org/10.1007/978-0-85729-652-8_1
59. Kalyanmoy Deb. Multi-Objective Optimization Using Evolutionary Algorithms: An Introduction.
60. Kalyanmoy Deb and Samir Agrawal. 1999. A Niched-Penalty Approach for Constraint Handling in Genetic Algorithms. In *Artificial Neural Nets and Genetic Algorithms*. Springer Vienna, Vienna, 235–243. https://doi.org/10.1007/978-3-7091-6384-9_40
61. Kalyanmoy Deb and Debayan Deb. 2014. Analysing mutation schemes for real-parameter genetic algorithms. *International Journal of Artificial Intelligence and Soft Computing* 4, 1: 1–28. https://doi.org/10.1504/IJAISC.2014.059280
62. Kalyanmoy Deb, Karthik Sindhya, and Tatsuya Okabe. 2007. Self-adaptive simulated binary crossover for real-parameter optimization. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation* (GECCO '07), 1187–1194. https://doi.org/10.1145/1276958.1277190
63. Xavier Desurmont, Isabel Martinez-Ponte, Jerome Meessen, and Jean-francois Delaigle. 2006. Nonintrusive viewpoint tracking for 3D for perception in smart video conference. In *Three-Dimensional Image Capture and Applications VII*, 106–117. https://doi.org/10.1117/12.642924
64. Gabriel Diaz, Joseph Cooper, Dmitry Kit, and Mary Hayhoe. 2013. Real-time recording and classification of eye movements in an immersive virtual environment. *Journal of Vision* 13, 12: 5. https://doi.org/10.1167/13.12.5
65. Alan Díaz-Manríquez, Gregorio Toscano, Jose Hugo Barron-Zambrano, and Edgar Tello-Leal. 2016. A Review of Surrogate Assisted Multiobjective Evolutionary Algorithms. *Computational Intelligence and Neuroscience* 2016: e9420460. https://doi.org/10.1155/2016/9420460
66. Anhai Doan, Raghu Ramakrishnan, and Alon Y. Halevy. 2011. Crowdsourcing systems on the World-Wide Web. *Communications of the ACM* 54, 4: 86–96. https://doi.org/10.1145/1924421.1924442
67. Dirk Donath and Holger Regenbrecht. 1996. Using Virtual Reality Aided Design Techniques for Three-dimensional Architectural Sketching. In *Design Computation: Collaboration, Reasoning, Pedagogy [ACADIA Conference Proceedings / ISBN 1-880250-05-5] Tucson (Arizona / USA) October 31 - November 2, 1996, pp. 199-212*. Retrieved

October 11, 2021 from http://papers.cumincad.org/cgi-bin/works/2015%20+dave=2:/Show?656d
68. Suyang Dong, Amir Behzadan, Chen Feng, and Vineet Kamat. 2013. Collaborative visualization of engineering processes using tabletop augmented reality. *Advances in Engineering Software* 55: 45–55. https://doi.org/10.1016/j.advengsoft.2012.09.001
69. Runliang Dou, Chao Zong, and Guofang Nan. 2016. Multi-stage interactive genetic algorithm for collaborative product customization. *Knowledge-Based Systems* 92: 43–54. https://doi.org/10.1016/j.knosys.2015.10.013
70. Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. 2012. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (CSCW '12), 1013–1022. https://doi.org/10.1145/2145204.2145355
71. Jing Du, Zhengbo Zou, Yangming Shi, and Dong Zhao. 2018. Zero latency: Real-time synchronization of BIM data in virtual reality for collaborative decision-making. *Automation in Construction* 85: 51–64. https://doi.org/10.1016/j.autcon.2017.10.009
72. José P. Duarte. 2005. A discursive grammar for customizing mass housing: the case of Siza's houses at Malagueira. *Automation in Construction* 14, 2: 265–275. https://doi.org/10.1016/j.autcon.2004.07.013
73. José P. Duarte, Gonçalo Ducla-Soares, Luisa G. Caldas, and João Rocha. 2006. An Urban Grammar For The Medina OF Marrakech. In *Design Computing and Cognition '06*, 483–502. https://doi.org/10.1007/978-1-4020-5131-9_25
74. A. Duchowski. 2003. Eye Tracking Methodology - Theory and Practice, Third Edition. https://doi.org/10.1007/978-3-319-57883-5
75. Phillip Dunston, Laura Arns, James Mcglothlin, Gregory Lasker, and Adam Kushner. 2011. An Immersive Virtual Reality Mock-Up for Design Review of Hospital Patient Rooms. . 167–176. https://doi.org/10.1007/978-94-007-0605-7_15
76. Charles M. Eastman. 1975. *Spatial Synthesis in Computer-Aided Building Design*. Applied Science Publishers. Retrieved October 11, 2021 from https://dl.acm.org/doi/abs/10.5555/540292
77. Maria K. Eckstein, Belén Guerra-Carrillo, Alison T. Miller Singley, and Silvia A. Bunge. 2017. Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience* 25: 69–91. https://doi.org/10.1016/j.dcn.2016.11.001
78. Eve A Edelstein, Klaus Gramann, Jurgen Schulze, Nima Bigdely Shamlo, Elke van Erp, Andrey Vankov, Scott Makeig, Laura Wolszon, and Eduardo Macagno. Neural Responses during Navigation in the Virtual Aided Design Laboratory: Brain Dynamics of Orientation in Architecturally Ambiguous Space. 7.
79. Mohamed Elgendi. 2012. On the Analysis of Fingertip Photoplethysmogram Signals. *Current Cardiology Reviews* 8, 1: 14–25. https://doi.org/10.2174/157340312801215782
80. Semiha Ergan, Ahmed Radwan, Zhengbo Zou, Hua-an Tseng, and Xue Han. 2018. Quantifying Human Experience in Architectural Spaces with Integrated Virtual Reality and Body Sensor Networks. *Journal of Computing in Civil Engineering* 33. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000812

81. Gabriel Evans, Jack Miller, Mariangely Iglesias Pena, Anastacia MacAllister, and Eliot Winer. 2017. Evaluating the Microsoft HoloLens through an augmented reality assembly application. 101970V. https://doi.org/10.1117/12.2262626
82. Ralph Evins. 2013. A review of computational optimisation methods applied to sustainable building design. *Renewable and Sustainable Energy Reviews* 22: 230–245. https://doi.org/10.1016/j.rser.2013.02.004
83. Ralph Evins, Philip Pointer, and Ravi Vaidyanathan. 2011. Multi-objective optimisation of the configuration and control of a double-skin facade.
84. C.M. Fonseca, L. Paquete, and M. Lopez-Ibanez. 2006. An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator. In *2006 IEEE International Conference on Evolutionary Computation*, 1157–1163. https://doi.org/10.1109/CEC.2006.1688440
85. Gerald Franz, Markus von der Heyde, and Heinrich H. Bülthoff. 2005. An empirical approach to the experience of architectural space in virtual reality—exploring relations between features and affective appraisals of rectangular indoor spaces. *Automation in Construction* 14, 2: 165–172. https://doi.org/10.1016/j.autcon.2004.07.009
86. John Frazer. 1995. *An Evolutionary Architecture*. Architectural Association Publications, London.
87. Laura Freina and Michela Ott. 2015. A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives.
88. Jean-Philippe Garric. 2016. The French Beaux-Arts. In *Companion to the History of Architecture*, Harry Francis Mallgrave (ed.). John Wiley & Sons, Inc., Hoboken, NJ, USA, 1–15. https://doi.org/10.1002/9781118887226.wbcha080
89. Fred W. Glover and Gary A. Kochenberger (eds.). 2003. *Handbook of Metaheuristics*. Springer US. https://doi.org/10.1007/b101874
90. David E. Goldberg. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., USA.
91. D. Gong, X. Sun, and J. Ren. 2009. Surrogate Models Based on Individual's Interval Fitness in Interactive Genetic Algorithms. *Chinese Journal of Electronics* 18: 689–694.
92. Dun-wei Gong, Xiaoyan Sun, and Jie Yuan. 2009. Interactive Genetic Algorithms with Individual's Uncertain Fitness. https://doi.org/10.5772/9605
93. Alexander Grasser. 2019. Towards an Architecture of Collaborative Objects. In *Sousa, JP, Xavier, JP and Castro Henriques, G (eds.), Architecture in the Age of the 4th Industrial Revolution - Proceedings of the 37th eCAADe and 23rd SIGraDi Conference - Volume 1, University of Porto, Porto, Portugal, 11-13 September 2019, pp. 325-332*. Retrieved November 28, 2021 from http://papers.cumincad.org/cgi-bin/works/paper/ecaadesigradi2019_218
94. Scott A. Green, Mark Billinghurst, XiaoQi Chen, and J. Geoffrey Chase. 2008. Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design. *International Journal of Advanced Robotic Systems* 5, 1: 1. https://doi.org/10.5772/5664
95. Fu Guo, Mingming Li, Mingcai Hu, Fengxiang Li, and Bozhao Lin. 2019. Distinguishing and quantifying the visual aesthetics of a product: An integrated approach of eye-tracking and EEG. *International Journal of Industrial Ergonomics* 71: 47–56. https://doi.org/10.1016/j.ergon.2019.02.006

96. Zhe Guo, Zihuan Zhang, Zao Li, Yi Hu, Yuandi Qian, Nengming Cheng, and Philip F. Yuan. 2024. Brain-computer interface based generative design framework: an empirical multi-domain application exploration based on human-factors and form-generation interactive mechanisms. *Architectural Intelligence* 3, 1: 3. https://doi.org/10.1007/s44223-024-00047-2
97. Kim Halskov and Nicolai Brodersen Hansen. 2015. The diversity of participatory design research practice at PDC 2002–2012. *International Journal of Human-Computer Studies* 74: 81–92. https://doi.org/10.1016/j.ijhcs.2014.09.003
98. John Harding and Cecilie Brandt-Olsen. 2018. Biomorpher: Interactive evolution for parametric design. *International Journal of Architectural Computing* 16, 2: 144–163. https://doi.org/10.1177/1478077118778579
99. Dominic Hawton, Ben Cooper-Wooley, Jorke Odolphi, Ben Doherty, Alessandra Fabbri, Nicole Gardner, and M Hank Haeusler. 2018. Shared Immersive Environments For Parametric Model Manipulation. In *the 23rd International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2018*, 10.
100. Oliver Heckmann. 2020. User-driven Parcellation of High-rise Units for Future Urban Habitation - Participatory Computational Design Tools for Future Urban Habitation. In *D. Holzer, W. Nakapan, A. Globa, I. Koh (eds.), RE: Anthropocene, Design in the Age of Humans - Proceedings of the 25th CAADRIA Conference - Volume 1, Chulalongkorn University, Bangkok, Thailand, 5-6 August 2020, pp. 751-760*. Retrieved April 16, 2024 from https://papers.cumincad.org/cgi-bin/works/paper/caadria2020_369
101. Bruno Herbelin, Helena Grillon, Pablo de Heras Ciechomski, and Daniel Thalmann. 2007. Coding gaze tracking data with chromatic gradients for VR Exposure Therapy. 7–14. https://doi.org/10.1109/ICAT.2007.28
102. Károly Hercegfi, Anita Komlódi, Bálint Szabó, Máté Köles, Emma Lógó, Balázs P. Hámornik, and Gyöngyi Rózsa. 2015. Experiences of virtual desktop collaboration experiments. In *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 375–379. https://doi.org/10.1109/CogInfoCom.2015.7390622
103. Anders Hermund, Lars Simon Klint, and Ture Slot Bundgaard. 2019. Neurological feedback from perception of architecture in virtual reality: a methodology study. In *53rd International Conference of the Architectural Science Association (ANZAScA)*, 10.
104. Christiane Herr and Thomas Fischer. 2001. Teaching Generative Design. *International Conference on Generative Art*.
105. Christiane M. Herr and Thomas Kvan. 2007. Adapting cellular automata to support the architectural design process. *Automation in Construction* 16, 1: 61–69. https://doi.org/10.1016/j.autcon.2005.10.005
106. Sarah Hirschman. 2021. Integrated Studio Design Process.
107. John H. Holland. 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. A Bradford Book, Cambridge, MA, USA.
108. Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. 2011. Eye Tracking: A Comprehensive Guide To Methods And Measures.

109. Mokter Hossain and Ilkka Kauranen. 2015. Crowdsourcing: a comprehensive literature review. *Strategic Outsourcing: An International Journal* 8, 1: 2–22. https://doi.org/10.1108/SO-12-2014-0029
110. Jeff Howe. 2006. The Rise of Crowdsourcing. *Wired Magazine*, 14: 5.
111. Ting-Wei Hsu, Ming-Han Tsai, Sabarish V. Babu, Pei-Hsien Hsu, Hsuan-Ming Chang, Wen-Chieh Lin, and Jung-Hong Chuang. 2020. Design and Initial Evaluation of a VR based Immersive and Interactive Architectural Design Discussion System. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 363–371. https://doi.org/10.1109/VR46266.2020.00056
112. Rupert Hurley, Joshua Galvarino, Emily Thackston, Andrew Ouzts, and Andy Pham. 2013. The Effect of Modifying Structure to Display Product Versus Graphical Representation on Packaging. *Packaging Technology and Science* 26. https://doi.org/10.1002/pts.1996
113. Wijnand A. IJsselsteijn, Huib de Ridder, Jonathan Freeman, and Steve E. Avons. 2000. Presence: concept, determinants, and measurement. In *Human Vision and Electronic Imaging V*, 520–529. https://doi.org/10.1117/12.387188
114. C. Ikehara and M. Crosby. 2005. Assessing Cognitive Load with Physiological Sensors. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. https://doi.org/10.1109/HICSS.2005.103
115. Bret Jackson and Daniel F. Keefe. 2016. Lift-Off: Using Reference Imagery and Freehand Sketching to Create 3D Models in VR. *IEEE Transactions on Visualization and Computer Graphics* 22, 4: 1442–1451. https://doi.org/10.1109/TVCG.2016.2518099
116. Patrick Hubert Theodoor Janssen. 2004. A design method and computational architecture for generating and evolving building designs. School of Design, The Hong Kong Polytechnic University.
117. Paul Jenkins and Leslie Forsyth (eds.). 2009. *Architecture, Participation and Society*. Taylor and Francis. https://doi.org/10.4324/9780203869499
118. Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies* 66, 9: 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004
119. Joel Jordan and M. Slater. 2009. An Analysis of Eye Scanpath Entropy in a Progressively Forming Virtual Environment. *PRESENCE: Teleoperators and Virtual Environments*. https://doi.org/10.1162/pres.18.3.185
120. Mohammad Keshavarzi. 2022. Contextual Spatial Computing: A Generative Approach. UC Berkeley. Retrieved April 23, 2024 from https://escholarship.org/uc/item/68q581bt
121. Mohammad Keshavarzi, Ardavan Bidgoli, and Hans Kellner. 2020. V-Dream: Immersive Exploration of Generative Design Solution Space. In *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence* (Lecture Notes in Computer Science), 477–494. https://doi.org/10.1007/978-3-030-60117-1_35
122. Mohammad Keshavarzi, Luisa Caldas, and Luis Santos. 2019. RadVR: A 6DOF Virtual Reality Daylighting Analysis Tool. *arXiv:1907.01652 [cs]*. Retrieved from http://arxiv.org/abs/1907.01652
123. Mohammad Keshavarzi, Allen Y. Yang, Woojin Ko, and Luisa Caldas. 2020. Optimization and Manipulation of Contextual Mutual Spaces for Multi-User Virtual and Augmented

Reality Interaction. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 353–362. https://doi.org/10.1109/VR46266.2020.00055
124. Farshad Kheiri. 2018. A review on optimization methods applied in energy-efficient building geometry and envelope design. *Renewable and Sustainable Energy Reviews* 92: 897–920. https://doi.org/10.1016/j.rser.2018.04.080
125. Dongyun Kim, George Guida, and Jose Luis García Del Castillo Y López. 2022. PlacemakingAI : Participatory Urban Design with Generative Adversarial Networks. 485–494. https://doi.org/10.52842/conf.caadria.2022.2.485
126. Hee-Su Kim and Sung-Bae Cho. 2001. An efficient genetic algorithm with less fitness evaluation by clustering. In *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, 887–894 vol. 2. https://doi.org/10.1109/CEC.2001.934284
127. Kangsoo Kim, Byung-Kuk Seo, Jae-Hyek Han, and Jong-Il Park. 2009. Augmented reality tour system for immersive experience of cultural heritage. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry* (VRCAI '09), 323–324. https://doi.org/10.1145/1670252.1670325
128. Geoff Kimm and Mark Burry. 2020. Encouraging Community Participation in Design Decision-Making Through Reactive Scripting. In *The 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA)*, 10.
129. Katja Knecht, Dimitrie Stefanescu, and Reinhard Koenig. 2019. Citizen Engagement through Design Space Exploration: Integrating citizen knowledge and expert design in computational urban planning. https://doi.org/10.5151/proceedings-ecaadesigradi2019_195
130. Máté Köles, Balázs Hámornik, Lógó Emma, Karoly Hercegfi, and Sarolta Tóvölgyi. 2015. Experiences of a combined psychophysiology and eye-tracking study in VR. *5th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2014 - Proceedings*: 391–395. https://doi.org/10.1109/CogInfoCom.2014.7020486
131. Arthur Kornblut and Lester Wertheimer. 2008. *Construction Documents & Services*. Dearborn Financial Publishing, Inc. Retrieved November 14, 2021 from https://www.buildersbook.com/construction-documents-services-by-kaplan-aec-education.html
132. Florian Kossak, Doina Petrescu, and Tatjana Schneider (eds.). 2010. *Agency: working with uncertain architectures*. Routledge. Retrieved October 11, 2021 from https://www.taylorfrancis.com/books/9780203860298
133. Panos Kourouthanassis, Costas Boletsis, Cleopatra Bardaki, and Dimitra Chasanidou. 2015. Tourists responses to mobile augmented reality travel guides: The role of emotions on adoption behavior. *Pervasive and Mobile Computing* 18: 71–87. https://doi.org/10.1016/j.pmcj.2014.08.009
134. M. Krejcirik. 1987. Computer-aided plant layout. *Computer-Aided Design* 2, 1: 7–19. https://doi.org/10.1016/S0010-4485(69)80028-X
135. Bess Krietemeyer, Amber Bartosh, and L. Covington. 2017. Shared Realities: A Method for Adaptive Design Incorporating Real-Time user Feedback using Virtual Reality and 3D Depth-Sensing Systems. *Proceedings of the Association for Computer Aided Design in Architecture (ACADIA)*: 330–339.

136. Antje Kunze, Jan Halatsch, Carlos Vanegas, Martina Maldaner, Benamy Turkienicz, and Gerhard Schmitt. 2011. A Conceptual Participatory Design Framework for Urban Planning. 895–903.
137. Thomas Kvan. 2001. The pedagogy of virtual design studios. *Automation in Construction* 10, 3: 345–353. https://doi.org/10.1016/S0926-5805(00)00051-0
138. Krystian Kwiecinski, Jacek Markusiewicz, and Agata Pasternak. 2017. Participatory Design Supported with Design System and Augmented Reality.
139. Krystian Kwieciński and Jan Słyk. 2023. Interactive generative system supporting participatory house design. *Automation in Construction* 145: 104665. https://doi.org/10.1016/j.autcon.2022.104665
140. Fan Li, Ching-Hung Lee, Shanshan Feng, Amy Trappey, and Fazal Gilani. 2021. Prospective on Eye-Tracking-based Studies in Immersive Virtual Reality. In *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 861–866. https://doi.org/10.1109/CSCWD49262.2021.9437692
141. Robin S Liggett. 2000. Automated facilities layout: past, present and future. *Automation in Construction* 9, 2: 197–215. https://doi.org/10.1016/S0926-5805(99)00005-9
142. Danny Lobos and Dirk Donath. 2008. Massing Study Support. A New Tool for Early Stages of Architectural Design.
143. Rachael Luck. 2018. What is it that makes participation in design participatory design? *Design Studies* 59: 1–8. https://doi.org/10.1016/j.destud.2018.10.002
144. David G. Luenberger and Yinyu Ye. 2008. *Linear and nonlinear programming*. Springer, New York, NY.
145. Kurt Luther, Amy Pavel, Wei Wu, Jari-lee Tolentino, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2014. CrowdCrit: crowdsourcing and aggregating visual design critique. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing* (CSCW Companion '14), 21–24. https://doi.org/10.1145/2556420.2556788
146. M.L. Maher and J. Poon. 1995. Co-evolution of the fitness function and design solution for design exploration. In *Proceedings of 1995 IEEE International Conference on Evolutionary Computation*, 240-. https://doi.org/10.1109/ICEC.1995.489152
147. Henry Mainsah and Andrew Morrison. Social media, design and civic engagement by youth: a cultural view. 9.
148. Anthony di Mari and Nora Yoo. 2013. *Operative Design: A Catalog of Spatial Verbs*. Laurence King Publishing.
149. Javier Marín-Morales, Carmen Llinares, Jaime Guixeres, and Mariano Alcañiz. 2020. Emotion Recognition in Immersive Virtual Reality: From Statistics to Affective Computing. *Sensors (Basel, Switzerland)* 20, 18: 5163. https://doi.org/10.3390/s20185163
150. R. Marler and Jasbir Arora. 2004. Survey of Multi-Objective Optimization Methods for Engineering. *Structural and Multidisciplinary Optimization* 26: 369–395. https://doi.org/10.1007/s00158-003-0368-6
151. Mariela Martinez, Giovanny Arbelaez Garces, Laurent Dupont, Anaëlle Hily, Mauricio Camargo, Christel Jacob, and Jérôme Dinet. 2020. Physiological assessment of User eXprience supported by Immersive Environments: First input from a literature review. In *ConVRgence (VRIC) Virtual Reality International Conference* (Proceedings of Virtual

Reality International Conference (VRIC)), 89–108. Retrieved October 5, 2021 from https://hal.archives-ouvertes.fr/hal-02551341
152. Jon McCormack, Alan Dorin, and Troy Innocent. 2004. Generative design: a paradigm for design research. In *Futureground - DRS International Conference 2004*, 8.
153. Georgina McNee and Dorina Pojani. 2022. NIMBYism as a barrier to housing and social mix in San Francisco. *Journal of Housing and the Built Environment* 37, 1: 553–573. https://doi.org/10.1007/s10901-021-09857-6
154. Markus Miessen and Shumon Basar (eds.). 2006. *Did Someone Say Participate?: An Atlas of Spatial Practice*. MIT Press, Cambridge, MA, USA.
155. W J Mitchell. 1975. The Theoretical Foundation of Computer-Aided Architectural Design. *Environment and Planning B: Planning and Design* 2, 2: 127–150. https://doi.org/10.1068/b020127
156. William J. Mitchell. 1977. Computer-Aided Architectural Design. In *573 p. : ill. New York: Mason Charter Pub. Inc., 1977. includes bibliography: p. 515-559 and index*. Retrieved November 8, 2021 from http://cumincad.scix.net/cgi-bin/works/Show?0d3a
157. William J. Mitchell. 1994. Three paradigms for computer-aided design. *Automation in Construction* 3, 2–3: 239–245. https://doi.org/10.1016/0926-5805(94)90023-X
158. Liraz Mor, Chao Liu, and Sebastian von Mammen. 2011. Interactive Biomimetic Space: An Interactive Installation to Explore Living Architecture. In *Applications of Evolutionary Computation* (Lecture Notes in Computer Science), 364–373. https://doi.org/10.1007/978-3-642-20520-0_37
159. Ali Motamedi, Zhe Wang, Nobuyoshi Yabuki, Tomohiro Fukuda, and Takashi Michikawa. 2017. Signage visibility analysis and optimization system using BIM-enabled virtual reality (VR) environments. *Advanced Engineering Informatics* 32: 248–262. https://doi.org/10.1016/j.aei.2017.03.005
160. Jana Annina Müller, Dorothea Wendt, Birger Kollmeier, and Thomas Brand. 2016. Comparing Eye Tracking with Electrooculography for Measuring Individual Sentence Comprehension Duration. *PLoS ONE* 11, 10: e0164627. https://doi.org/10.1371/journal.pone.0164627
161. Michael Muller and Sarah Kuhn. 1993. Participatory Design. *Communications of the ACM* 36: 24–28. https://doi.org/10.1145/153571.255960
162. Norman Murray, Dave Roberts, Anthony Steed, Paul Sharkey, Paul Dickerson, John Rae, and Robin Wolff. 2009. Eye gaze in virtual environments: evaluating the need and initial work on implementation. *Concurrency and Computation: Practice and Experience* 21, 11: 1437–1449. https://doi.org/10.1002/cpe.1396
163. Ivan Mutis and Abhijeet Ambekar. 2020. Challenges and enablers of augmented reality technology for in situ walkthrough applications. *Journal of Information Technology in Construction (ITcon)* 25, 3: 55–71. https://doi.org/10.36680/j.itcon.2020.003
164. Danil Nagy, Damon Lau, John Locke, James Stoddart, Lorenzo Villaggi, Ray Wang, Dale Zhao, and David Benjamin. 2017. *Project Discover: An Application of Generative Design for Architectural Space Planning*. https://doi.org/10.22360/simaud.2017.simaud.007
165. Aki Namioka and Douglas Schuler. 1990. Participatory Design Conference Proceedings PDC' 90. 16.

166. Worawan Natephra, A. Motamedi, T. Fukuda, and N. Yabuki. 2017. Integrating building information modeling and virtual reality development engines for building indoor lighting design. https://doi.org/10.1186/S40327-017-0058-X
167. A. Y. C. Nee, S. K. Ong, G. Chryssolouris, and D. Mourtzis. 2012. Augmented reality applications in design and manufacturing. *CIRP Annals* 61, 2: 657–679. https://doi.org/10.1016/j.cirp.2012.05.010
168. Scott Davidson created this Ning Network. Grasshopper. Retrieved from https://www.grasshopper3d.com/
169. Anh-Tuan Nguyen, Sigrid Reiter, and Philippe Rigo. 2014. A review on simulation-based optimization methods applied to building performance analysis. *Applied Energy* 113: 1043–1058. https://doi.org/10.1016/j.apenergy.2013.08.061
170. Mojtaba Noghabaei, Arsalan Heydarian, Vahid Balali, and Kevin Han. 2020. Trend Analysis on Adoption of Virtual and Augmented Reality in the Architecture, Engineering, and Construction Industry. *Data* 5, 1: 26. https://doi.org/10.3390/data5010026
171. Ariel Noyman, Yasushi Sakai, and Kent Larson. 2019. CityScopeAR: Urban Design and Crowdsourced Engagement Platform. Retrieved April 23, 2024 from http://arxiv.org/abs/1907.08586
172. Yeonjoo Oh, Suguru Ishizaki, Mark D. Gross, and Ellen Yi-Luen Do. 2013. A theoretical framework of design critiquing in architecture studios. *Design Studies* 34, 3: 302–325. https://doi.org/10.1016/j.destud.2012.08.004
173. Iuliia Osintseva, Reinhard Koenig, Andreas Berst, Martin Bielik, and Sven Schneider. Automated Parametric Building Volume Generation: A Case Study for Urban Blocks. In *SimAUD 2020*, 8.
174. José Antonio Parejo, Antonio Ruiz-Cortés, Sebastián Lozano, and Pablo Fernandez. 2012. Metaheuristic optimization frameworks: a survey and benchmarking. *Soft Computing* 16, 3: 527–561. https://doi.org/10.1007/s00500-011-0754-8
175. Gitte Pedersen and Koumaditis Konstantinos. 2020. Virtual Reality (VR) in the Computer Supported Cooperative Work (CSCW) Domain: A Mapping and a Pre-study on Functionality and Immersion. . 136–153. https://doi.org/10.1007/978-3-030-49698-2_10
176. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 85: 2825–2830.
177. Thies Pfeiffer. 2012. 3D Attention Volumes for usability studies in virtual reality. In *2012 IEEE Virtual Reality Workshops (VRW)*, 117–118. https://doi.org/10.1109/VR.2012.6180910
178. Adriana Debora Piemonti. Interactive Genetic Algorithms for Watershed Planning: An investigation of Usability and Human-centered Design. 221.
179. Ben Polly, Shanti Pless, Sammy Houssainy, Paul Torcellini, William Livingood, Sarah Zaleski, Matt Jungclaus, Tom Hootman, and Mindy Craig. 2020. *A Guide to Energy Master Planning of High-Performance Districts and Communities*. https://doi.org/10.2172/1734654

180. Tobii Pro. 2020. Latest in wearable eye tracking | Choose Tobii Pro Glasses 3. Retrieved November 18, 2021 from https://www.tobiipro.com/product-listing/tobii-pro-glasses-3/
181. Juan C. Quiroz, Sushil J. Louis, Amit Banerjee, and Sergiu M. Dascalu. 2009. Towards creative design using collaborative interactive genetic algorithms. In *2009 IEEE Congress on Evolutionary Computation*, 1849–1856. https://doi.org/10.1109/CEC.2009.4983166
182. S. D. Rajan. 1995. Sizing, Shape, and Topology Design Optimization of Trusses Using Genetic Algorithm. *Journal of Structural Engineering* 121, 10: 1480–1487. https://doi.org/10.1061/(ASCE)0733-9445(1995)121:10(1480)
183. Inge Rechenberg. 1965. *Cybernetic Solution Path of an Experimental Problem*. Ministry of Aviation.
184. Daniel Říha, Klára Bártová, and Jakub Binter. 2016. Use of Virtual Reality and Human-Computer Interface for Diagnostic and Treatment Purposes in Human Sexuality Research. In *Design, User Experience, and Usability: Technological Contexts* (Lecture Notes in Computer Science), 294–305. https://doi.org/10.1007/978-3-319-40406-6_28
185. Edward Robbins, Stephen Sanders, Adrian Williams, and Peter Allan. 2009. The use of virtual reality and intelligent database systems for procedure planning, visualisation, and real-time component tracking in remote handling operations. *Fusion Engineering and Design* 84, 7: 1628–1632. https://doi.org/10.1016/j.fusengdes.2009.01.049
186. T.P. Runarsson and Xin Yao. 2000. Stochastic ranking for constrained evolutionary optimization. *IEEE Transactions on Evolutionary Computation* 4, 3: 284–294. https://doi.org/10.1109/4235.873238
187. David Rutten. 2013. Galapagos: On the Logic and Limitations of Generic Solvers. *Architectural Design* 83, 2: 132–135. https://doi.org/10.1002/ad.1568
188. Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '00*, 71–78. https://doi.org/10.1145/355017.355028
189. Christian Sandor and Gudrun Klinker. 2005. A Rapid Prototyping Software Infrastructure for User Interfaces in Ubiquitous Augmented Reality. *Personal and Ubiquitous Computing* 9: 169–185. https://doi.org/10.1007/s00779-004-0328-1
190. Luis Santos, António Leitão, and Luisa Caldas. 2018. A comparison of two light-redirecting fenestration systems using a modified modeling technique for Radiance 3-phase method simulations. *Solar Energy* 161: 47–63. https://doi.org/10.1016/j.solener.2017.12.020
191. Gerhard Schmitt. 1999. *Information Architecture: Basics and Future of Caad*. Springer Science & Business Media.
192. M. A. Schnabel, Thomas Kvan, Ernst Kruijff, and Dirk Donath. 2001. The First Virtual Environment Design Studio. In *Architectural Information Management [19th eCAADe Conference Proceedings / ISBN 0-9523687-8-1] Helsinki (Finland) 29-31 August 2001, pp. 394-400*. Retrieved August 18, 2021 from http://papers.cumincad.org/cgi-bin/works/BrowseTreefield=seriesorder=AZ/Show?1d5a
193. Donald A. Schön. 1985. *The Design Studio: An Exploration of Its Traditions and Potentials*. RIBA Publications for RIBA Building Industry Trust.
194. Gerhard Schubert, David Schattel, Marcus Tönnis, Gudrun Klinker, and Frank Petzold. 2015. Tangible Mixed Reality On-Site: Interactive Augmented Visualisations from Architectural Working Models in Urban Design. In *Computer-Aided Architectural Design*

Futures. The Next City - New Technologies and the Future of the Built Environment (Communications in Computer and Information Science), 55–74. https://doi.org/10.1007/978-3-662-47386-3_4
195. scooley. HoloLens 2 hardware. Retrieved November 18, 2021 from https://docs.microsoft.com/en-us/hololens/hololens2-hardware
196. Stacey Scott, Neal Lesh, and Gunnar Klau. 2002. Investigating Human-Computer Optimization. *Proc. of CHI 2002* 4. https://doi.org/10.1145/503376.503405
197. S. S. Shapiro and M. B. Wilk. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52, 3/4: 591–611. https://doi.org/10.2307/2333709
198. Avishag Shemesh, Gerry Leisman, Moshe Bar, and Yasha Jacob Grobman. 2021. A neurocognitive study of the emotional impact of geometrical criteria of architectural space. *Architectural Science Review* 64, 4: 394–407. https://doi.org/10.1080/00038628.2021.1940827
199. Avishag Shemesh, Ronen Talmon, Ofer Karp, Idan Amir, Moshe Bar, and Jacob Grobman. 2016. Affective response to architecture – investigating human reaction to spaces with different geometry. *Architectural Science Review* 60. https://doi.org/10.1080/00038628.2016.1266597
200. Joel Simon. Evolving Floorplans. Retrieved November 25, 2021 from http://www.joelsimon.net/evo_floorplans.html
201. James Simpson, Megan Freeth, Kimberley J. Simpson, and Kevin Thwaites. 2019. Visual engagement with urban street edges: insights using mobile eye-tracking. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* 12, 3: 259–278. https://doi.org/10.1080/17549175.2018.1552884
202. Jordan W. Smith. 2015. Immersive Virtual Environment Technology to Supplement Environmental Perception, Preference and Behavior Research: A Review with Applications. *International Journal of Environmental Research and Public Health* 12, 9: 11486–11505. https://doi.org/10.3390/ijerph120911486
203. Agata Marta Soccini. 2017. Gaze estimation based on head movements in virtual reality applications using deep learning. In *2017 IEEE Virtual Reality (VR)*, 413–414. https://doi.org/10.1109/VR.2017.7892352
204. Aditya Soman, Shervin Azadi, and Pirouz Nourian. 2022. *DeciGenArch A generative design methodology for participatory architectural configuration via multi-criteria decision analysis*. https://doi.org/10.52842/conf.ecaade.2022.1.459
205. Hadas Sopher, Dafna Fisher-Gewirtzman, and Yehuda E Kalay. 2018. Use of Immersive Virtual Environment in the Design Studio. 10.
206. Hadas Sopher and John S Gero. 2021. Effect of Immersive VR on Communication Patterns in Architectural Design Critiques. In *Towards a new configurable architecture*, 8.
207. sostel. Eye tracking - Mixed Reality. Retrieved November 18, 2021 from https://docs.microsoft.com/en-us/windows/mixed-reality/design/eye-tracking
208. Robert Spang and Jan-Niklas Voigt-Antons. 2020. Accuracy Assessment of ARKit 2 Based Gaze Estimation. . 439–449. https://doi.org/10.1007/978-3-030-49059-1_32
209. N. Srinivas and Kalyanmoy Deb. 1994. Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation* 2, 3: 221–248. https://doi.org/10.1162/evco.1994.2.3.221

210. Marc Steen. 2011. Tensions in human-centred design. *CoDesign* 7, 1: 45–60. https://doi.org/10.1080/15710882.2011.563314
211. Nicolai Steinø, Miray Baş Yıldırım, and Mine Özkar. 2013. Parametric Design Strategies for Collaborative and Participatory Urban Design. In *eCAADe 2013: Computation and Performance*, 10.
212. Sophie Stellmach, Lennart Nacke, and Raimund Dachselt. 2010. Advanced gaze visualizations for three-dimensional virtual environments. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (ETRA '10), 109–112. https://doi.org/10.1145/1743666.1743693
213. L. Suchman. 1993. *Participatory design: Principles and practices*. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US.
214. Ayoung Suh and Jane Prophet. 2018. The state of immersive technology research: A literature analysis. *Computers in Human Behavior* 86: 77–90. https://doi.org/10.1016/j.chb.2018.04.019
215. Liang Sun, Weikang Tan, Yanbing Ren, Xiang Ji, Z. Wang, and Pei Li. 2020. Research on Visual Comfort of Underground Commercial Streets' Pavement in China on the Basis of Virtual Simulation. *Int. J. Pattern Recognit. Artif. Intell.* https://doi.org/10.1142/S0218001420500056
216. Xiaoyan Sun, Dunwei Gong, Yaochu Jin, and Shanshan Chen. 2012. A New Surrogate-Assisted Interactive Genetic Algorithm With Weighted Semisupervised Learning. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* 43. https://doi.org/10.1109/TSMCB.2012.2214382
217. Xiaoyan Sun, Dunwei Gong, Yaochu Jin, and Shanshan Chen. 2012. A New Surrogate-Assisted Interactive Genetic Algorithm With Weighted Semisupervised Learning. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* 43. https://doi.org/10.1109/TSMCB.2012.2214382
218. Xiaoyan Sun, Jie Ren, and Dunwei Gong. 2010. Interval Fitness Interactive Genetic Algorithms with Variational Population Size Based on Semi-supervised Learning. In *Advances in Neural Networks - ISNN 2010*, Liqing Zhang, Bao-Liang Lu and James Kwok (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 288–295. https://doi.org/10.1007/978-3-642-13278-0_37
219. Patrick Surry and Nicholas Radcliffe. 1997. The COMOGA method: Constrained optimisation by multi-objective genetic algorithms. *Control and Cybernetics* 26.
220. Ivan Edward Sutherland. 1963. Sketchpad, a man-machine graphical communication system. Massachusetts Institute of Technology. Retrieved October 11, 2021 from https://dspace.mit.edu/handle/1721.1/14979
221. H. Takagi. 2001. Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE* 89, 9: 1275–1296. https://doi.org/10.1109/5.949485
222. Xinxing Tang, Hironao Yamada, Lingtao Huang, and Ahmad Yusof. 2010. Virtual reality-based teleoperation construction robot control system with 3Dvisor device. 384–388. https://doi.org/10.1109/ICMA.2010.5588360

223. A. Alper Tasmektepligil and Erkan Gunpinar. 2022. SplineLearner: Generative learning system of design constraints for models represented using B-spline surfaces. *Advanced Engineering Informatics* 51: 101478. https://doi.org/10.1016/j.aei.2021.101478
224. Benjamin W Tatler, Nicholas J Wade, Hoi Kwan, John M Findlay, and Boris M Velichkovsky. 2010. Yarbus, eye movements, and vision. *i-Perception* 1, 1: 7–27. https://doi.org/10.1068/i0382
225. Gustav Taxén. 2004. Introducing participatory design in museums. In *Proceedings of the eighth conference on Participatory design Artful integration: interweaving media, materials and practices - PDC 04*, 204. https://doi.org/10.1145/1011870.1011894
226. Peter Testa, Una-May O'Reilly, Devyn Weiser, and Ian Ross. 2001. Emergent Design: A Crosscutting Research Program and Design Curriculum Integrating Architecture and Artificial Intelligence. *Environment and Planning B: Planning and Design* 28, 4: 481–498. https://doi.org/10.1068/b2702
227. Tarja Tiainen and Tiina Jouppila. 2019. Use of Virtual Environment and Virtual Prototypes in Co-Design: The Case of Hospital Design. *Computers* 8: 44. https://doi.org/10.3390/computers8020044
228. Tuan Q. Tran, Ronald L. Boring, Donald D. Dudenhoeffer, Bruce P. Hallbert, M.David Keller, and Tessa M. Anderson. 2007. Advantages and disadvantages of physiological assessment for next generation control room design. In *2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting*, 259–263. https://doi.org/10.1109/HFPP.2007.4413216
229. Tom Tullis and Bill Albert. 2013. *Measuring the User Experience - 2nd Edition*. Morgan Kaufmann, Boston. Retrieved October 8, 2021 from https://www.elsevier.com/books/measuring-the-user-experience/albert/978-0-12-415781-1
230. Christopher J. Turner, Windo Hutabarat, John Oyekan, and Ashutosh Tiwari. 2016. Discrete Event Simulation and Virtual Reality Use in Industry: New Opportunities and Future Trends. *IEEE Transactions on Human-Machine Systems* 46, 6: 882–894. https://doi.org/10.1109/THMS.2016.2596099
231. Belkis Uluoğlu. 2000. Design knowledge communicated in studio critiques. *Design Studies* 21, 1: 33–58. https://doi.org/10.1016/S0142-694X(99)00002-2
232. Z. Uykan, C. Guzelis, M.E. Celebi, and H.N. Koivo. 2000. Analysis of input-output clustering for determining centers of RBFN. *IEEE Transactions on Neural Networks* 11, 4: 851–858. https://doi.org/10.1109/72.857766
233. Robert Vierlinger. 2013. Multi Objective Design Interface. Technischen Universität Wien. https://doi.org/10.13140/RG.2.1.3401.0324
234. Sophia Vyzoviti. 2003. *Folding Architecture: Spatial, Structural and Organizational Diagrams*. BIS Publishers B.V.
235. Sophia Vyzoviti. 2006. *Super Surfaces: Generating Forms for Architecture, Products and Fashion*. BIS Publishers B.V.
236. Julian Wang, Luisa Caldas, Lei Huo, and Yehao Song. 2016. Simulation-Based Optimization of Window Properties Based on Existing Products.
237. Weimin Wang, Hugues Rivard, and Radu Zmeureanu. 2005. An object-oriented framework for simulation-based green building design optimization with genetic algorithms. *Advanced Engineering Informatics* 19, 1: 5–23. https://doi.org/10.1016/j.aei.2005.03.002

238. Nick Wates and Charles Knevitt. 1987. *Community Architecture: How People are Creating Their Own Environment*. Penguin.
239. Bartholomäus Wissmath, Daniel Stricker, David Weibel, Eva Siegenthaler, and Fred W. Mast. 2010. The Illusion of Being Located in Dynamic Virtual Environments. Can Eye Movement Parameters Predict Spatial Presence? *Journal of Eye Movement Research* 3, 5. https://doi.org/10.16910/jemr.3.5.2
240. Thomas Wortmann and Jonathan Natanian. 2020. Multi-Objective Optimization for Zero-Energy Urban Design in China: A Benchmark. In *SimAUD 2020*.
241. Anbang Xu and Brian Bailey. 2012. What do you think? a case study of benefit, expectation, and interaction in a large online critique community. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (CSCW '12), 295–304. https://doi.org/10.1145/2145204.2145252
242. Anbang Xu, Shih-Wen Huang, and Brian Bailey. 2014. Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (CSCW '14), 1433–1444. https://doi.org/10.1145/2531602.2531604
243. Muchen Yan and Martin Tamke. 2021. Augmented Reality for Experience-centered Spatial Design - A quantitative assessment method for architectural space. In *Stojakovic, V and Tepavcevic, B (eds.), Towards a new, configurable architecture - Proceedings of the 39th eCAADe Conference - Volume 1, University of Novi Sad, Novi Sad, Serbia, 8-10 September 2021, pp. 173-180*. Retrieved October 5, 2021 from http://papers.cumincad.org/cgi-bin/works/paper/ecaade2021_069
244. Wei Yan. WP-BIM: Web-based Parametric BIM Towards Online Collaborative Design and Optimization. 8.
245. Alvin Yuan, Kurt Luther, Markus Krause, Sophie Isabel Vennix, Steven P Dow, and Bjorn Hartmann. 2016. Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16), 1005–1017. https://doi.org/10.1145/2818048.2819953
246. Lelin Zhang, Yu Mike Chi, Eve Edelstein, Jurgen Schulze, Klaus Gramann, Alvaro Velasquez, Gert Cauwenberghs, and Eduardo Macagno. 2010. Wireless physiological monitoring and ocular tracking: 3D calibration in a fully-immersive virtual health care environment. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* 2010: 4464–4467. https://doi.org/10.1109/IEMBS.2010.5625969
247. Yuxuan Zhang, Hexu Liu, Shih-Chung Kang, and Mohamed Al-Hussein. 2020. Virtual reality applications for the built environment: Research trends and opportunities. *Automation in Construction* 118: 103311. https://doi.org/10.1016/j.autcon.2020.103311
248. Ximing Zhong, Immanuel Koh, and Prof Dr Pia Fricker. 2023. Building-GNN: Exploring a co-design framework for generating controllable 3D building prototypes by graph and recurrent neural networks. In *Digital Design Reconsidered: Proceedings of the 41st Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2023)*, 431–440. https://doi.org/10.52842/conf.ecaade.2023.2.431

249. Xiaojin Zhu and Andrew B. Goldberg. 2009. Introduction to Semi-Supervised Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 3, 1: 1–130. https://doi.org/10.2200/S00196ED1V01Y200906AIM006
250. Eckart Zitzler and Lothar Thiele. 1999. An evolutionary algorithm for multiobjective optimization: the strength Pareto approach. *Technical Report: TIK43*.
251. 2015. How do Tobii Eye Trackers work? | Learn more with Tobii Pro. Retrieved August 9, 2021 from https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/
252. 2016. Living Building Challenge Certification | Living-Future.org. Retrieved November 15, 2021 from https://living-future.org/lbc-3_1/certification/
253. 2016. Zero Energy Certification | Living-Future.org. Retrieved November 15, 2021 from https://living-future.org/zero-energy/certification/
254. 2019. Eye-Tracking Boston City Hall to Better Understand Human Perception and the Architectural Experience. *Imotions*. Retrieved November 27, 2021 from https://imotions.com/publications/eye-tracking-boston-city-hall-to-better-understand-human-perception-and-the-architectural-experience/
255. 2020. Microsoft HoloLens 2: Improved Research Mode to facilitate computer vision research. *Microsoft Research*. Retrieved November 27, 2021 from https://www.microsoft.com/en-us/research/blog/microsoft-hololens-2-improved-research-mode-to-facilitate-computer-vision-research/
256. 2021. James Turrell: Into The Light | MASS MoCA. Retrieved September 29, 2021 from https://massmoca.org/event/james-turrell-skyspace/
257. 2021. Product Lineup | FOVE Official Website. Retrieved November 18, 2021 from https://fove-inc.com/product/, https://fove-inc.com/product/
258. Amsterdam 3D printed steel bridge. Retrieved from https://www.arup.com/en/projects/mx3d-bridge
259. RIBA Plan of Work. Retrieved November 14, 2021 from https://www.architecture.com/knowledge-and-resources/resources-landing-page/riba-plan-of-work
260. B101- 2017 Standard Form of Agreement Between Owner and Architect. Retrieved November 14, 2021 from https://www.aiacontracts.org/contract-documents/25141-owner-architect-agreement
261. The Net-Zero Standard. *Science Based Targets*. Retrieved November 15, 2021 from https://sciencebasedtargets.org/net-zero
262. Standard 90.1. Retrieved November 28, 2021 from https://www.ashrae.org/technical-resources/bookstore/standard-90-1
263. TimesMachine: Sunday March 19, 1922 - NYTimes.com. *The New York Times*. Retrieved November 14, 2021 from http://timesmachine.nytimes.com/timesmachine/1922/03/19/issue.html
264. 2019 California Building Code, Title 24. Retrieved November 15, 2021 from https://codes.iccsafe.org/content/CBC2019P4
265. DBF: AI Generative Design and Spatial Analytics. Retrieved November 12, 2023 from https://www.digitalbluefoam.com/
266. Learn. *Dynamo BIM*. Retrieved November 28, 2021 from https://dynamobim.org/

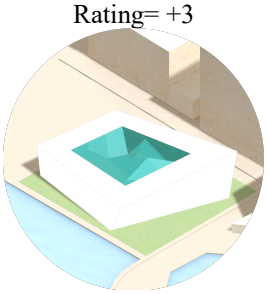
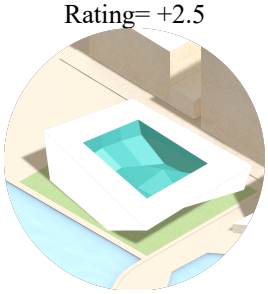
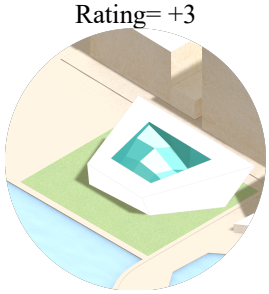
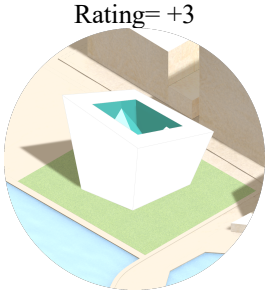
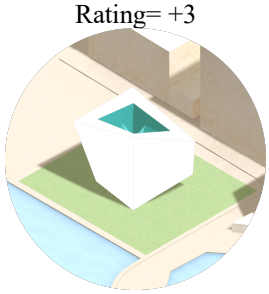
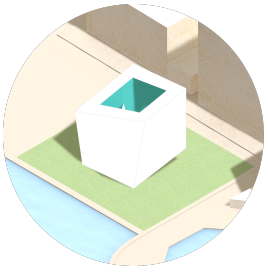
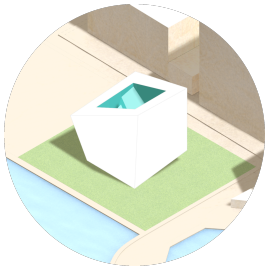
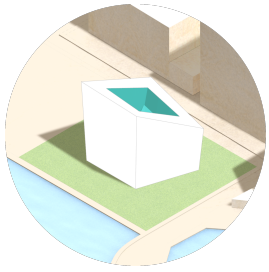
267. Rechenraum GmbH - goat. Retrieved November 1, 2021 from https://www.rechenraum.com/en/goat.html
268. Project Refinery Beta | Autodesk. Retrieved November 8, 2021 from https://www.autodesk.com/campaigns/refinery-beta
269. Visualizing Results in Generative Design. Retrieved November 8, 2021 from https://www.generativedesign.org/03-hello-gd-for-revit/03-04_visualizing-results-in-gd-for-revit
270. EnergyPlus. Retrieved April 22, 2024 from https://energyplus.net/
271. An Online Whiteboard & Visual Collaboration Platform for Teamwork | Miro. *https://miro.com/*. Retrieved September 7, 2021 from https://miro.com/
272. Amazon Mechanical Turk. Retrieved November 24, 2021 from https://www.mturk.com/
273. virtual reality, n. *OED Online*. Retrieved September 20, 2021 from http://www.oed.com/view/Entry/328583
274. Oculus Device Specifications | Oculus Developers. Retrieved October 14, 2021 from https://developer.oculus.com/resources/oculus-device-specs/
275. Meta Quest 3: New Mixed Reality VR Headset - Shop Now. Retrieved November 7, 2023 from https://www.meta.com/quest/quest-3/
276. Tilt Brush by Google. Retrieved October 12, 2021 from https://www.tiltbrush.com/
277. Fologram. Retrieved from https://fologram.com/
278. End of the line for Foster's Tulip after UK government rejects project due to climate concerns. *Archinect*. Retrieved November 19, 2021 from https://archinect.com/news/article/150287967/end-of-the-line-for-foster-s-tulip-after-uk-government-rejects-project-due-to-climate-concerns
279. VIVE Pro Eye Overview | VIVE European Union. Retrieved from https://www.vive.com/eu/product/vive-pro-eye/overview/www.vive.com/eu/product/vive-pro-eye/overview/
280. The HTC Vive Pro Eye Brings Eye Tracking to Engineers in VR | designnews.com. Retrieved from https://www.designnews.com/electronics-test/htc-vive-pro-eye-brings-eye-tracking-engineers-vr
281. HTC Vive Pro Eye - VR Headset with Eye Tracking integration. *Tobii VR*. Retrieved October 14, 2021 from https://vr.tobii.com/integrations/htc-vive-pro-eye/
282. Apple Vision Pro. *Apple*. Retrieved November 6, 2023 from https://www.apple.com/apple-vision-pro/
283. Eye tracking. *Varjo for developers*. Retrieved November 21, 2021 from https://developer.varjo.com/docs/get-started/eye-tracking-with-varjo-headset
284. ARKit | Apple Developer Documentation. Retrieved November 18, 2021 from https://developer.apple.com/documentation/arkit
285. Tracking and Visualizing Faces | Apple Developer Documentation. Retrieved September 29, 2021 from https://developer.apple.com/documentation/arkit/content_anchors/tracking_and_visualizing_faces
286. EY Seren – International design & transformation consultancy. Retrieved November 20, 2021 from https://www.ey-seren.com/

287. INSIGHT - 5 Channel EEG Brainwear®. *EMOTIV*. Retrieved November 20, 2021 from https://www.emotiv.com/insight/
288. Cloud Storage for Firebase. *Firebase*. Retrieved October 11, 2021 from https://firebase.google.com/docs/storage
289. Firebase. *Firebase*. Retrieved October 8, 2020 from https://firebase.google.com/
290. About AR Foundation | AR Foundation | 4.0.8. Retrieved from https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@4.0/manual/index.html
291. ARKit | Apple Developer Documentation. Retrieved October 8, 2020 from https://developer.apple.com/documentation/arkit/
292. VIVE Pro Eye Specs | VIVE European Union. Retrieved from https://www.vive.com/eu/product/vive-pro-eye/specs/www.vive.com/eu/product/vive-pro-eye/specs/
293. Cloud Computing Services - Amazon Web Services (AWS). Retrieved August 20, 2023 from https://aws.amazon.com/
294. SEC. 249.80. MISSION ROCK SPECIAL USE DISTRICT. *American Legal Publishing*. Retrieved June 1, 2023 from https://codelibrary.amlegal.com/codes/san_francisco/latest/sf_planning/0-0-0-58323
295. AR tracked image manager | AR Foundation | 4.0.12. Retrieved July 29, 2023 from https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@4.0/manual/tracked-image-manager.html
296. DEAP documentation — DEAP 1.4.1 documentation. Retrieved March 18, 2024 from https://deap.readthedocs.io/en/master/

Appendix A. Highly Rated Solutions by Experts and Non-Experts

A.1 Experts Highly Rated Solutions

Table A.1.1. Expert AA1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
AA1	Experiment 1 Intervention 1	Option 8 AV= 13% NPZ= 20% UOS= 51% Rating= +3 	Option 7 AV = 11% NPZ= 26% UOS= 43% Rating= +2.5 	Option 4 AV= 22% NPZ= 30% UOS= 69% Rating= +3 
	Experiment 1 Intervention 2	Option 9 AV= 2% NPZ= 24% UOS= 83% Rating = +2.5 	Option 13 AV= 6% NPZ= 25 UOS = 83% Rating= +3 	Option 2 AV= 23% NPZ= 1% UOS= 88% Rating= +3 
	Experiment 2 Intervention 1	Option 7 AV= 7% NPZ= 4% UOS= 87% 	Option 6 AV= 27% NPZ= 2% UOS= 88% 	Option 9 AV= 31% NPZ= 0.0 UOS= 80% 

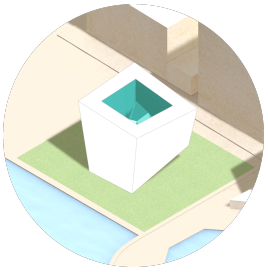
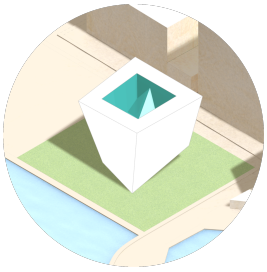
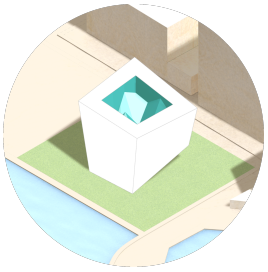
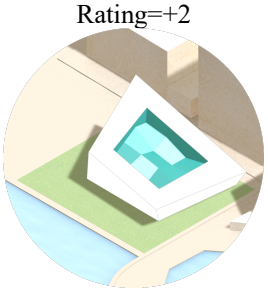
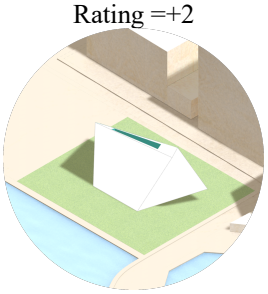
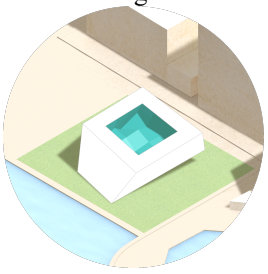
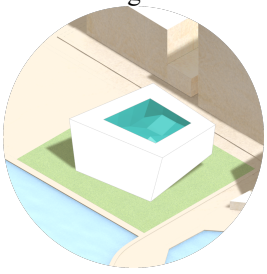
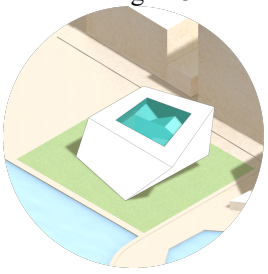
Participant	Experiment	Design Options		
	Experiment 2 Intervention 2	Option 7 AV= 15% NPZ= 1% UOS= 86% 	Option 9 AV=11% NPZ= 4% UOS= 88% 	Option 6 AV= 10% NPZ= 1% UOS= 86% 

Table A.1.2. Expert BB1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
BB1	Experiment 1 Intervention 1	Option 26 AV= 3% NPZ=47% UOS= 67% Rating=+2 	Option 9 AV= 22% NPZ= 8% UOS= 84% Rating= +2 	Option 3 AV= 48% NPZ= 3% UOS= 88% Rating =+2 
	Experiment 1 Intervention 2	Option 18 AV= 15% NPZ=12% UOS= 75% Rating=+2 	Option 25 AV= 12% NPZ= 17% UOS=79% Rating = +2 	Option 1 AV= 9% NPZ= 31% UOS= 73% Rating= +3 
	Experiment 2	Option 12 AV= 16%	Option 11 AV= 0	Option 13 AV= 31%

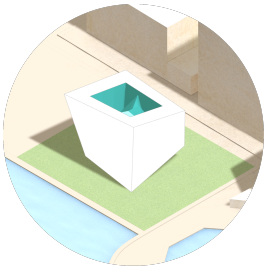
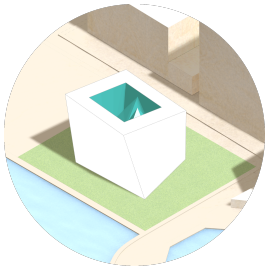
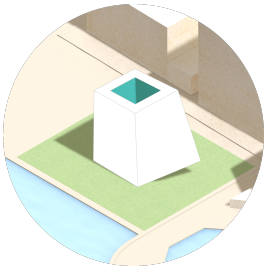
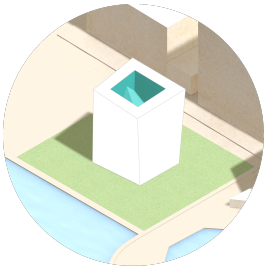
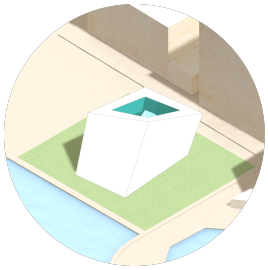
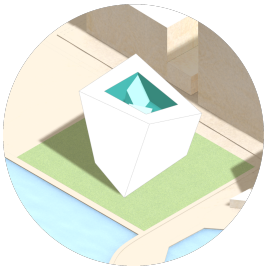
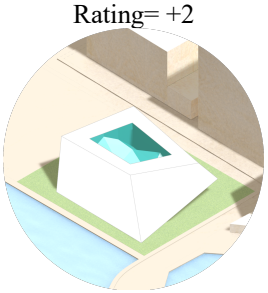
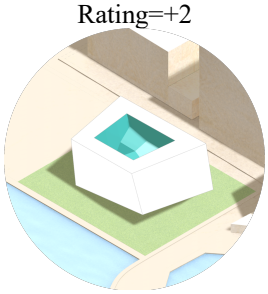
Participant	Experiment	Design Options		
	Intervention 1	NPZ= 6% UOS= 85% 	NPZ= 29% UOS= 74% 	NPZ= 0 UOS= 88% 
	Experiment 2 Intervention 2	Option 7 AV= 34% NPZ= 14% UOS= 87% 	Option 5 AV= 20% NPZ= 0 UOS= 80% 	Option 6 AV= 7% NPZ= 6% UOS= 84% 

Table A.1.3. Expert CC1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
CC1	Experiment 1 Intervention 1	Option 16 AV= 7% NPZ=18% UOS = 67% Rating= +2 	Option 2 AV= 5% NPZ= 29% UOS= 76% Rating=+2 	Option 7 AV= 5% NPZ= 55% UOS= 70% Rating= +2 
	Experiment 1 Intervention 2	Option 6 AV= 2% NPZ= 12% UOS= 80%	Option 1 AV= 1% NPZ= 40% UOS= 71%	Option 0 AV= 24% NPZ= 1% UOS= 86%

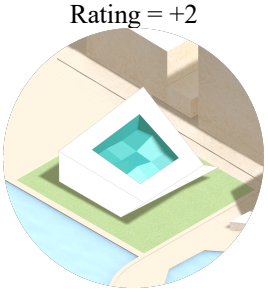
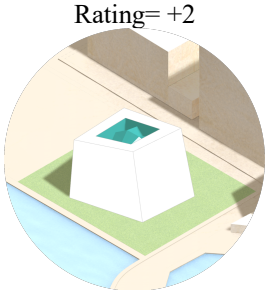
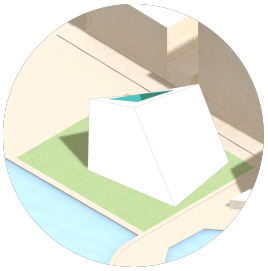
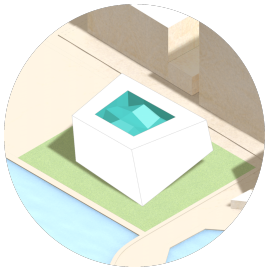
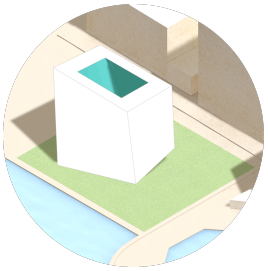
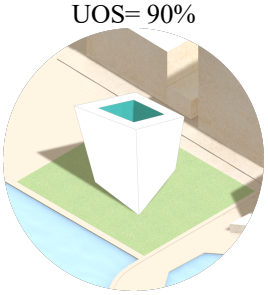
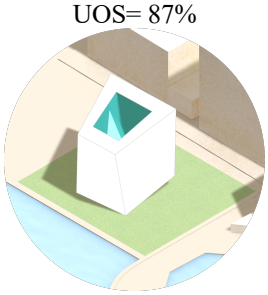
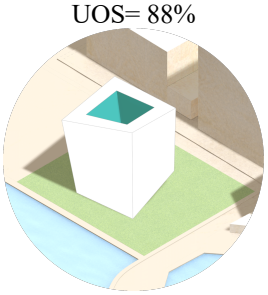
Participant	Experiment	Design Options		
		Rating = +2 	Rating= +2 	Rating= +2 
	Experiment 2 Intervention 1	Option 11 AV=0 NPZ= 13% UOS= 63% 	Option 8 AV= 2% NPZ= 17% UOS= 79% 	Option 9 AV= 12% NPZ= 3% UOS= 87% 
	Experiment 2 Intervention 2	Option 21 AV= 25% NPZ= 1% UOS= 90% 	Option 19 AV= 4% NPZ= 14% UOS= 87% 	Option 20 AV= 5% NPZ= 4% UOS= 88% 

Table A.1.4. Expert DD1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
DD1	Experiment 1 Intervention 1	Option 8 AV= 5% NPZ= 42% UOS= 49% Rating= +3	Option 19 AV= 17% NPZ= 34% UOS= 43% Rating= +2	Option 1 AV= 49% NPZ= 34% UOS= 43% Rating= +3

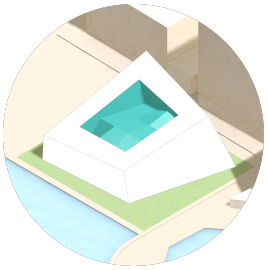
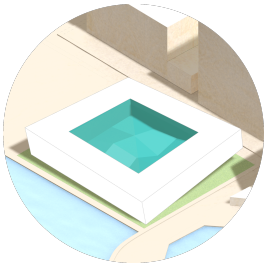
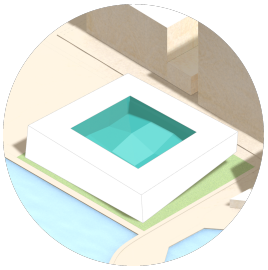
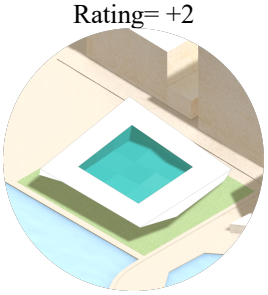
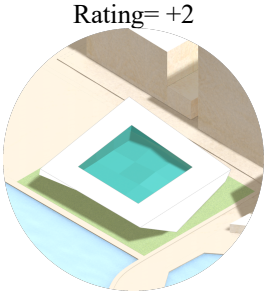
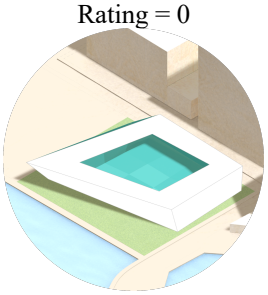
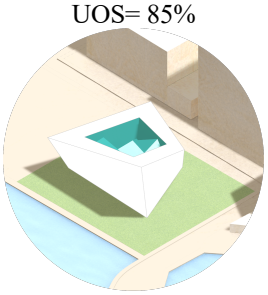
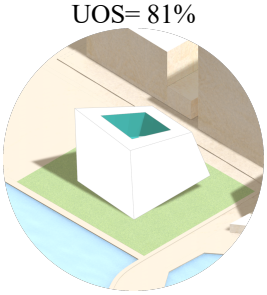
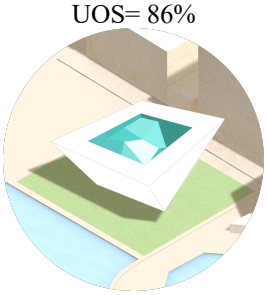
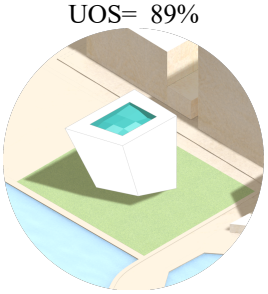
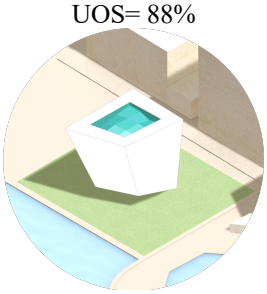
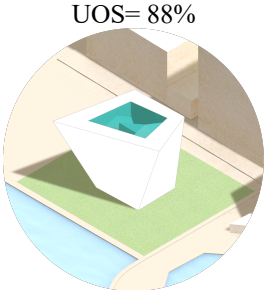
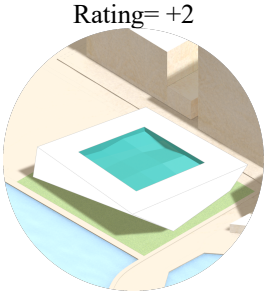
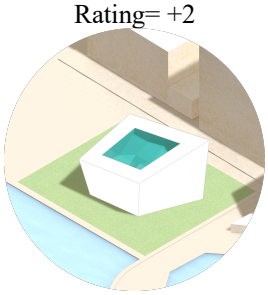
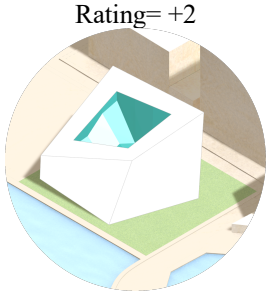
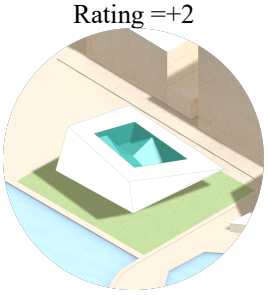
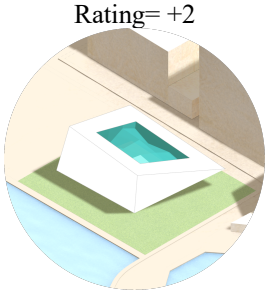
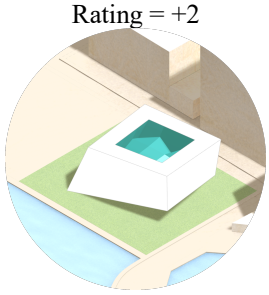
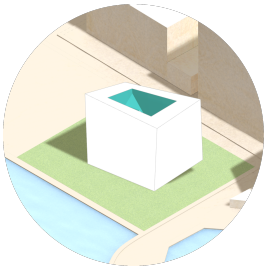
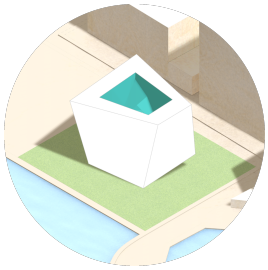
Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 2 AV = 13 NPZ= 47 UOS=59% Rating= +2 	Option 8 AV= 13% NPZ= 47% UOS= 41% Rating= +2 	Option 1 AV= 29% NPZ= 27% UOS= 37% Rating = 0 
	Experiment 2 Intervention 1	Option 23 AV= 4% NPZ= 6% UOS= 85% 	Option 0 AV=6% NPZ= 6% UOS= 81% 	Option 22 AV= 6% NPZ= 36% UOS= 86% 
	Experiment 2 Intervention 2	Option 10 AV= 12% NPZ= 43% UOS= 89% 	Option 0 AV= 4% NPZ= 43% UOS= 88% 	Option 22 AV= 12% NPZ= 13% UOS= 88% 

Table A.1.5. Expert EE1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
EE1	Experiment 1 Intervention 1	Option 4 AV= 6% NPZ= 61% UOS= 46% Rating= +2 	Option 18 AV = 28% NPZ= 42% UOS= 76% Rating= +2 	Option 8 AV= 26% NPZ= 54% UOS= 63% Rating= +2 
	Experiment 1 Intervention 2	Option 13 AV= 0 NPZ= 40% UOS= 69% Rating =+2 	Option 4 AV= 3% NPZ= 48% UOS= 69% Rating= +2 	Option 18 AV= 2% NPZ= 45% UOS= 72% Rating = +2 
	Experiment 2 Intervention 1	Option 10 AV= 18% NPZ= 3% UOS= 88% 	Option 3 AV= 5% NPZ= 10% UOS= 87% 	Option 13 AV= 5% NPZ= 20% UOS= 88% 
	Experiment 2 Intervention 2	Option 18 AV= 50% NPZ= 0 UOS= 93%	Option 16 AV= 17% NPZ= 1% UOS= 89%	Option 17 AV= 10% NPZ= 3% UOS= 89%

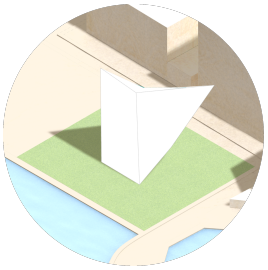
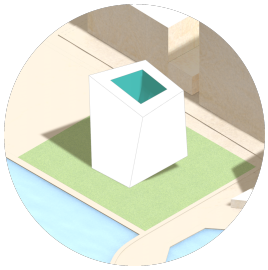
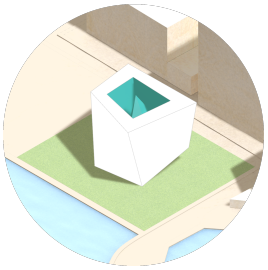
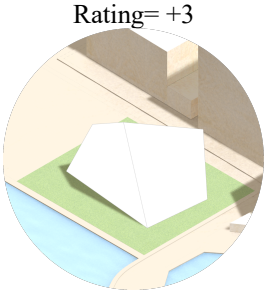
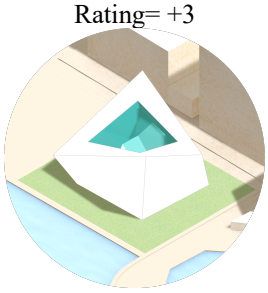
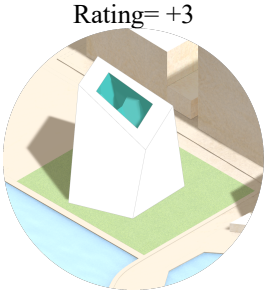
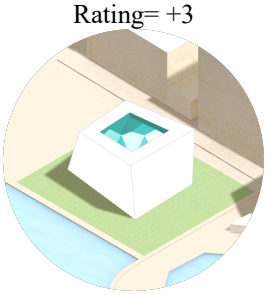
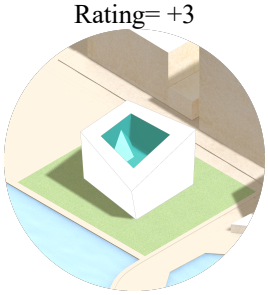
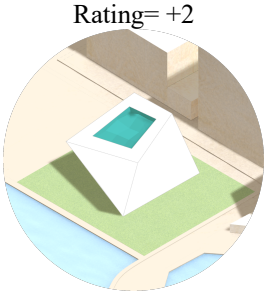
Participant	Experiment	Design Options		
				

Table A.1.6. Expert FF1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
FF1	Experiment 1 Intervention 1	Option 26 AV=16% NPZ= 33% UOS= 72% Rating= +3 	Option 5 AV= 10 NPZ= 50% UOS= 66% Rating= +3 	Option 11 AV= 39% NPZ= 29% UOS= 76% Rating= +3 
	Experiment 1 Intervention 2	Option 10 AV= 5% NPZ= 31% UOS= 73% Rating= +3 	Option 5 AV= 16% NPZ= 33% UOS= 76% Rating= +3 	Option 9 AV= 31% NPZ= 39% UOS= 81% Rating= +2 
	Experiment 2 Intervention 1	Option 12 AV= 7% NPZ= 12% UOS= 89%	Option 0 AV= 11% NPZ= 7% UOS= 87%	Option 3 AV= 3% NPZ= 12% UOS= 85%

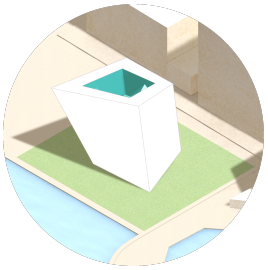
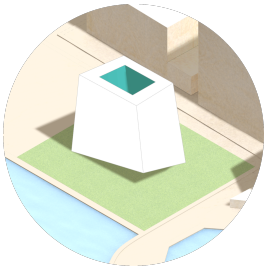
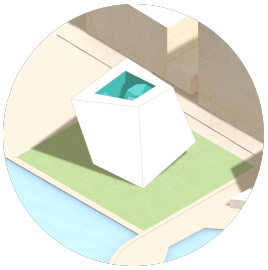
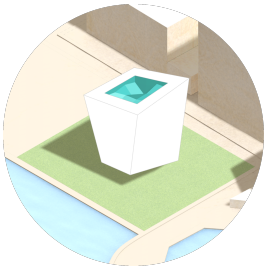
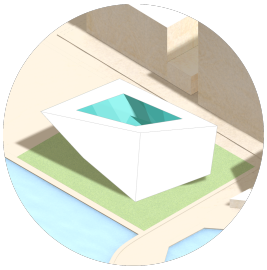
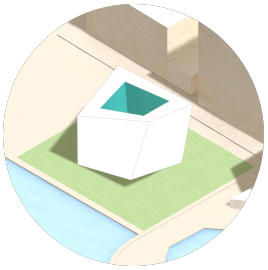
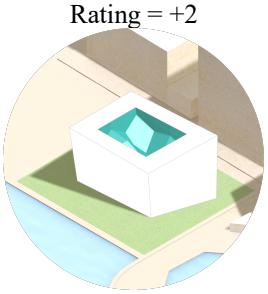
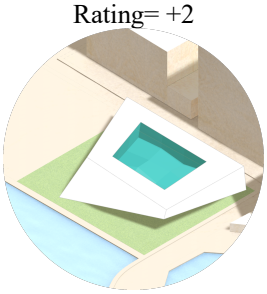
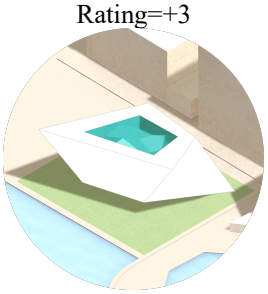
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 3 AV= 27% NPZ= 32% UOS= 91% 	Option 1 AV= 4% NPZ= 10% UOS= 77% 	Option 2 AV= 14% NPZ= 3% UOS= 87% 

Table A.1.7. Expert GG1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
GG1	Experiment 1 Intervention 1	Option 25 AV= 21% NPZ= 41% UOS= 72% Rating = +2 	Option 1 AV= 14% NPZ= 45% UOS= 55% Rating= +2 	Option 5 AV= 17% NPZ= 53% UOS= 78% Rating=+3 
	Experiment 1 Intervention 2	Option 4 AV= 0 NPZ= 15% UOS= 73% Rating= +2	Option 28 AV= 24% NPZ= 26% UOS= 84% Rating= +2	Option 22 AV= 25% NPZ= 26% UOS= 81% Rating= +2

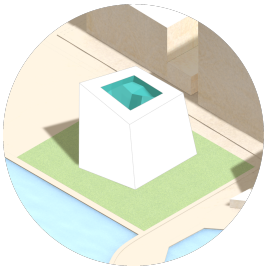
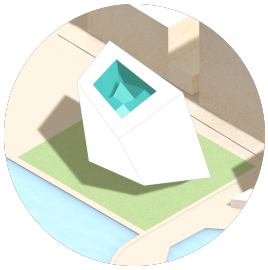
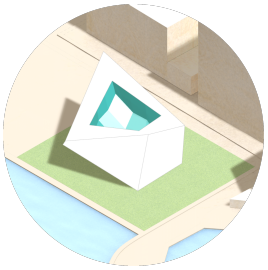
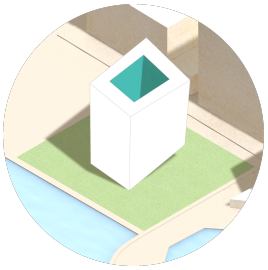
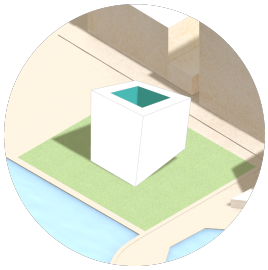
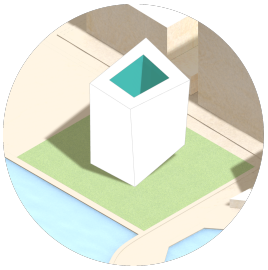
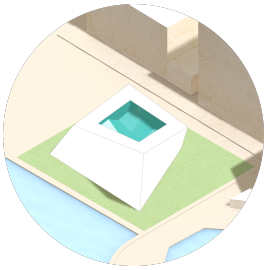
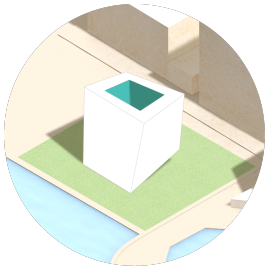
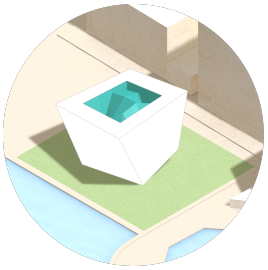
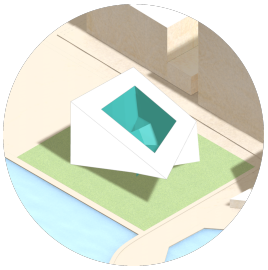
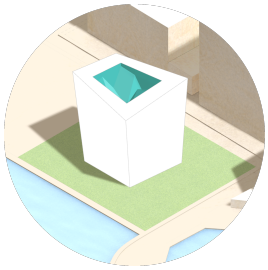
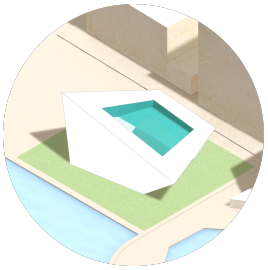
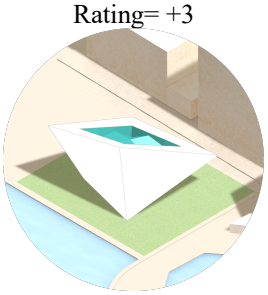
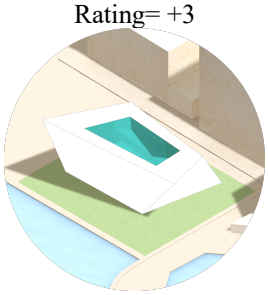
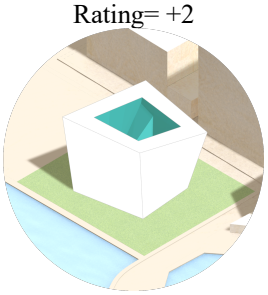
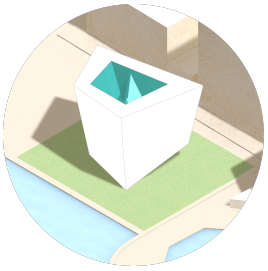
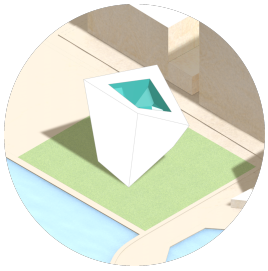
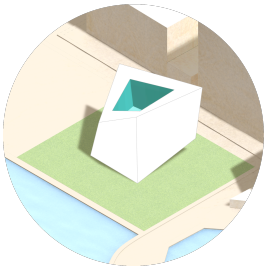
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 1	Option 10 AV= 1% NPZ= 1% UOS= 89% 	Option 6 AV= 36% NPZ= 0 UOS= 90% 	Option 18 AV= 2% NPZ= 1% UOS= 89% 
	Experiment 2 Intervention 2	Option 21 AV= 0 NPZ= 11% UOS= 70% 	Option 17 AV= 18% NPZ= 1% UOS= 86% 	Option 20 AV= 5% NPZ= 8% UOS= 88% 

Table A.1.8. Expert HH1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
HH1	Experiment 1 Intervention 1	Option 15 AV= 16% NPZ= 4% UOS= 86% Rating= +3	Option 16 AV= 4% NPZ= 33% UOS= 84% Rating= +3	Option 5 AV= 9% NPZ= 44% UOS= 70% Rating= +3

Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 11 AV= 3% NPZ= 8% UOS= 84% Rating= +3 	Option 1 AV= 2% NPZ= 15% UOS= 74% Rating= +3 	Option 8 AV= 18% NPZ= 2% UOS= 76% Rating= +2 
	Experiment 2 Intervention 1	Option 18 AV= 9% NPZ= 10% UOS= 87% 	Option 17 AV= 36% NPZ= 19% UOS= 91% 	Option 20 AV= 40% NPZ= 7% UOS= 90% 
	Experiment 2 Intervention 2	Option 10 AV= 24% NPZ= 1% UOS= 88%	Option 0 AV= 29% NPZ= 3% UOS= 89%	Option 14 AV= 1% NPZ= 4% UOS= 88%

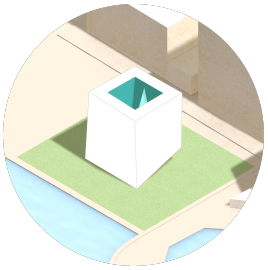
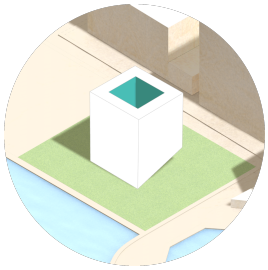
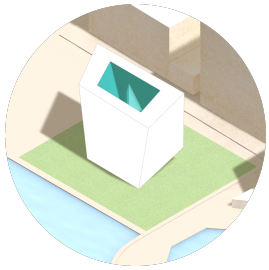
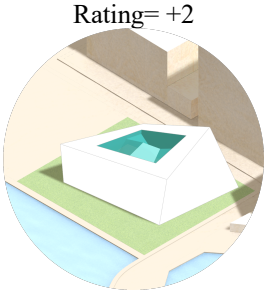
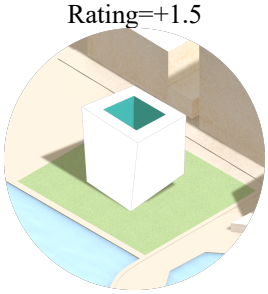
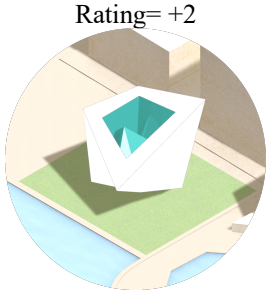
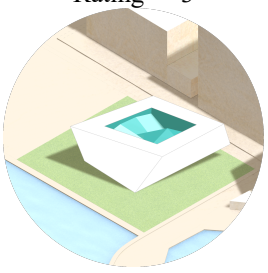
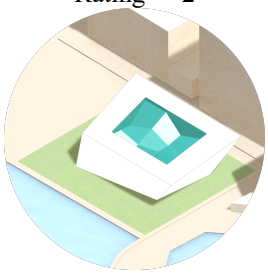
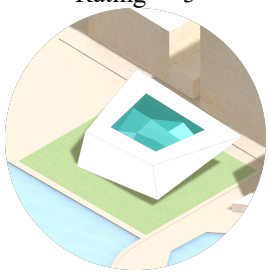
Participant	Experiment	Design Options		
				

Table A.1.9. Expert II1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
II1	Experiment 1 Intervention 1	Option 8 AV= 8 NPZ= 27% UOS= 57% Rating= +2 	Option 4 AV= 13% NPZ= 60% UOS= 63% Rating=+1.5 	Option 7 AV= 24% NPZ= 37% UOS= 65% Rating= +2 
	Experiment 1 Intervention 2	Option 11 AV= 20% NPZ= 16% UOS= 80% Rating= +3 	Option 9 AV= 9% NPZ= 41% UOS= 74% Rating= +2 	Option 5 AV= 16% NPZ= 27% UOS= 77% Rating= +3 
	Experiment 2 Intervention 1	Option 10 AV= 4% NPZ= 5% UOS= 87%	Option 9 AV= 7% NPZ= 4% UOS= 88%	Option 11 AV= 25% NPZ= 4% UOS= 88%

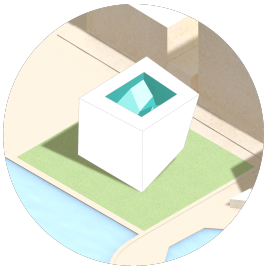
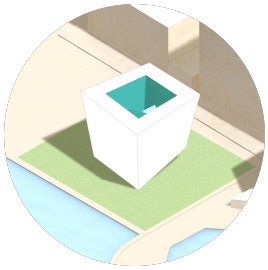
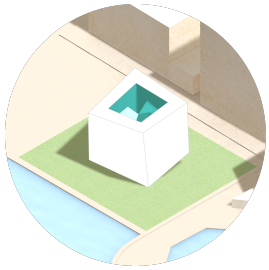
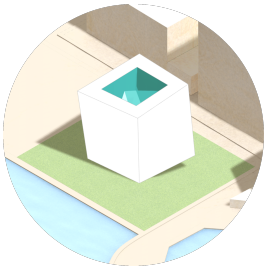
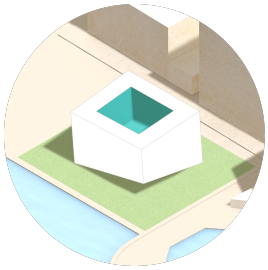
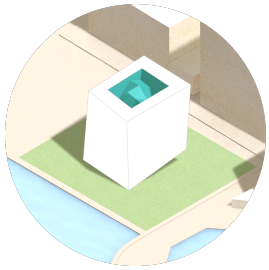
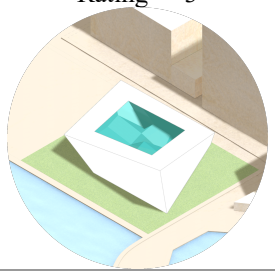
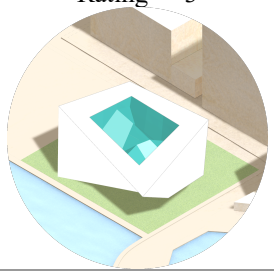
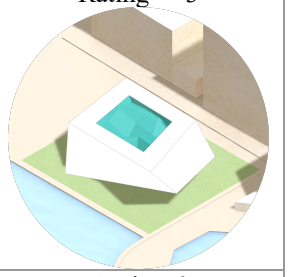
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 11 AV= 6% NPZ= 4% UOS= 89% 	Option 0 AV= 0 NPZ= 32% UOS= 79% 	Option 21 AV= 4% NPZ= 1% UOS= 88% 

Table A.1.10. Expert JJ1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
JJ1	Experiment 1 Intervention 1	Option 13 AV= 8% NPZ= 15% UOS= 78% Rating= +3 	Option 4 AV= 2% NPZ= 28% UOS= 72% Rating= +3 	Option 21 AV= 6% NPZ= 54% UOS= 73% Rating= +3 
	Experiment 1 Intervention 2	Option 17 AV= 19% NPZ= 34% UOS= 79% Rating= +2	Option 13 AV= 40% NPZ= 14% UOS= 75% Rating= +2	Option 6 AV= 46% NPZ= 2% UOS= 72% Rating= +2

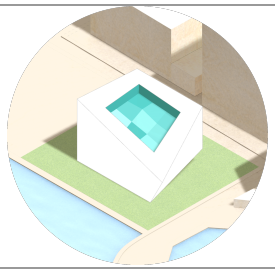
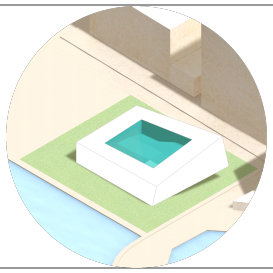
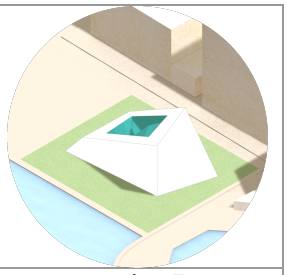
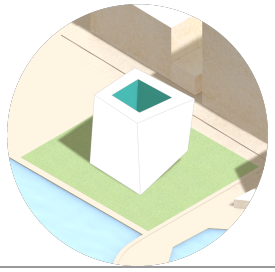
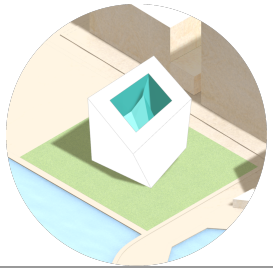
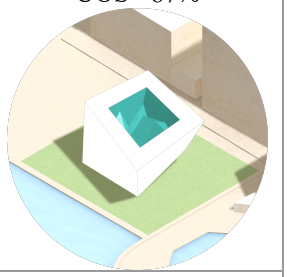
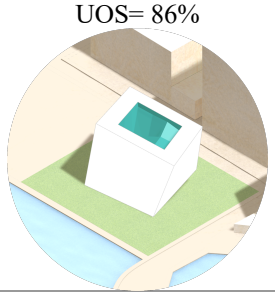
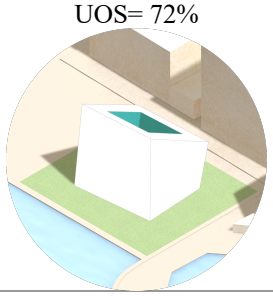
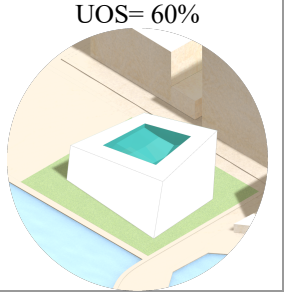
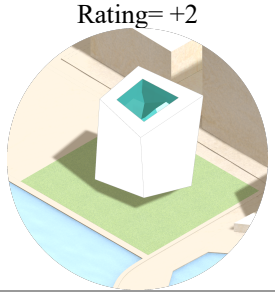
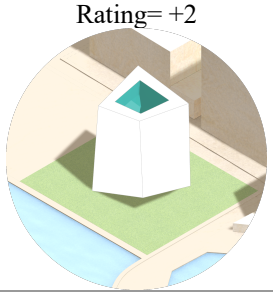
				
	Experiment 2 Intervention 1	Option 6 AV= 15% NPZ= 7% UOS= 86% 	Option 4 AV= 42% NPZ= 2% UOS= 89% 	Option 7 AV= 2% NPZ= 12% UOS= 87% 
	Experiment 2 Intervention 2	Option 14 A= 4% NPZ= 6% UOS= 86% 	Option 1 AV= 7% NPZ= 3% UOS= 72% 	Option 27 AV= 0 NPZ= 48% UOS= 60% 

Table A.1.11. Expert KK1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
KK1	Experiment 1 Intervention 1	Option 28 AV= 4% NPZ= 13% UOS= 88% Rating= +2 	Option 7 AV= 10% NPZ= 14% UOS= 88% Rating= +2 	Option 3 AV= 17% NPZ= 41% UOS= 86% Rating= +3 

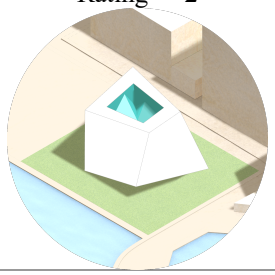
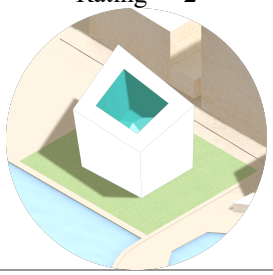
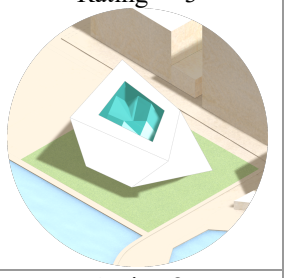
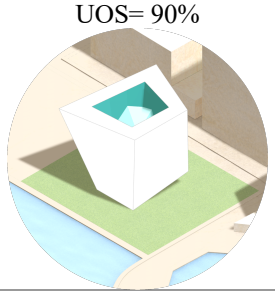
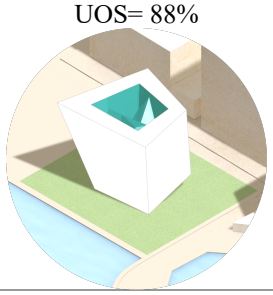
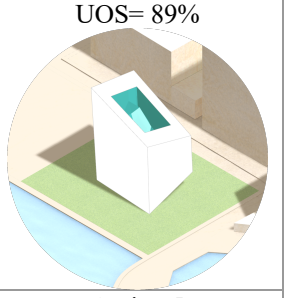
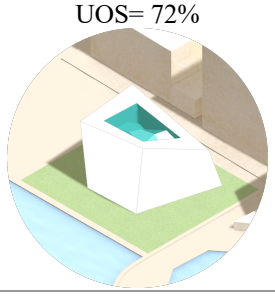
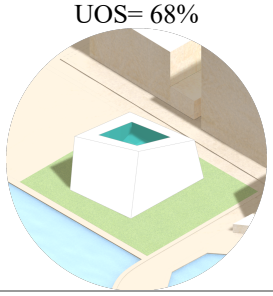
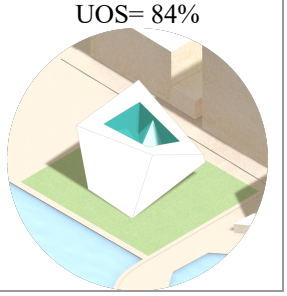
Participant	Experiment	Design Options		
	Experiment 1 Intervention 2	Option 14 AV= 2% NPZ= 39% UOS= 79% Rating= +2 	Option 13 AV= 32% NPZ= 29% UOS= 78% Rating= +2 	Option 4 AV= 32% NPZ= 36% UOS= 76% Rating= +3 
	Experiment 2 Intervention 1	Option 7 AV= 1% NPZ= 2% UOS= 90% 	Option 6 AV= 12% NPZ= 1% UOS= 88% 	Option 8 AV= 26% NPZ= 0 UOS= 89% 
	Experiment 2 Intervention 2	Option 6 AV= 0 NPZ= 43% UOS= 72% 	Option 4 AV= 0 NPZ= 40% UOS= 68% 	Option 5 AV= 0 NPZ= 8% UOS= 84% 

Table A.1.12. Expert LL1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
LL1	Experiment 1 Intervention 1	Option 6 AV=1% NPZ= 8% UOS= 65% Rating= +3	Option 7 AV= 16% NPZ= 27% UOS= 71% Rating= +3	Option 8 AV= 16% NPZ= 37% UOS= 72% Rating=+2

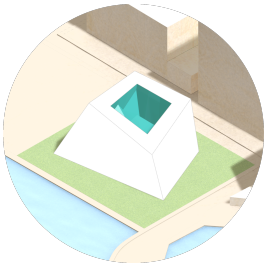
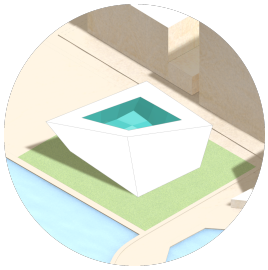
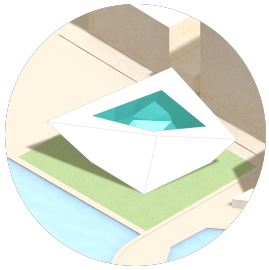
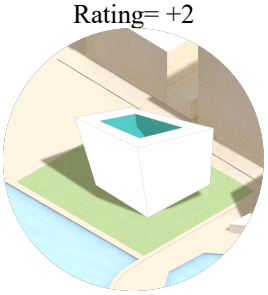
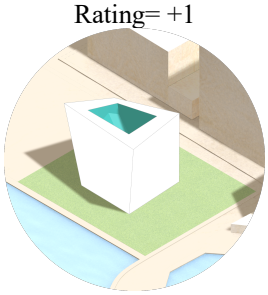
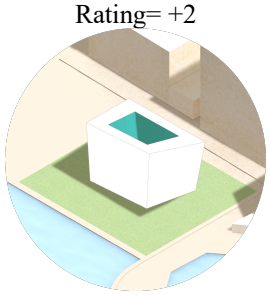
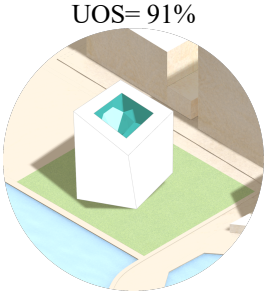
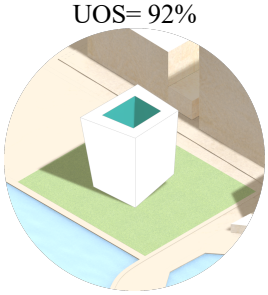
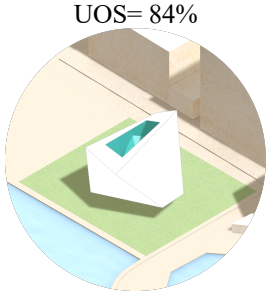
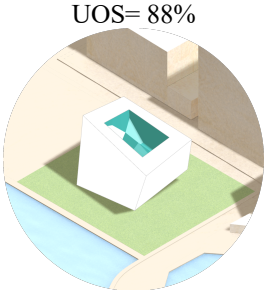
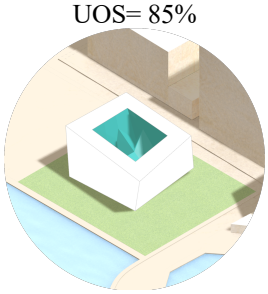
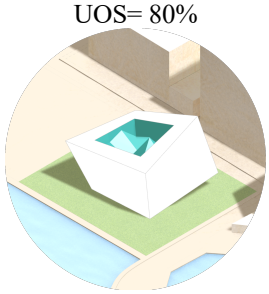
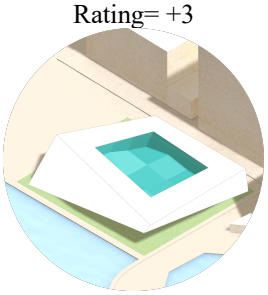
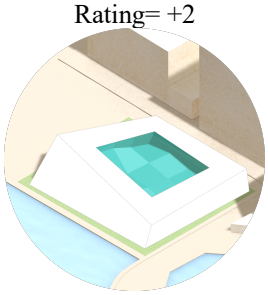
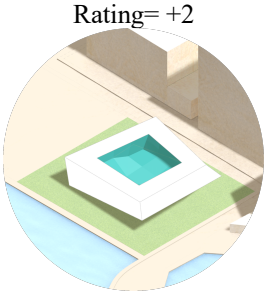
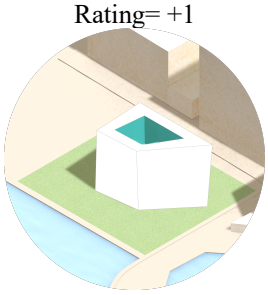
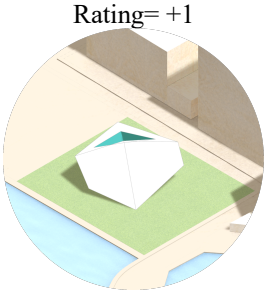
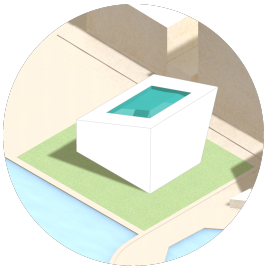
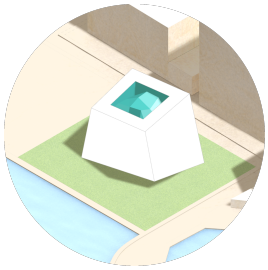
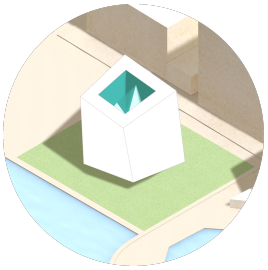
Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 10 AV= 7% NPZ= 8% UOS= 86% Rating= +2 	Option 8 AV= 7% NPZ= 24% UOS= 82% Rating= +1 	Option 2 AV= 19% NPZ= 1% UOS= 86% Rating= +2 
	Experiment 2 Intervention 1	Option 9 AV= 4% NPZ= 2% UOS= 91% 	Option 8 AV= 34% NPZ= 1% UOS= 92% 	Option 10 AV= 24% NPZ= 0 UOS= 84% 
	Experiment 2 Intervention 2	Option 9 AV= 9% NPZ= 5% UOS= 88% 	Option 1 AV= 9% NPZ= 3% UOS= 85% 	Option 20 AV= 0 NPZ= 16% UOS= 80% 

Table A.1.13. Expert MM1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
MM1	Experiment 1 Intervention 1	Option 1 AV= 2% NPZ= 51% UOS= 44% Rating= +3 	Option 4 AV= 2% NPZ= 51% UOS= 20 Rating= +2 	Option 8 AV= 45% NPZ= 28% UOS= 70% Rating= +2 
	Experiment 1 Intervention 2	Option 1 AV= 25% NPZ= 32% UOS= 77% Rating= +2 	Option 20 AV= 32% NPZ= 20% UOS= 82% Rating= +1 	Option 0 AV= 49% NPZ=0 UOS= 90% Rating= +1 
	Experiment 2 Intervention 1	Option 4 AV= 15% NPZ= 11% UOS= 77% 	Option 1 AV= 13% NPZ= 11% UOS= 83% 	Option 13 AV= 7% NPZ= 20% UOS= 87% 
	Experiment 2 Intervention 2	Option 14 AV= 48% NPZ= 0 UOS= 89%	Option 10 AV= 0 NPZ= 15% UOS= 82%	Option 12 AV= 7% NPZ= 0 UOS= 92%

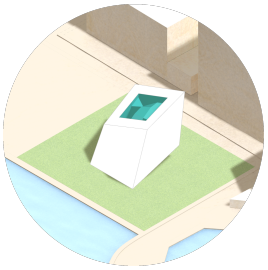
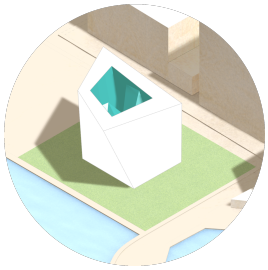
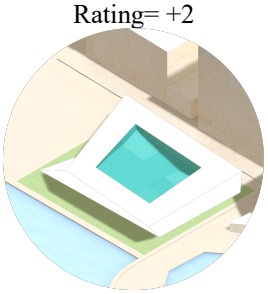
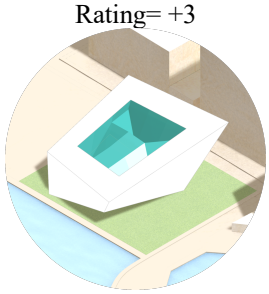
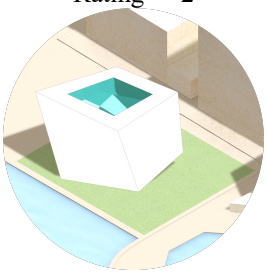
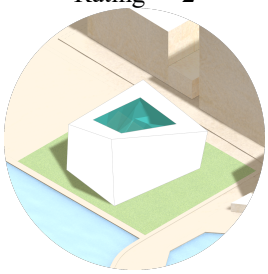
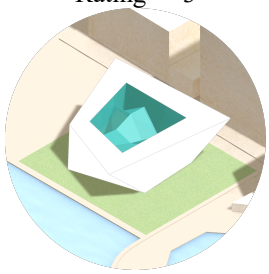
Participant	Experiment	Design Options		
				

Table A.1.14. Expert NN1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
NN1	Experiment 1 Intervention 1	Option 1 AV= 2% NPZ= 28% UOS= 36% Rating= +2 	Option 2 AV= 16% NPZ= 38% UOS= 67% Rating= +3 	Option 24 AV= 15% NPZ= 53% UOS= 73% Rating= +3 
	Experiment 1 Intervention 2	Option 11 AV= 21% NPZ= 7% UOS= 85% Rating= +2 	Option 1 AV= 8% NPZ= 40% UOS= 61% Rating= +2 	Option 3 AV= 20% NPZ= 37% UOS= 68% Rating= +3 
	Experiment 2 Intervention 1	Option 18 AV= 12% NPZ= 24% UOS= 89%	Option 15 AV= 6% NPZ= 25% UOS= 87%	Option 16 AV= 7% NPZ= 8% UOS= 88%

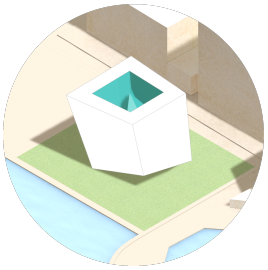
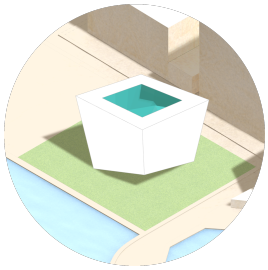
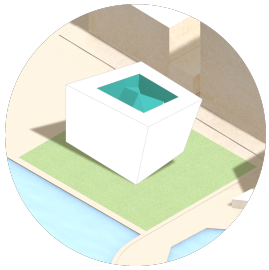
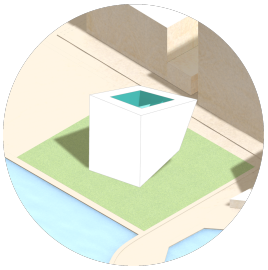
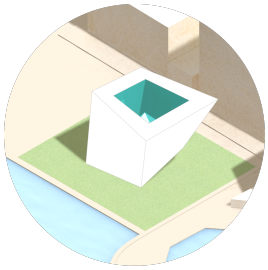
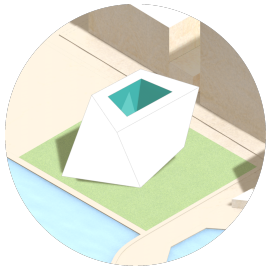
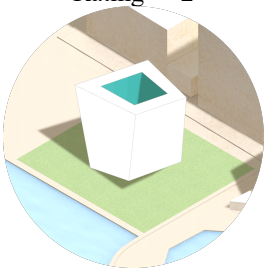
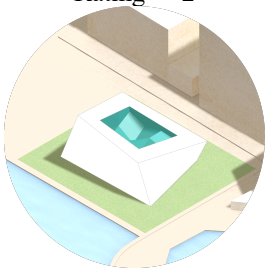
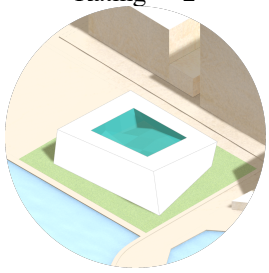
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 5 AV= 32% NPZ= 4% UOS= 89% 	Option 3 AV= 25% NPZ= 5% UOS= 88% 	Option 4 AV= 3% NPZ= 4% UOS= 85% 

Table A.1.15. Expert OO1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
OO1	Experiment 1 Intervention 1	Option 18 AV= 12% NPZ= 17% UOS= 88% Rating= +2 	Option 5 AV= 3% NPZ= 33% UOS= 64% Rating= +2 	Option 0 AV= 5% NPZ= 40% UOS= 54% Rating= +2 
	Experiment 1 Intervention 2	Option 16 AV= 2% NPZ= 6% UOS= 77% Rating= +3	Option 4 AV= 17% NPZ= 14% UOS= 85% Rating= +2	Option 10 AV= 28% NPZ= 2% UOS= 90% Rating= +2

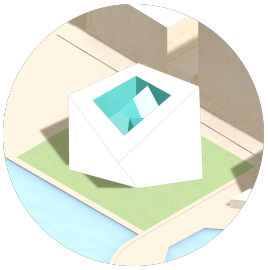
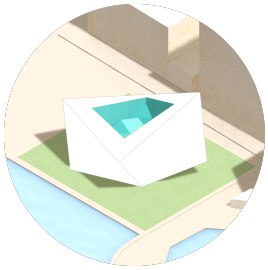
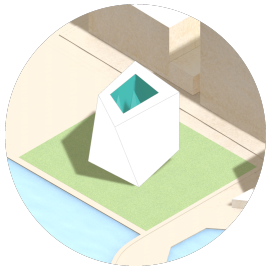
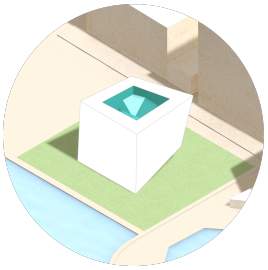
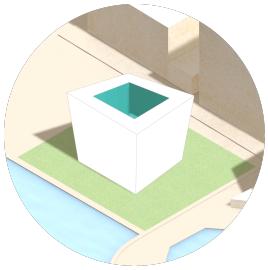
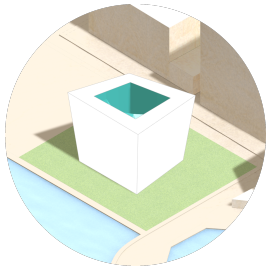
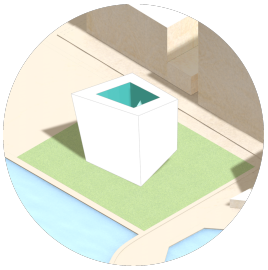
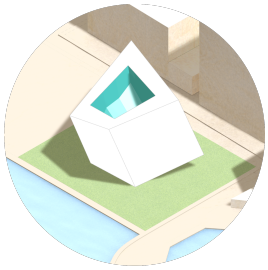
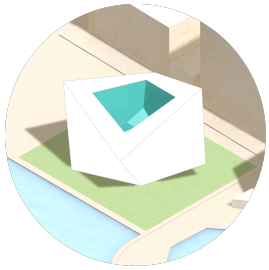
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 1	Option 14 AV= 12% NPZ= 5% UOS= 85% 	Option 13 AV= 1% NPZ= 24% UOS= 78% 	Option 26 AV= 1% NPZ= 24% UOS= 78% 
	Experiment 2 Intervention 2	Option 9 AV= 35% NPZ= 0 UOS= 90% 	Option 10 AV= 7% NPZ= 5% UOS= 88% 	Option 11 AV= 4% NPZ= 12% UOS= 86% 

Table A.1.16. Expert PP1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
PP1	Experiment 1 Intervention 1	Option 19 AV= 21% NPZ= 1% UOS= 86% Rating= +2	Option 5 AV= 5% NPZ= 38% UOS= 39% Rating= +2	Option 11 AV= 19% NPZ= 32% UOS= 76% Rating= +2

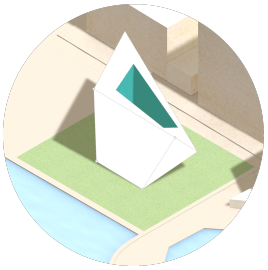
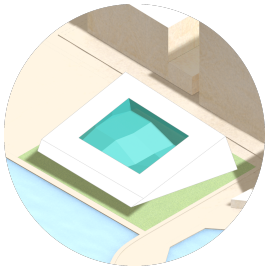
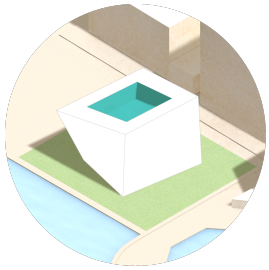
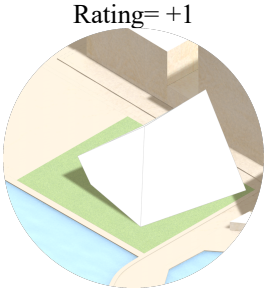
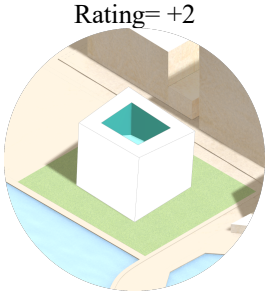
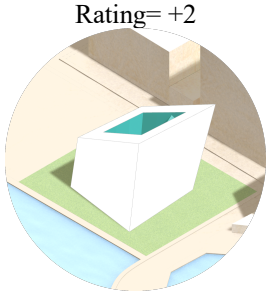
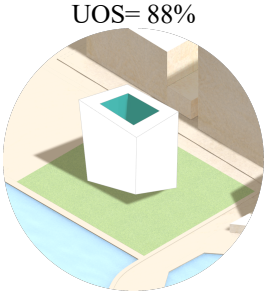
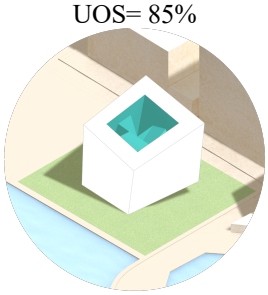
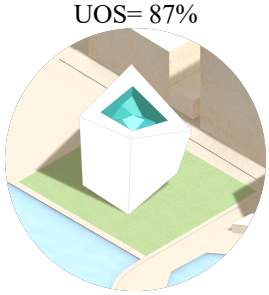
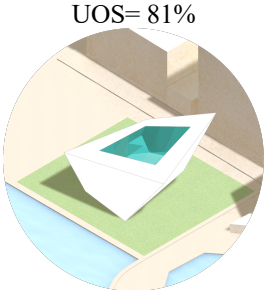
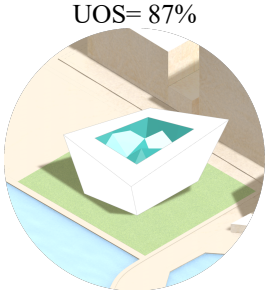
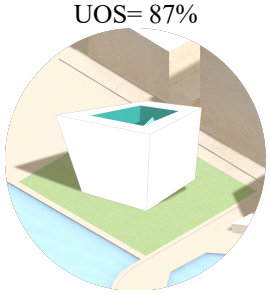
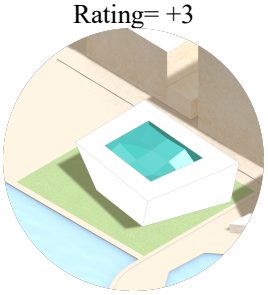
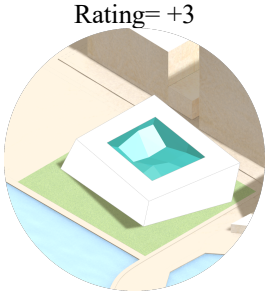
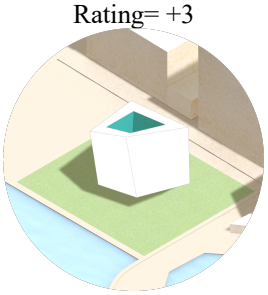
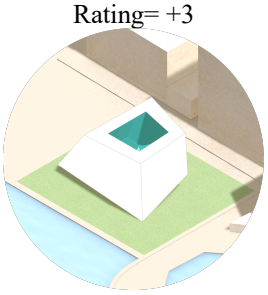
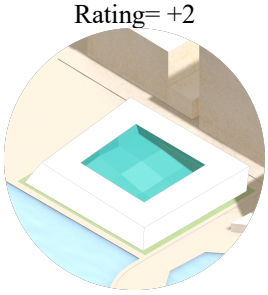
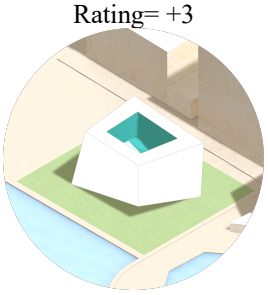
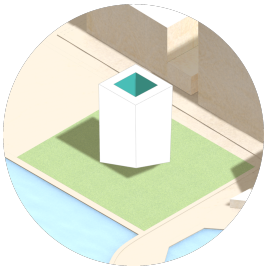
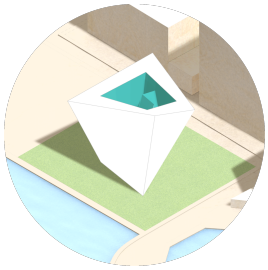
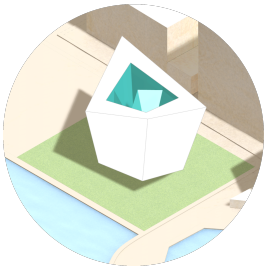
Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 9 AV= 2% NPZ= 11% UOS= 65% Rating= +1 	Option 6 AV=0 NPZ= 27% UOS= 75% Rating= +2 	Option 88 AV= 15% NPZ= 17% UOS= 73% Rating= +2 
	Experiment 2 Intervention 1	Option 9 AV= 17% NPZ= 15% UOS= 88% 	Option 11 AV= 3% NPZ= 9% UOS= 85% 	Option 12 AV= 8% NPZ=1% UOS= 87% 
	Experiment 2 Intervention 2	Option 0 AV= 7% NPZ= 5% UOS= 81% 	Option 7 AV= 13% NPZ= 7% UOS= 87% 	Option 12 AV= 1% NPZ= 6% UOS= 87% 

Table A.1.17. Expert QQ1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
QQ1	Experiment 1 Intervention 1	Option 6 AV= 0 NPZ= 50% UOS= 67% Rating= +3 	Option 4 AV= 4% NPZ= 51% UOS= 65% Rating= +3 	Option 2 AV= 27% NPZ= 23% UOS= 89% Rating= +3 
	Experiment 1 Intervention 2	Option 9 AV= 16% NPZ= 5% UOS= 83 Rating= +3 	Option 6 AV= 1% NPZ= 42% UOS= 26% Rating= +2 	Option 7 AV= 3% NPZ= 32% UOS= 80% Rating= +3 
	Experiment 2 Intervention 1	Option 6 AV= 51% NPZ= 9% UOS= 93% 	Option 26 AV= 18% NPZ= 3% UOS= 91% 	Option 27 AV= 2% NPZ= 10% UOS= 90% 
	Experiment 2 Intervention 2	Option 12 AV= 6% NPZ= 4% UOS= 85%	Option 13 AV= 2% NPZ= 30% UOS= 80%	Option 19 AV= 10% NPZ= 15% UOS= 86%

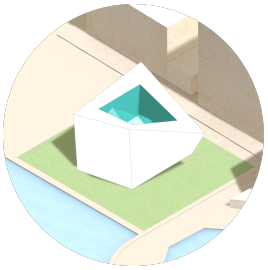
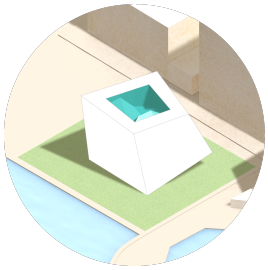
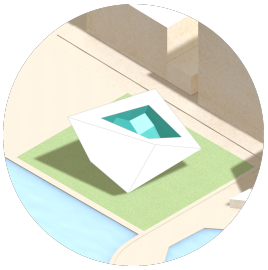
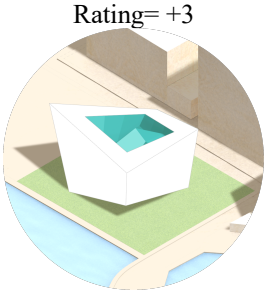
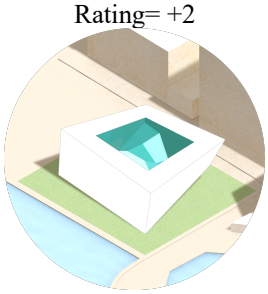
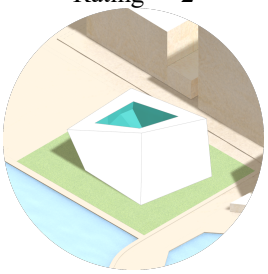
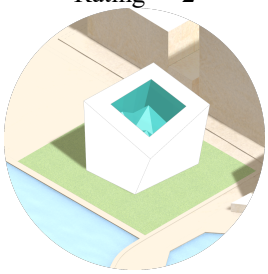
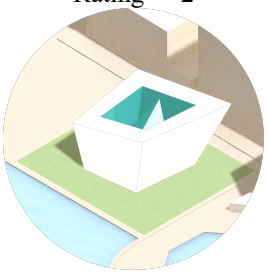
Participant	Experiment	Design Options		
				

Table A.1.18. Expert RR1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
RR1	Experiment 1 Intervention 1	Option 10 AV= 4% NPZ= 50% UOS= 80% Rating= +3 	Option 7 AV= 25% NPZ= 20% UOS= 69% Rating= +2 	Option 4 AV= 54% NPZ= 23% UOS= 46% Rating= +3 
	Experiment 1 Intervention 2	Option 1 AV= 11% NPZ= 8% UOS= 70% Rating= +2 	Option 6 AV= 3% NPZ= 24% UOS= 76% Rating= +2 	Option 28 AV= 23% NPZ= 20% UOS= 80% Rating= +2 
	Experiment 2 Intervention 1	Option 13 AV= 12% NPZ= 7% UOS= 83%	Option 14 AV= 4% NPZ= 8% UOS= 87%	Option 15 AV= 2% NPZ= 25% UOS= 83%

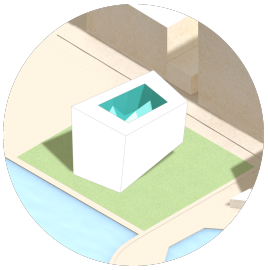
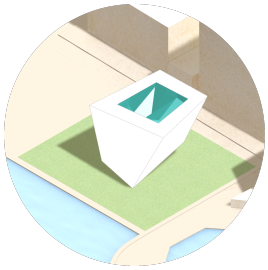
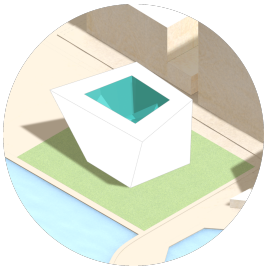
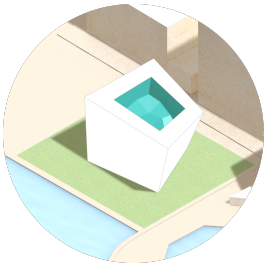
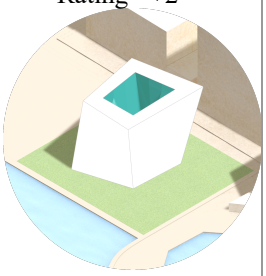
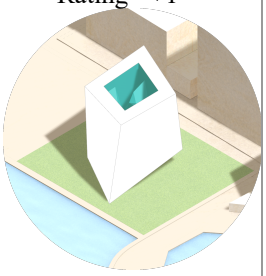
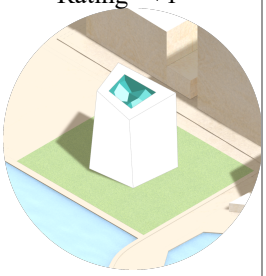
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 0 AV= 9% NPZ= 39% UOS= 88% 	Option 5 AV= 10% NPZ= 4% UOS= 83% 	Option 6 AV= 6% NPZ= 6% UOS= 87% 

Table A.1.19. Expert SS1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
SS1	Experiment 1 Intervention 1	Option 14 AV= 0 NPZ= 22% UOS= 80% Rating= +2 	Option 11 AV= 22% NPZ= 4% UOS= 81% Rating= +1 	Option 12 AV= 29% NPZ= 24% UOS= 87% Rating= +1 
	Experiment 1 Intervention 2	Option 0 AV= 2% NPZ= 35% UOS= 45% Rating= +2	Option 4 AV=1% NPZ= 48% UOS= 56 Rating= +2	Option 19 AV= 11% NPZ= 33% UOS= 81% Rating= +2

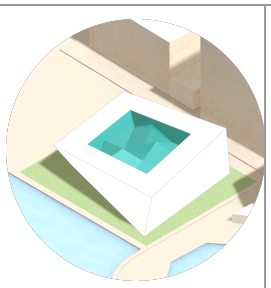
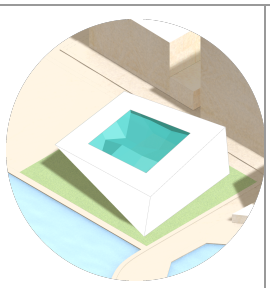
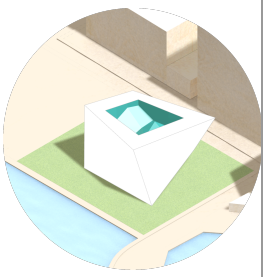
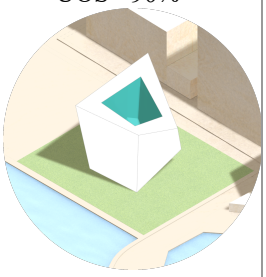
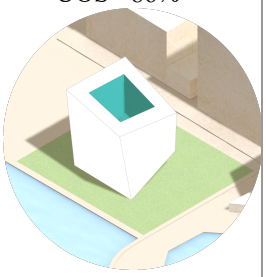
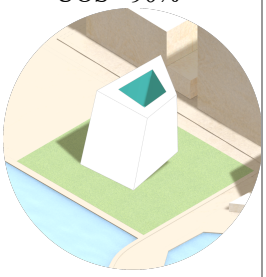
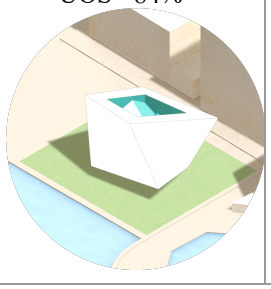
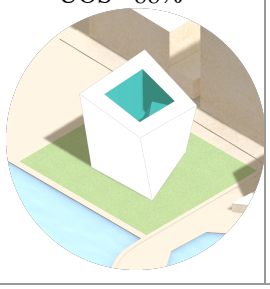
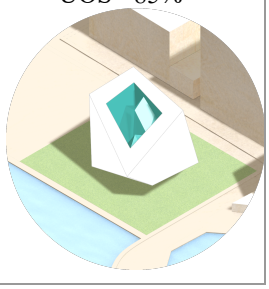
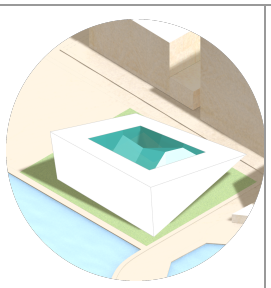
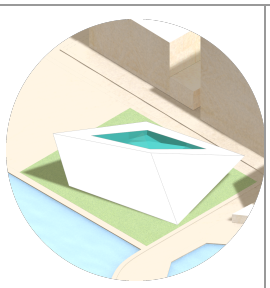
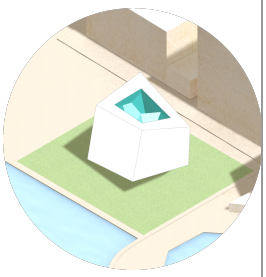
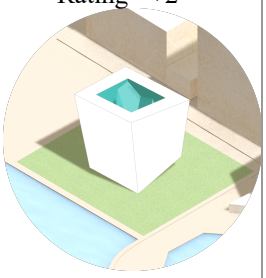
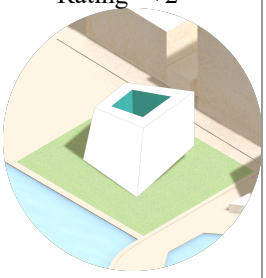
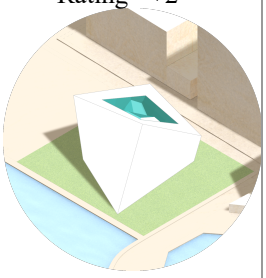
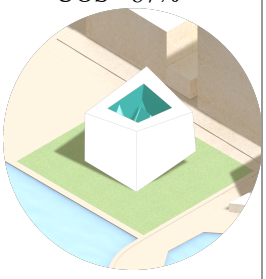
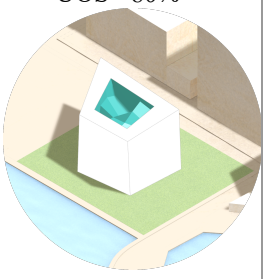
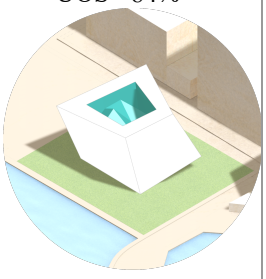
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 1	Option 5 AV= 34% NPZ= 1% UOS= 90% 	Option 6 AV= 10% NPZ= 6% UOS= 86% 	Option 7 AV= 20% NPZ= 4% UOS= 90% 
	Experiment 2 Intervention 2	Option 4 AV= 0% NPZ= 42% UOS= 84% 	Option 5 AV= 7% NPZ= 3% UOS= 88% 	Option 6 AV= 42% NPZ= 0% UOS= 85% 

Table A.1.20. Expert TT1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
TT1	Experiment 1 Intervention 1	Option 6 AV= 8% NPZ= 17% UOS= 54% Rating= +3	Option 14 AV= 9% NPZ= 27% UOS= 71% Rating= +3	Option 3 AV= 48% NPZ= 41% UOS= 86% Rating= +3

Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 12 AV= 15% NPZ= 8% UOS= 90% Rating= +2 	Option 4 AV= 16% NPZ= 7% UOS= 84% Rating= +2 	Option 6 AV= 8% NPZ= 41% UOS= 66% Rating= +2 
	Experiment 2 Intervention 1	Option 1 AV= 24% NPZ= 6% UOS= 87% 	Option 24 AV= 1% NPZ= 26% UOS= 80% 	Option 25 AV= 5% NPZ= 8% UOS= 84% 
	Experiment 2 Intervention 2	Option 0 AV= 12% NPZ= 0 UOS= 87%	Option 6 AV= 20% NPZ= 3% UOS= 90%	Option 9 AV= 9% NPZ= 9% UOS= 88%

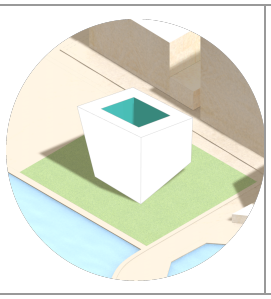
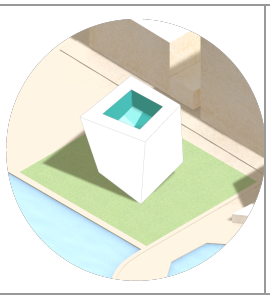
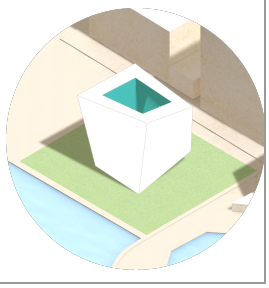
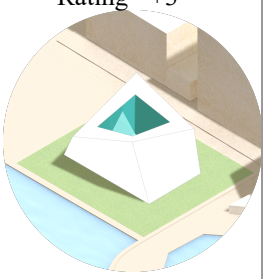
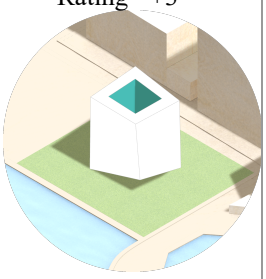
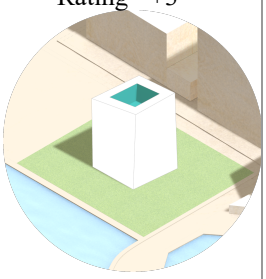
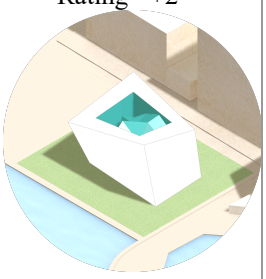
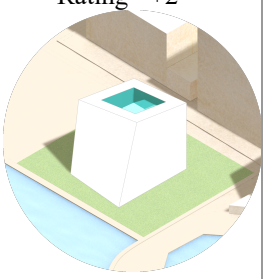
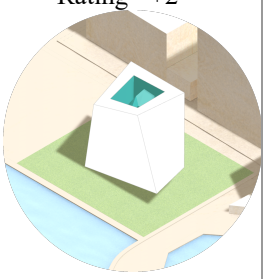
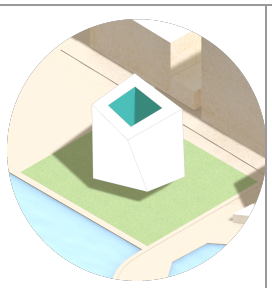
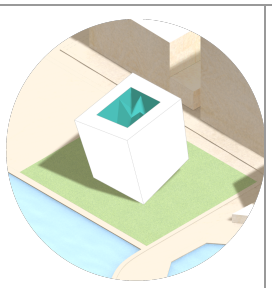
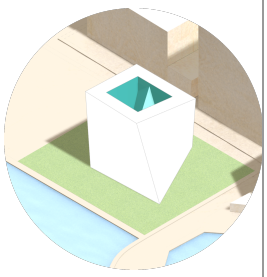
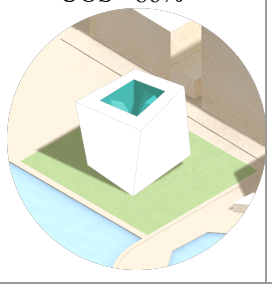
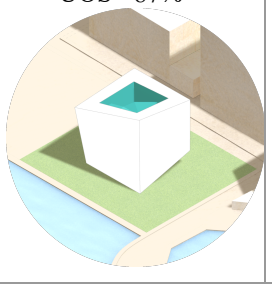
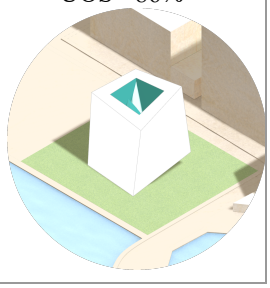
Participant	Experiment	Design Options		
				

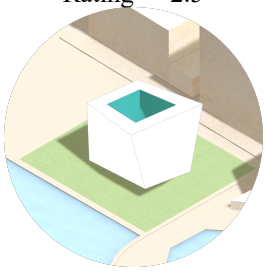
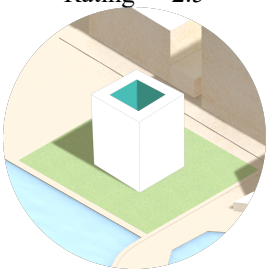
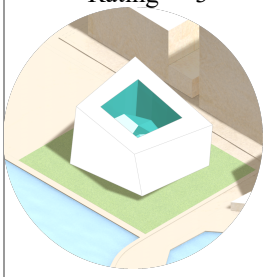
Table A.1.21. Expert UU1 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
UU1	Experiment 1 Intervention 1	Option 6 AV= 3% NPZ= 5% UOS= 73% Rating= +3 	Option 8 AV= 28% NPZ= 15% UOS= 89% Rating= +3 	Option 3 AV= 42% NPZ= 13% UOS= 87% Rating= +3 
	Experiment 1 Intervention 2	Option 2 AV=4% NPZ= 23% UOS= 77% Rating= +2 	Option 8 AV= 3% NPZ= 26% UOS= 75% Rating= +2 	Option 11 AV= 7% NPZ= 23% UOS= 85% Rating= +2 
	Experiment 2 Intervention 1	Option 16 AV= 18% NPZ= 19% UOS= 88%	Option 17 AV= 9% NPZ= 4% UOS= 88%	Option 19 AV= 6% NPZ= 8% UOS= 85%

Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 4 AV= 2% NPZ= 8% UOS= 86% 	Option 6 AV= 3% NPZ= 2% UOS= 87% 	Option 7 AV= 8% NPZ= 1% UOS= 86% 

A.2 Non-Expert Highly Rated Solutions

Table A.2.1. Non-Expert 2A2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2A2	Experiment 1 Intervention 1	Option 9 AV= 20% NPZ= 5% UOS= 84% Rating= +2.5 	Option 27 AV= 26% NPZ= 18% UOS= 85% Rating= +2.5 	Option 2 AV= 21% NPZ= 35% UOS= 78% Rating= +3 
	Experiment 1 Intervention 2	Option 5 AV= 12% NPZ= 10%	Option 19 AV= 3% NPZ= 29%	Option 0 AV= 0.5% NPZ= 53%

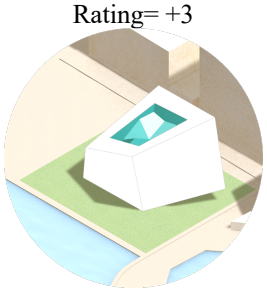
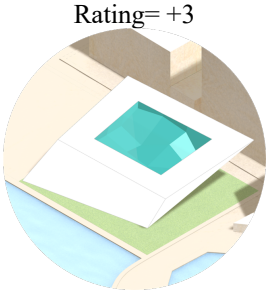
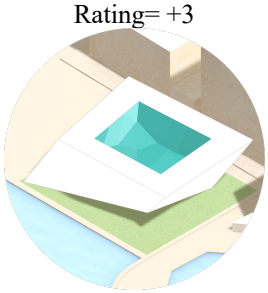
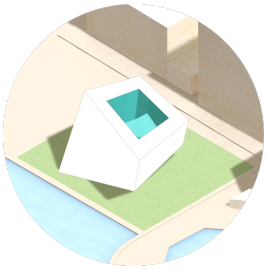
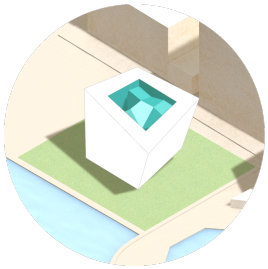
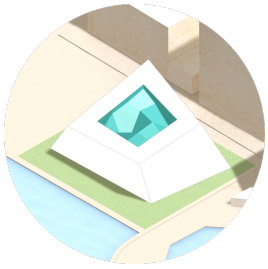
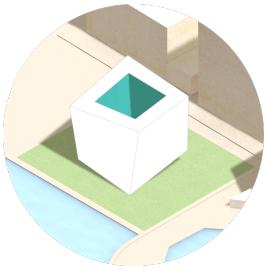
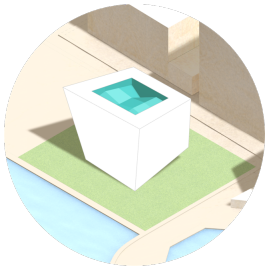
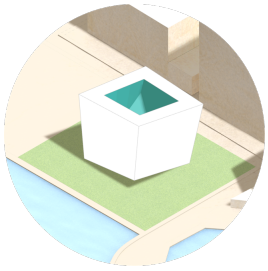
Participant	Experiment	Design Options		
		UOS= 82% Rating= +3 	UOS= 73% Rating= +3 	UOS= 71% Rating= +3 
	Experiment 2 Intervention 1	Option 2 AV= 0 NPZ= 6% UOS= 85% 	Option 3 AV=2% NPZ= 4% UOS= 88% 	Option 4 AV=0 NPZ= 44% UOS= 61% 
	Experiment 2 Intervention 2	Option 0 AV= 5% NPZ= 5% UOS= 86% 	Option 8 AV= 8% NPZ= 39% UOS= 86% 	Option 17 AV= 0 NPZ= 24% UOS= 84% 

Table A.2.2. Non-Expert 2B2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2B2	Experiment 1 Intervention 1	Option 4 AV= 2% NPZ= 8% UOS= 84% Rating= +2	Option 1 AV= 3% NPZ= 57% UOS= 45% Rating= +3	Option 13 AV= 15% NPZ= 11% UOS= 74% Rating= +3

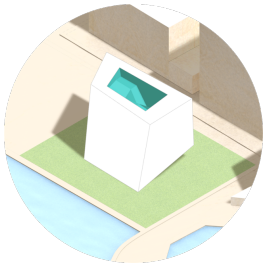
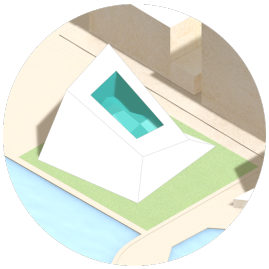
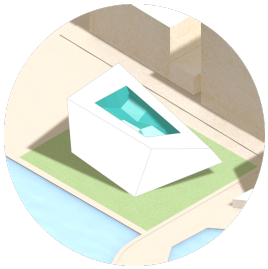
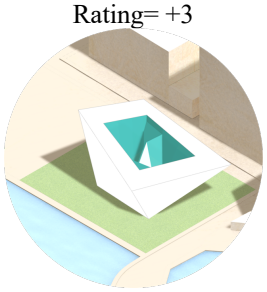
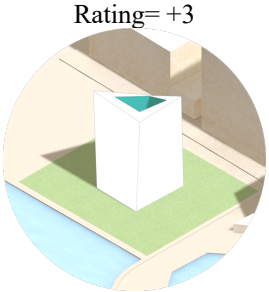
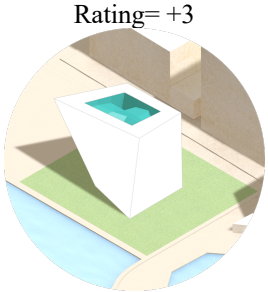
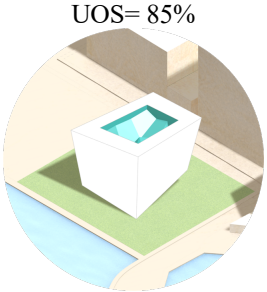
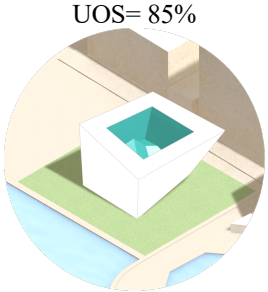
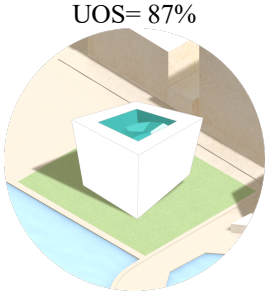
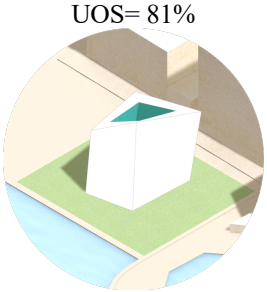
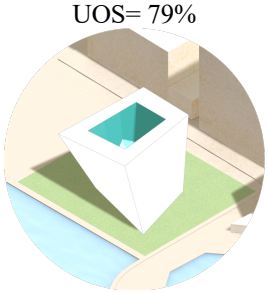
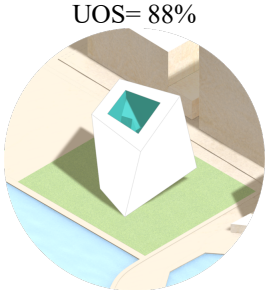
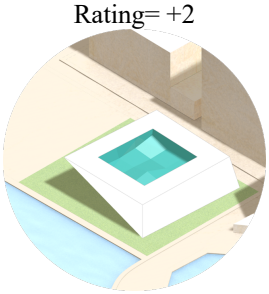
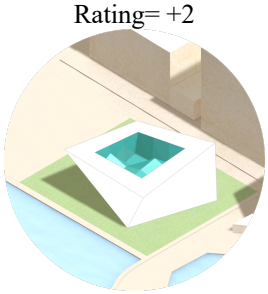
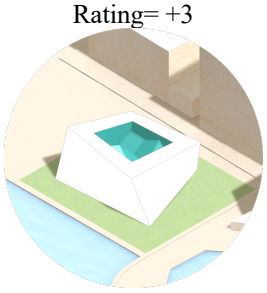
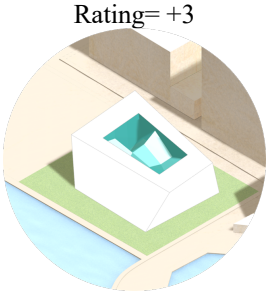
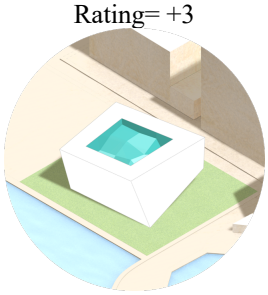
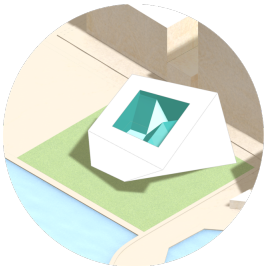
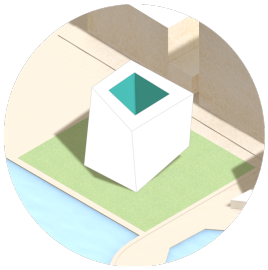
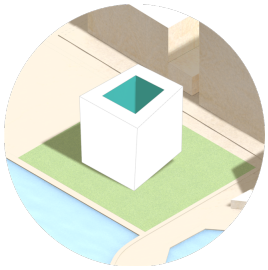
Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 8 AV= 9% NPZ= 2% UOS= 80% Rating= +3 	Option 2 AV= 30% NPZ= 16% UOS= 84% Rating= +3 	Option 20 AV= 4% NPZ= 37% UOS= 79% Rating= +3 
	Experiment 2 Intervention 1	Option 13 AV= 2% NPZ= 8% UOS= 85% 	Option 20 AV= 14% NPZ= 4% UOS= 85% 	Option 29 AV= 7% NPZ= 28% UOS= 87% 
	Experiment 2 Intervention 2	Option 1 AV= 23% NPZ= 0 UOS= 81% 	Option 6 AV= 6% NPZ= 0 UOS= 79% 	Option 17 AV= 3% NPZ= 6% UOS= 88% 

Table A.2.3. Non-Expert 2C2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2C2	Experiment 1 Intervention 1	Option 5 AV= 28% NPZ= 10% UOS= 85% Rating= +2 	Option 0 AV= 3% NPZ= 54% UOS= 56% Rating= +2 	Option 8 AV= 20% NPZ= 30% UOS= 70% Rating= +2 
	Experiment 1 Intervention 2	Option 25 AV= 12% NPZ= 12% UOS= 75% Rating= +3 	Option 6 AV= 10% NPZ= 22% UOS= 59% Rating= +3 	Option 11 AV= 14% NPZ= 13% UOS= 75% Rating= +3 
	Experiment 2 Intervention 1	Option 1 AV= 0 NPZ= 10% UOS= 77% 	Option 10 AV= 5% NPZ= 4% UOS= 86% 	Option 19 AV= 7% NPZ= 4% UOS= 87% 
	Experiment 2 Intervention 2	Option 11 AV= 19% NPZ= 3% UOS= 85%	Option 12 AV= 0 NPZ= 16% UOS= 80%	Option 13 AV= 9% NPZ= 5% UOS= 86%

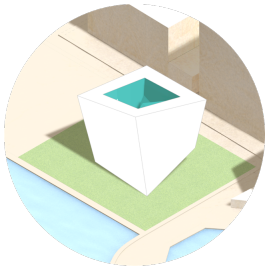
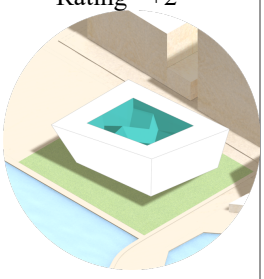
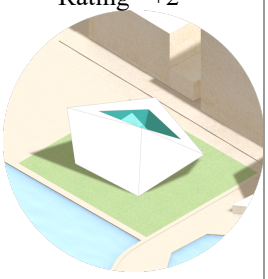
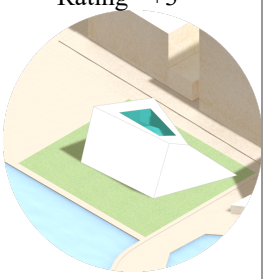
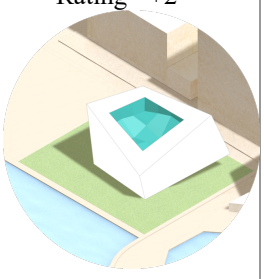
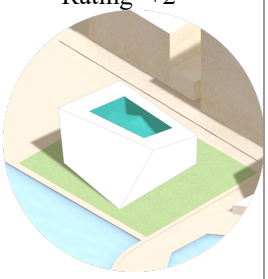
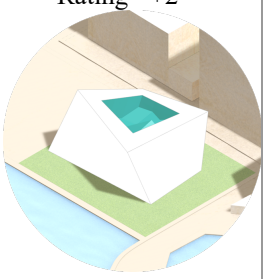
Participant	Experiment	Design Options		
				

Table A.2.4. Non-Expert 2D2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2D2	Experiment 1 Intervention 1	Option 7 AV= 5% NPZ= 17% UOS= 75% Rating= +2 	Option 11 AV= 28% NPZ= 19% UOS= 78% Rating= +2 	Option 6 AV= 30% NPZ= 35% UOS= 67% Rating= +3 
	Experiment 1 Intervention 2	Option 12 AV= 12% NPZ= 18% UOS= 81% Rating= +2 	Option 3 AV= 5% NPZ= 52% UOS= 68% Rating=+2 	Option 11 AV= 7% NPZ=49% UOS= 69% Rating= +2 
	Experiment 2 Intervention 1	Option 4 AV= 7% NPZ= 3% UOS= 82%	Option 20 AV= 42% NPZ= 0 UOS= 91%	Option 24 AV= 11% NPZ= 7% UOS= 87%

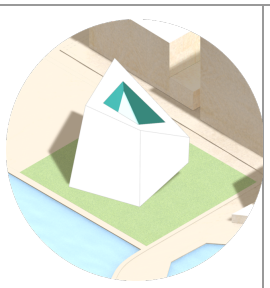
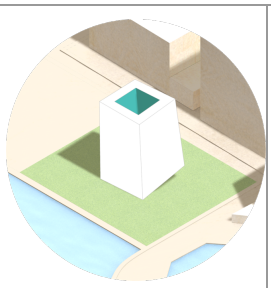
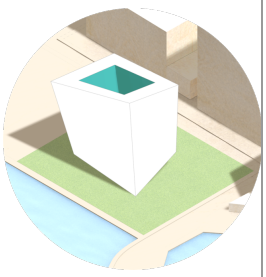
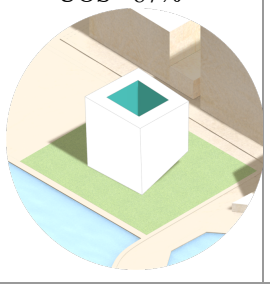
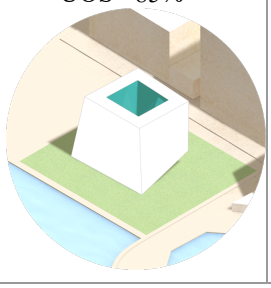
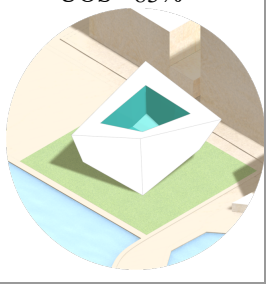
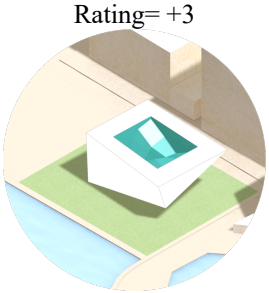
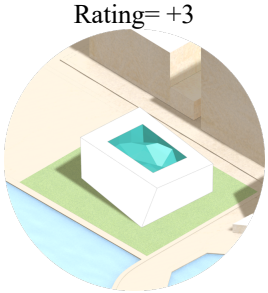
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 0 AV= 17% NPZ= 1% UOS= 87% 	Option 13 AV= 11% NPZ= 5% UOS= 85% 	Option 14 AV= 1% NPZ= 21% UOS= 83% 

Table A.2.5. Non-Expert 2E2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2E2	Experiment 1 Intervention 1	Option 13 AV= 1% NPZ= 22% UOS= 83% Rating= +3 	Option 15 AV= 20% NPZ= 21% UOS= 84% Rating= +3 	Option 12 AV= 32% NPZ= 20% UOS= 69% Rating= +3 
	Experiment 1 Intervention 2	Option 7 AV= 1% NPZ= 8% UOS= 84% Rating= +3	Option 11 AV= 19% NPZ= 37% UOS= 84 Rating= +2	Option 15 AV= 40% NPZ= 12% UOS= 88% Rating= +2

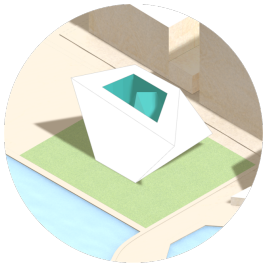
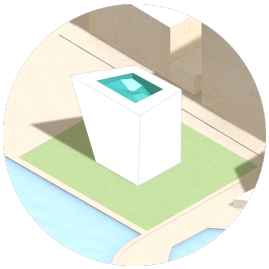
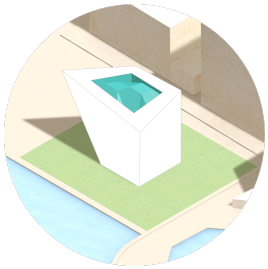
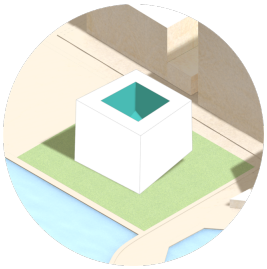
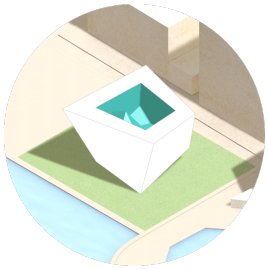
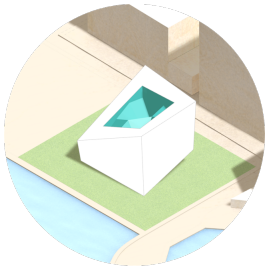
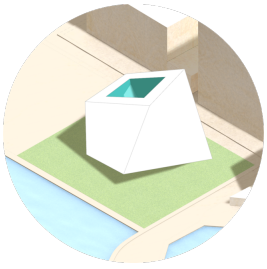
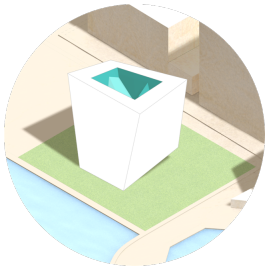
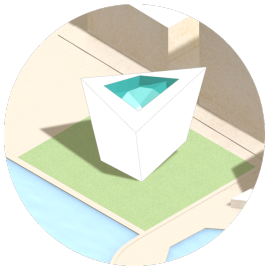
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 1	Option 12 AV= 0 NPZ= 7% UOS= 83% 	Option 13 AV= 6% NPZ= 7% UOS= 86% 	Option 14 AV= 43% NPZ= 5% UOS= 85% 
	Experiment 2 Intervention 2	Option 4 AV= 0 NPZ= 8% UOS= 82% 	Option 5 AV= 3% NPZ= 3% UOS= 89% 	Option 6 AV= 18% NPZ= 3% UOS= 90% 

Table A.2.6. Non-Expert 2G2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2G2	Experiment 1 Intervention 1	Option 0 AV= 2% NPZ= 49% UOS= 50% Rating= +2	Option 1 AV= 20% NPZ= 42% UOS= 32% Rating= +2	Option 7 AV= 33% NPZ= 38% UOS= 45% Rating= +2

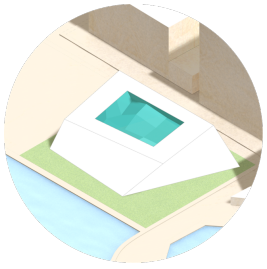
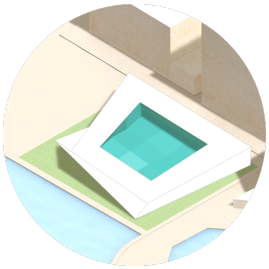
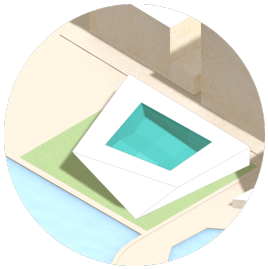
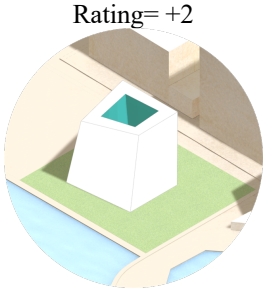
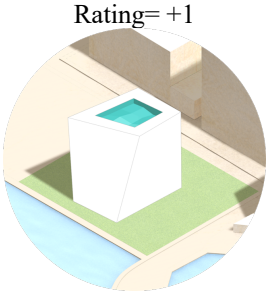
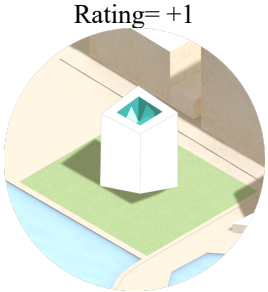
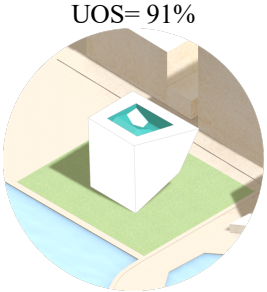
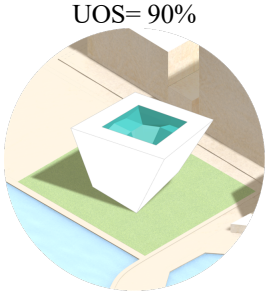
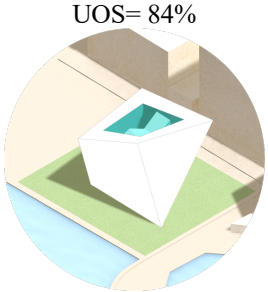
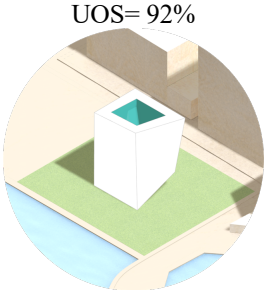
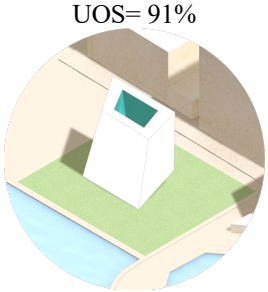
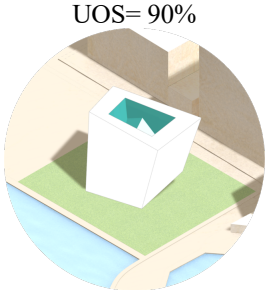
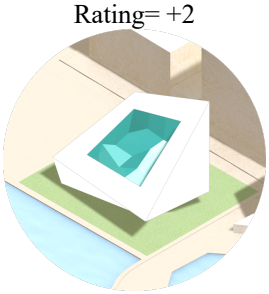
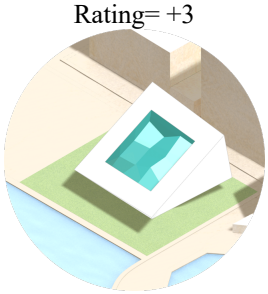
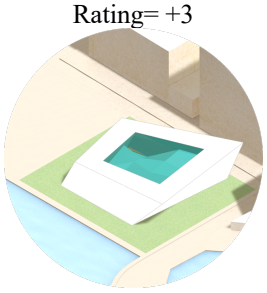
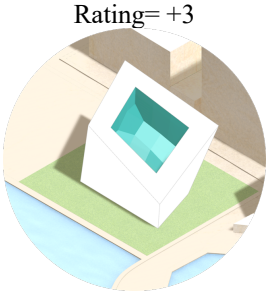
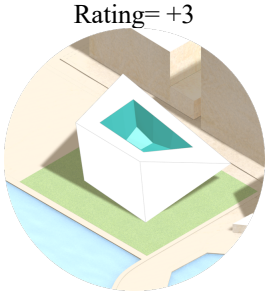
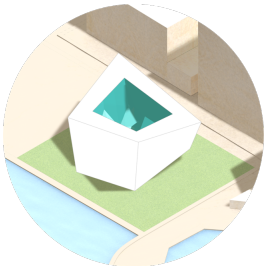
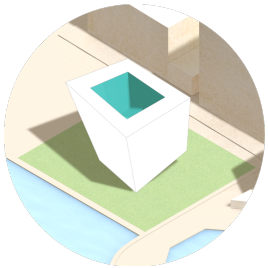
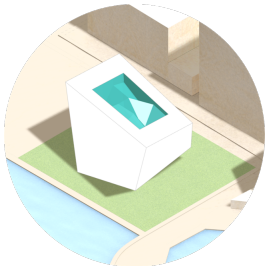
Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 1 AV= 13% NPZ= 19% UOS= 82% Rating= +2 	Option 8 AV= 11% NPZ= 42% UOS= 81% Rating= +1 	Option 5 AV= 46% NPZ= 13% UOS= 92% Rating= +1 
	Experiment 2 Intervention 1	Option 8 AV= 23% NPZ= 38% UOS= 91% 	Option 9 AV= 3% NPZ= 4% UOS= 90% 	Option 10 AV= 0 NPZ= 5% UOS= 84% 
	Experiment 2 Intervention 2	Option 15 AV= 38% NPZ= 0 UOS= 92% 	Option 20 AV= 32% NPZ= 2% UOS= 91% 	Option 21 AV= 11% NPZ= 11% UOS= 90% 

Table A.2.7. Non-Expert 2H2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2H2	Experiment 1 Intervention 1	Option 11 AV= 1% NPZ= 15% UOS= 77% Rating= +2 	Option 23 AV= 6% NPZ= 16% UOS= 79% Rating= +2 	Option 3 AV= 8% NPZ= 15% UOS= 79% Rating= +3 
	Experiment 1 Intervention 2	Option 19 AV= 3% NPZ= 19% UOS= 47% Rating= +3 	Option 1 AV= 2% NPZ= 45% UOS= 74% Rating= +3 	Option 11 AV= 16% NPZ= 43% UOS= 77% Rating= +3 
	Experiment 2 Intervention 1	Option 1 AV= 1% NPZ= 11% UOS= 80% 	Option 19 AV= 4% NPZ= 6% UOS= 89% 	Option 20 AV= 0 NPZ= 44% UOS= 86% 
	Experiment 2 Intervention 2	Option 6 AV= 1% NPZ= 5% UOS= 83%	Option 7 AV= 1% NPZ= 5% UOS= 83%	Option 8 AV= 29% NPZ= 1% UOS= 77%

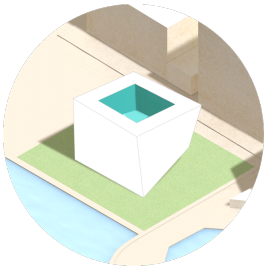
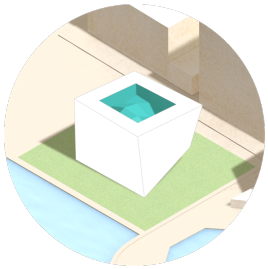
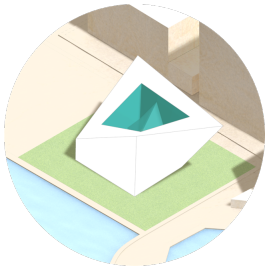
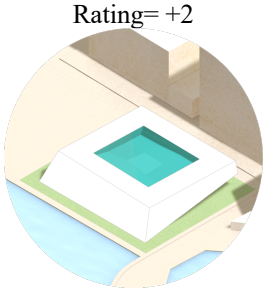
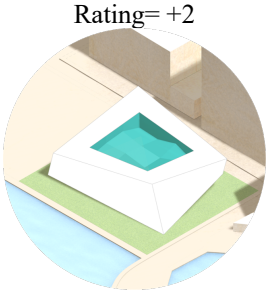
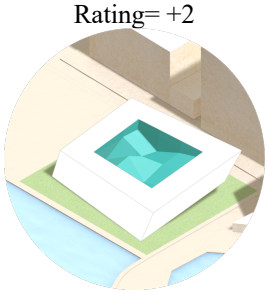
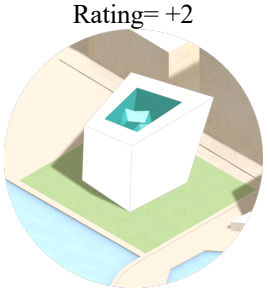
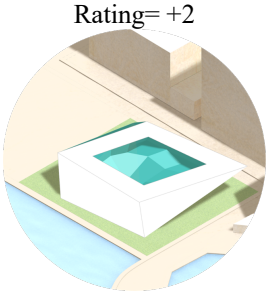
Participant	Experiment	Design Options		
				

Table A.2.8. Non-Expert 2I2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2I2	Experiment 1 Intervention 1	Option 4 AV= 5% NPZ= 29% UOS= 49% Rating= +2 	Option 16 AV= 18% NPZ= 38% UOS= 64% Rating= +2 	Option 19 AV= 28% NPZ= 26% UOS= 60 Rating= +2 
	Experiment 1 Intervention 2	Option 13 AV= 9% NPZ= 26% UOS= 80% Rating= +2 	Option 3 AV= 15% NPZ= 24% UOS= 59% Rating= +2 	Option 7 AV= 28 NPZ= 9% UOS= 84% Rating= +1 
	Experiment 2 Intervention 1	Option 10 AV= 6% NPZ= 12% UOS= 80%	Option 12 AV= 9% NPZ= 9% UOS= 84%	Option 13 AV= 22% NPZ= 17% UOS= 88%

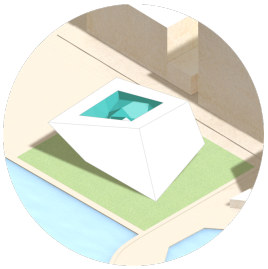
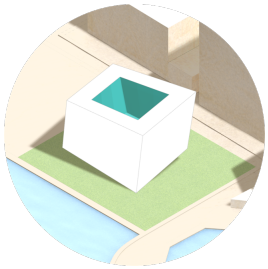
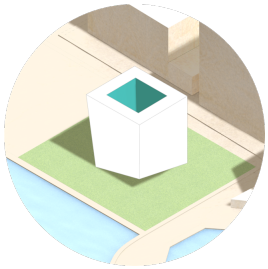
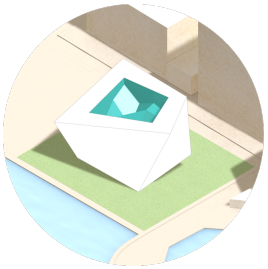
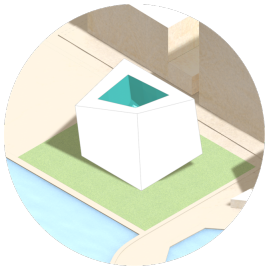
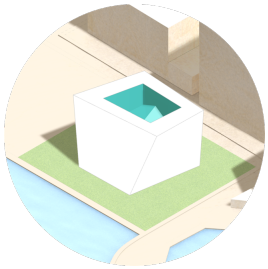
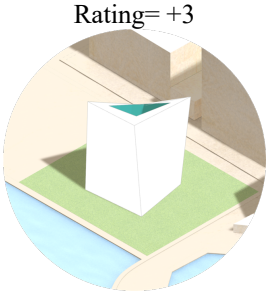
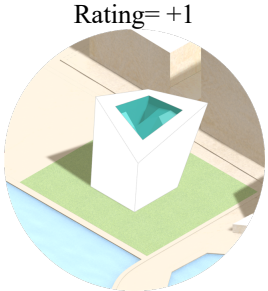
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 5 AV= 2% NPZ= 24% UOS= 86% 	Option 8 AV= 10% NPZ= 7% UOS= 85% 	Option 9 AV= 0 NPZ= 33% UOS= 75% 

Table A.2.9. Non-Expert 2J2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2J2	Experiment 1 Intervention 1	Option 8 AV= 25% NPZ= 3% UOS= 84% Rating= +2 	Option 5 AV= 18% NPZ= 28% UOS= 80% Rating= +3 	Option 28 AV= 34% NPZ= 38% UOS= 85% Rating= +1 
	Experiment 1 Intervention 2	Option 7 AV= 16% NPZ= 31% UOS= 73% Rating= +2	Option 16 AV= 7% NPZ= 44% UOS= 73% Rating= +3	Option 0 AV= 17% NPZ= 31% UOS= 73% Rating= +3

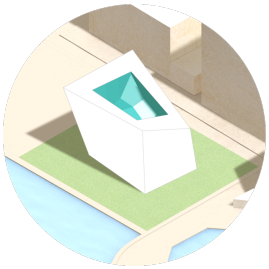
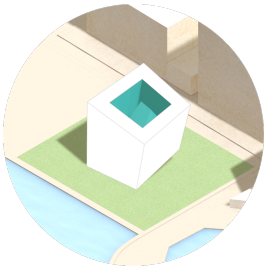
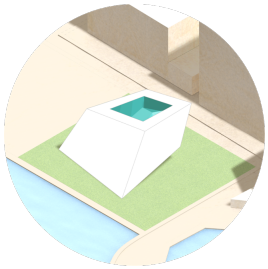
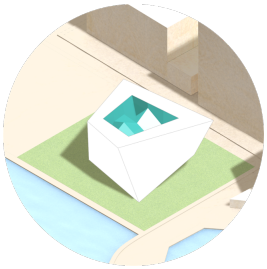
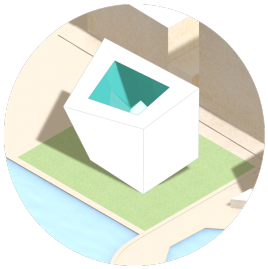
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 1	Option 19 AV= 32% NPZ= 0 UOS= 87% 	Option 20 AV= 12% NPZ= 11% UOS= 84% 	Option 21 AV= 1% NPZ= 25% UOS= 80% 
	Experiment 2 Intervention 2	Option 4 AV= 37% NPZ= 5% UOS= 82% 	Option 5 AV= 0 NPZ= 7% UOS= 84% 	Option 6 AV= 15% NPZ= 13% UOS= 84% 

Table A.2.10. Non-Expert 2K2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2K2	Experiment 1 Intervention 1	Option 10 AV=0 NPZ= 26% UOS= 73% Rating= 0	Option 7 AV= 15% NPZ= 35% UOS= 75% Rating= 0	Option 3 AV= 9% NPZ= 52% UOS= 74% Rating= 0

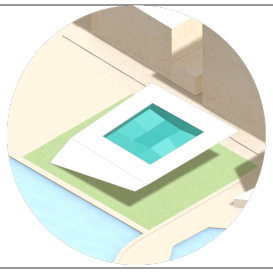
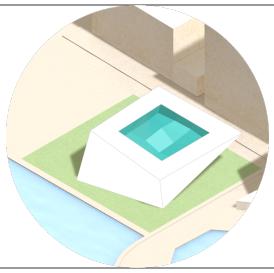
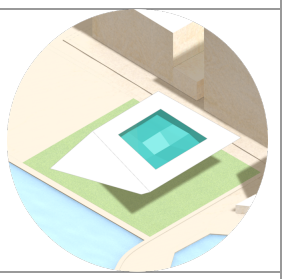
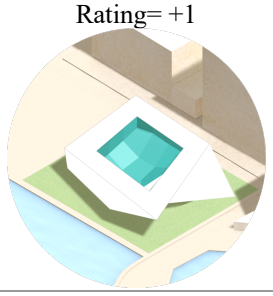
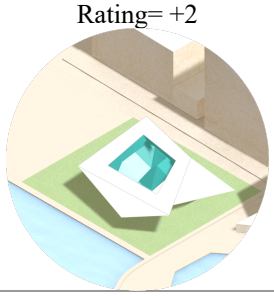
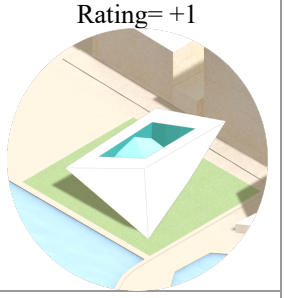
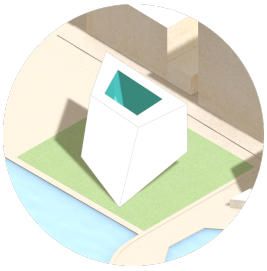
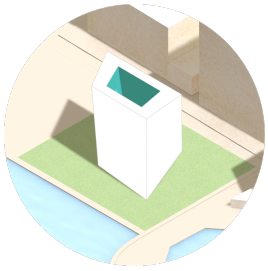
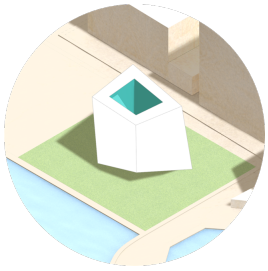
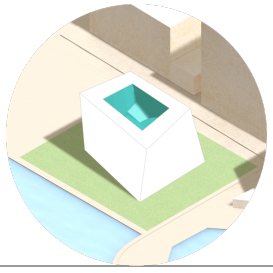
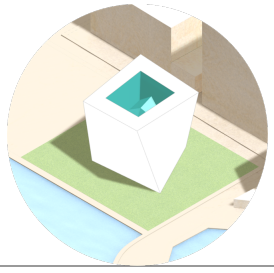
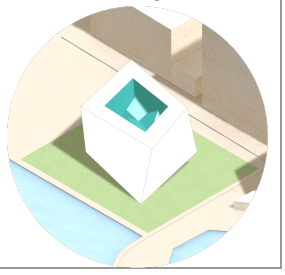
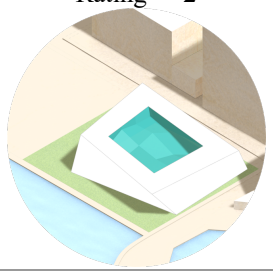
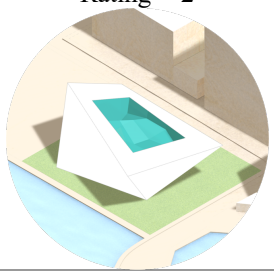
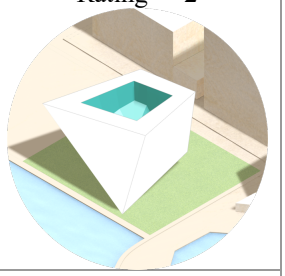
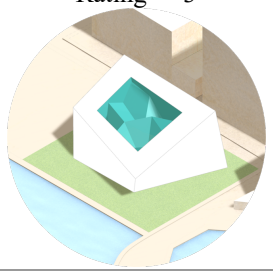
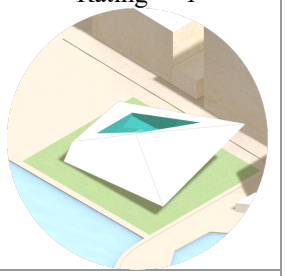
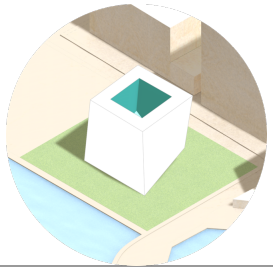
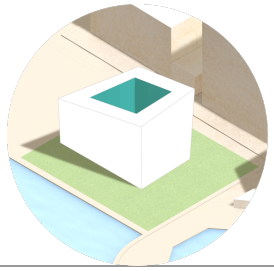
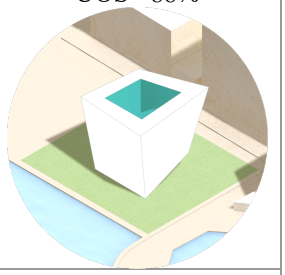
Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 10 AV= 14% NPZ= 42% UOS= 64% Rating= +1 	Option 1 AV= 39% NPZ= 21% UOS= 74% Rating= +2 	Option 14 AV= 35% NPZ= 45% UOS= 76% Rating= +1 
	Experiment 2 Intervention 1	Option 9 AV= 1% NPZ= 7% UOS= 83% 	Option 10 AV= 9% NPZ= 2% UOS= 88% 	Option 11 AV= 37% NPZ= 20% UOS= 90% 
	Experiment 2 Intervention 2	Option 6 AV= 1% NPZ= 32% UOS= 74% 	Option 7 AV= 2% NPZ= 5% UOS= 88% 	Option 8 AV= 2% NPZ= 9% UOS= 82% 

Table A.2.11. Non-Expert 2L2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2L2	Experiment 1 Intervention 1	Option 3 AV= 6% NPZ= 33% UOS= 49% Rating= +2 	Option 27 AV= 24% NPZ= 15% UOS= 79% Rating=+2 	Option 14 AV= 31% NPZ= 10% UOS= 84% Rating= +2 
	Experiment 1 Intervention 2	Option 1 AV= 4% NPZ= 13% UOS= 63% Rating= +3 	Option 10 AV= 14% NPZ= 17% UOS= 71% Rating= +2 	Option 15 AV= 7% NPZ= 36% UOS= 71% Rating= +1 
	Experiment 2 Intervention 1	Option 0 AV= 9% NPZ= 2% UOS= 86% 	Option 2 AV= 1% NPZ= 24% UOS= 78% 	Option 15 AV= 1% NPZ= 6% UOS= 88% 
	Experiment 2 Intervention 2	Option 2 AV= 25% NPZ= 15% UOS = 91%	Option 9 AV= 3% NPZ= 5% UOS= 90%	Option 15 AV= 25% NPZ= 15% UOS= 91%

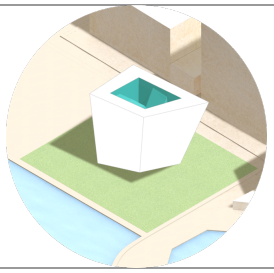
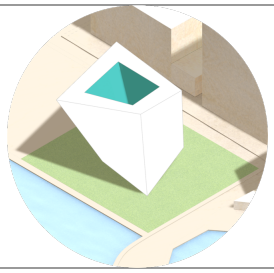
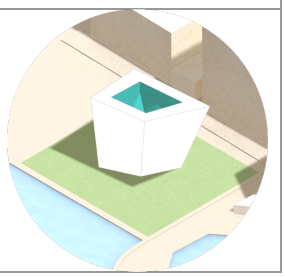
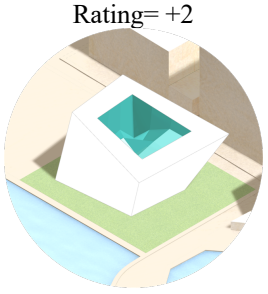
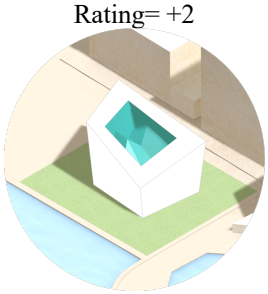
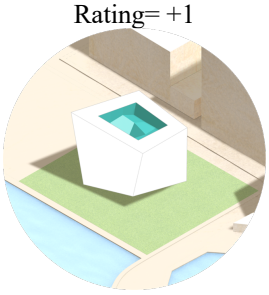
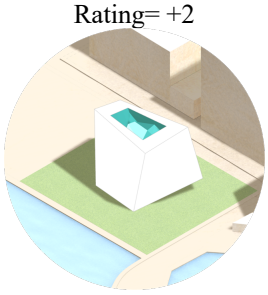
Participant	Experiment	Design Options		
				

Table A.2.12. Non-Expert 2M2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2M2	Experiment 1 Intervention 1	Option 7 AV=3% NPZ= 31% UOS= 69% Rating= +2 	Option 2 AV= 20% NPZ= 46% UOS= 74% Rating= +3 	Option 20 AV= 28% NPZ= 33% UOS= 76% Rating= +2 
	Experiment 1 Intervention 2	Option 7 AV= 1% NPZ= 7% UOS= 70% Rating= +1 	Option 2 AV=0 NPZ= 44% UOS= 86% Rating= +1 	Option 20 AV= 8% NPZ= 1% UOS= 83% Rating= +2 
	Experiment 2 Intervention 1	Option 1 AV= 0 NPZ= 41% UOS= 67%	Option 2 AV= 3% NPZ= 4% UOS= 84%	Option 3 AV= 10% NPZ= 4% UOS= 80%

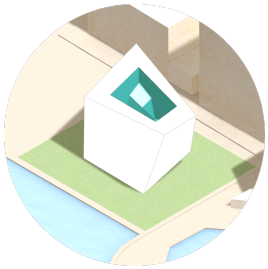
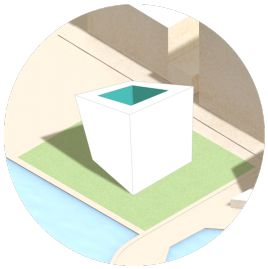
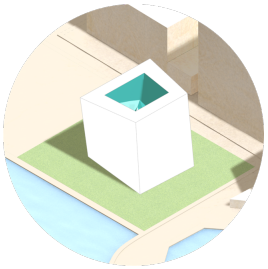
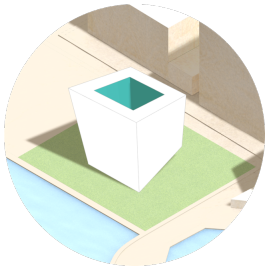
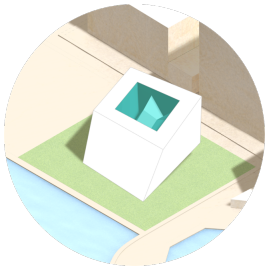
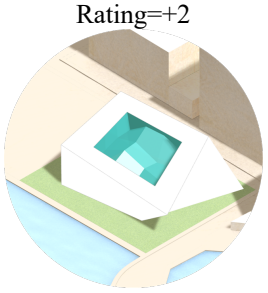
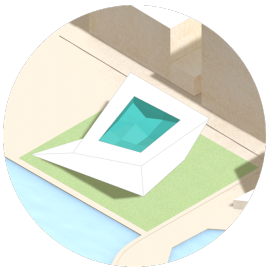
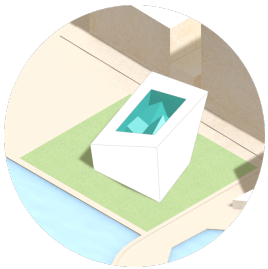
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 12 AV= 1% NPZ= 26% UOS= 78% 	Option 14 AV= 4% NPZ= 7% UOS= 87% 	Option 27 AV= 3% NPZ= 6% UOS= 82% 

Table A.2.13. Non-Expert 2N2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2N2	Experiment 1 Intervention 1	Option 3 AV= 4% NPZ= 55% UOS= 58% Rating=+2 	Option 8 AV= 27% NPZ= 46% Rating= +2 UOS= 66% 	Option 9 AV= 38% NPZ= 22% Rating= +2 UOS= 78% 
	Experiment 1 Intervention 2	Option 5 AV= 0 NPZ= 44% UOS= 72% Rating= +2	Option 14 AV= 9% NPZ= 39% UOS= 87% Rating= +2	Option 0 AV= 47% NPZ= 0.0 UOS= 82% Rating= +2

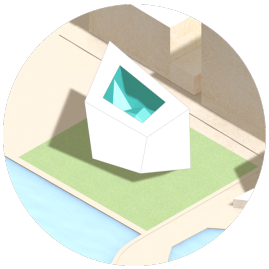
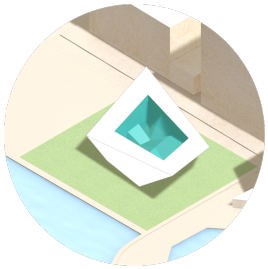
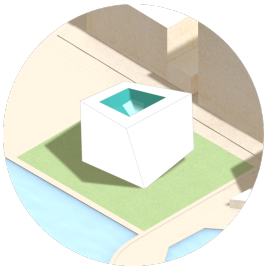
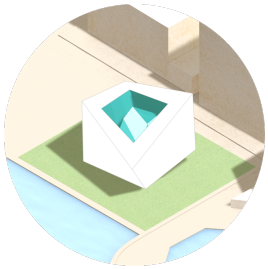
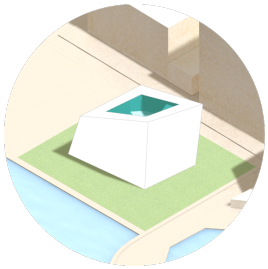
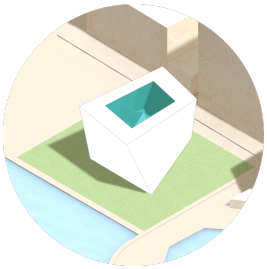
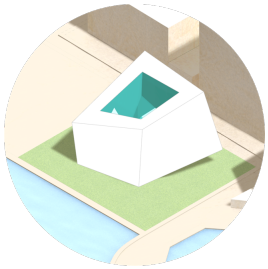
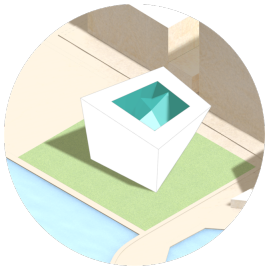
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 1	Option 10 AV= 5% NPZ= 22% UOS= 84% 	Option 12 AV= 3% NPZ= 6% UOS= 84% 	Option 13 AV= 0 NPZ= 32% UOS= 74% 
	Experiment 2 Intervention 2	Option 1 AV= 1% NPZ= 4% UOS= 83% 	Option 3 AV= 1% NPZ= 12% UOS= 80% 	Option 10 AV= 4% NPZ= 9% UOS= 84% 

Table A.2.14. Non-Expert 2O2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2O2	Experiment 1 Intervention 1	Option 18 AV= 14% NPZ= 12% UOS= 60% Rating= +3	Option 3 AV= 15% NPZ= 10% UOS= 80% Rating= +2	Option 4 AV= 7% NPZ= 45% UOS= 88% Rating= +2

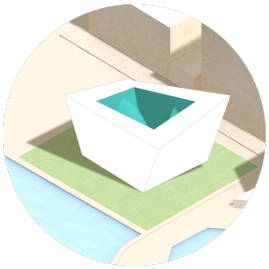
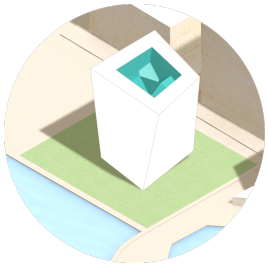
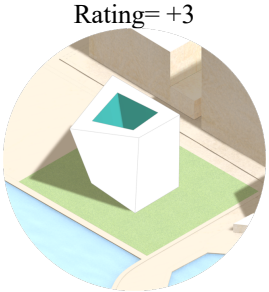
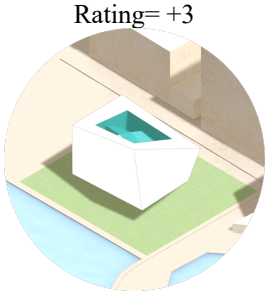
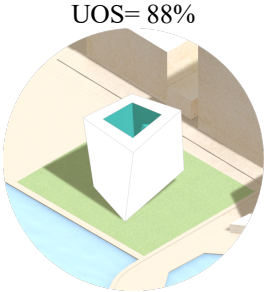
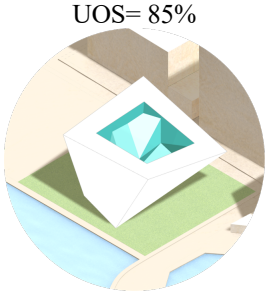
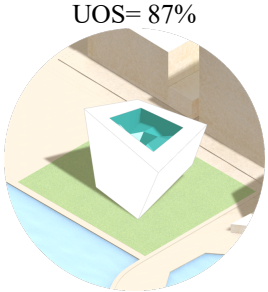
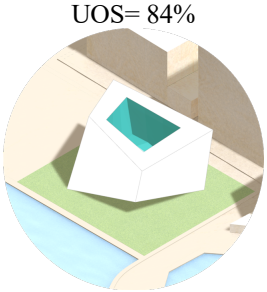
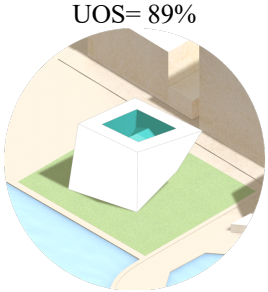
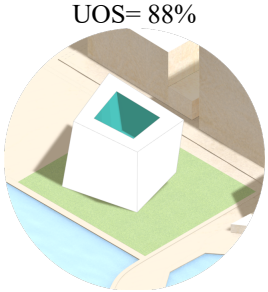
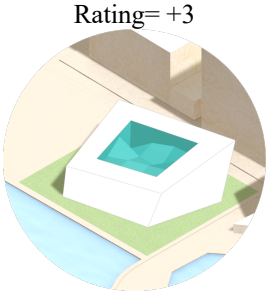
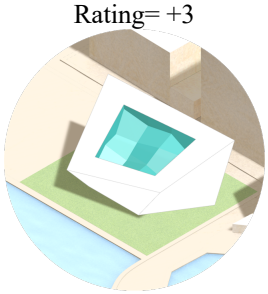
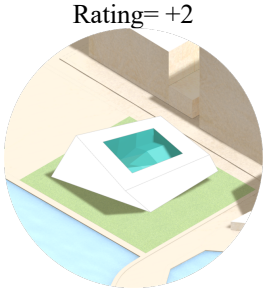
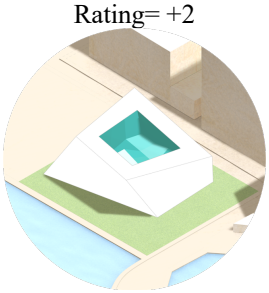
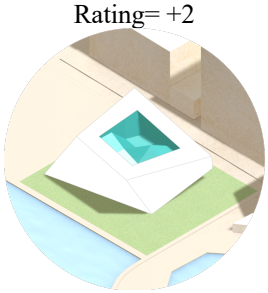
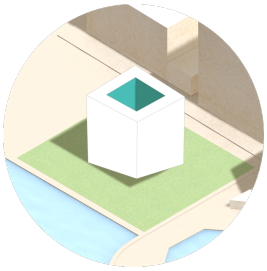
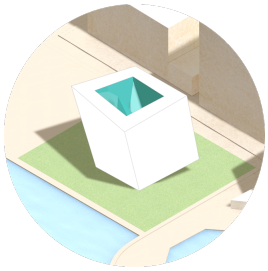
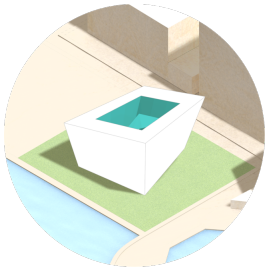
Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 10 AV= 5% NPZ= 32% UOS= 71% Rating= +3 	Option 19 AV= 11% NPZ= 23% UOS= 85% Rating= +3 	Option 6 AV= 19% NPZ= 28% UOS= 81% Rating= +3 
	Experiment 2 Intervention 1	Option 4 AV= 20% NPZ= 2% UOS= 88% 	Option 5 AV= 5% NPZ= 9% UOS= 85% 	Option 6 AV= 10% NPZ= 3% UOS= 87% 
	Experiment 2 Intervention 2	Option 4 AV= 0 NPZ= 31% UOS= 84% 	Option 16 1AV= 11% NPZ= 3% UOS= 89% 	Option 18 AV= 7% NPZ= 7% UOS= 88% 

Table A.2.15. Non-Expert 2P2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2P2	Experiment 1 Intervention 1	Option 6 AV= 0 NPZ= 53% UOS= 55% Rating= +2 	Option 21 AV= 3% NPZ= 48% UOS= 63% Rating= +3 	Option 4 AV= 5% NPZ= 46% UOS= 75% Rating= +3 
	Experiment 1 Intervention 2	Option 11 AV= 11% NPZ= 14% UOS= 76% Rating= +2 	Option 12 AV= 6% NPZ= 52% UOS= 67% Rating= +2 	Option 3 AV= 9% NPZ= 52% UOS= 78% Rating= +2 
	Experiment 2 Intervention 1	Option 0 AV= 25% NPZ= 20% UOS= 88% 	Option 9 AV= 9% NPZ= 4% UOS= 88% 	Option 14 AV= 0 NPZ= 6% UOS= 84% 
	Experiment 2 Intervention 2	Option 3 AV= 0 NPZ= 38% UOS= 87%	Option 4 AV= 1% NPZ= 6% UOS= 88%	Option 7 AV= 5% NPZ= 2% UOS= 89%

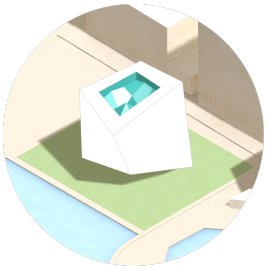
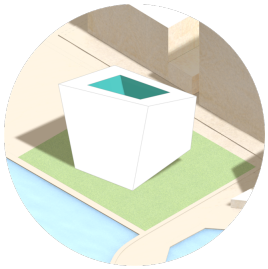
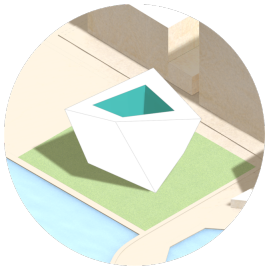
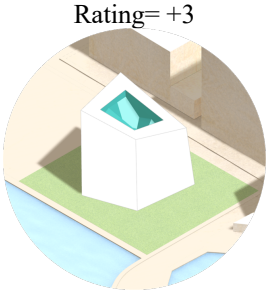
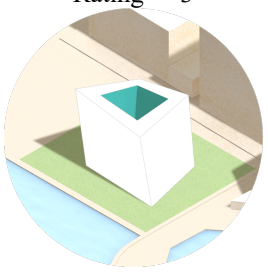
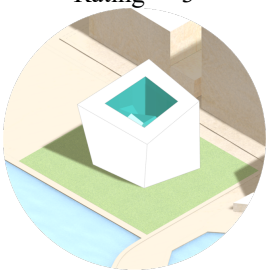
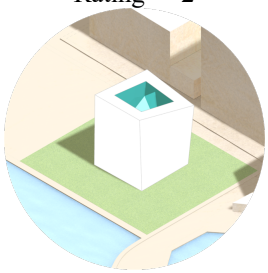
Participant	Experiment	Design Options		
				

Table A.2.16. Non-Expert 2Q2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2Q2	Experiment 1 Intervention 1	Option 13 AV= 5% NPZ= 18% UOS= 79% Rating= +3 	Option 15 AV= 9% NPZ= 33% UOS= 81% Rating= +3 	Option 5 AV= 6% NPZ= 44% UOS= 80% Rating= +3 
	Experiment 1 Intervention 2	Option 5 AV= 4% NPZ= 24% UOS= 75% Rating= +3 	Option 14 AV= 19% NPZ= 22% UOS= 85% Rating= +3 	Option 6 AV= 31% NPZ= 17% UOS= 85% Rating= +2 
	Experiment 2 Intervention 1	Option 12 AV= 6% NPZ= 3% UOS= 88%	Option 13 AV= 10% NPZ= 8% UOS= 88%	Option 14 AV= 1% NPZ= 3% UOS= 86%

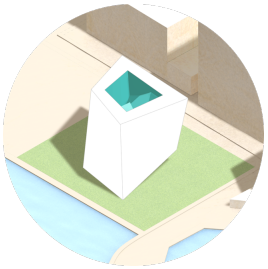
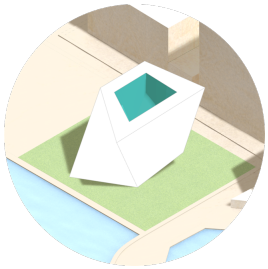
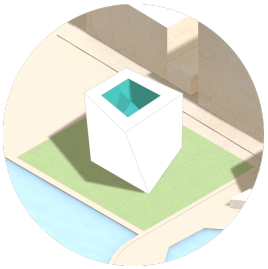
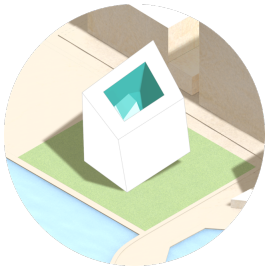
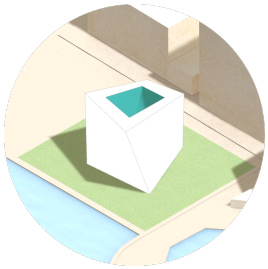
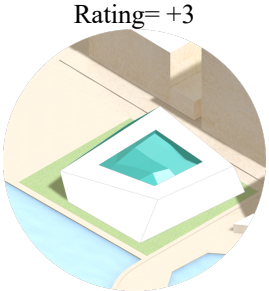
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 14 AV= 11% NPZ= 5% UOS= 88% 	Option 15 AV= 6% NPZ= 6% UOS= 86% 	Option 21 AV= 18% NPZ= 6% UOS= 88% 

Table A.2.17. Non-Expert 2R2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2R2	Experiment 1 Intervention 1	Option 14 AV= 15% NPZ= 20% UOS= 66% Rating= +3 	Option 4 AV= 5% NPZ= 47% UOS= 36% Rating= +3 	Option 1 AV= 23% NPZ= 53% UOS= 72% Rating= +3 
	Experiment 1 Intervention 2	Option 2 AV= 7% NPZ= 9% UOS= 75% Rating= +1	Option 15 AV=0 NPZ= 35% UOS= 66% Rating= +1	Option 14 AV= 16% NPZ=18% UOS= 88% Rating= +1

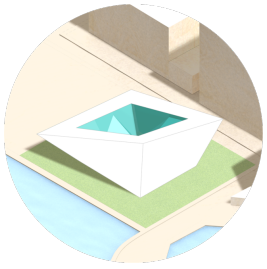
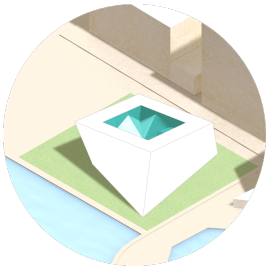
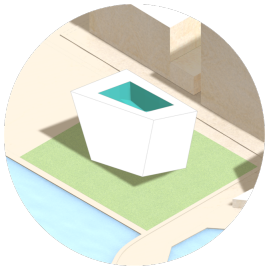
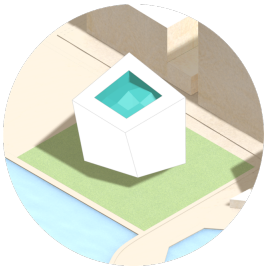
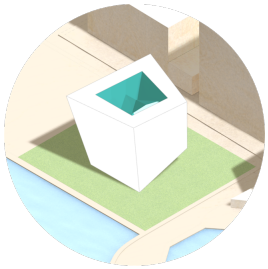
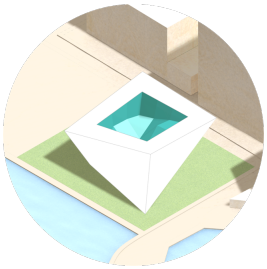
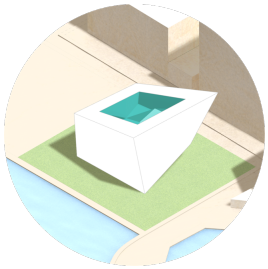
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 1	Option 1 AV= 3% NPZ= 44% UOS= 87% 	Option 24 AV= 8% NPZ= 4% UOS= 87% 	Option 25 AV= 5% NPZ= 5% UOS= 86% 
	Experiment 2 Intervention 2	Option 12 AV= 2% NPZ= 8% UOS= 82% 	Option 14 AV= 3% NPZ= 9% UOS= 85% 	Option 15 AV= 0 NPZ= 42% UOS= 77% 

Table A.2.18. Non-Expert 2S2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2S2	Experiment 1 Intervention 1	Option 3 AV= 5% NPZ= 45% UOS= 44% Rating= +3	Option 5 AV= 51% NPZ= 52% UOS= 63% Rating= +3	Option 4 AV= 5-% NPZ= 65% UOS= 49% Rating= +2

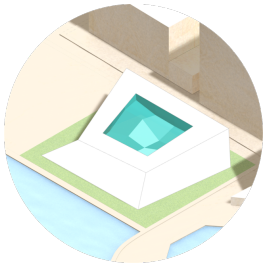
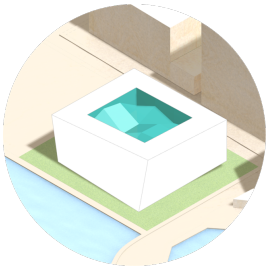
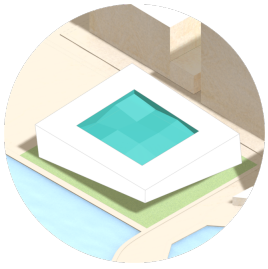
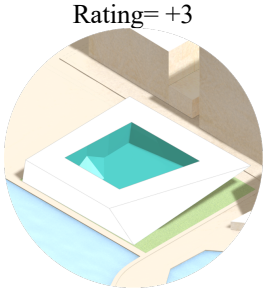
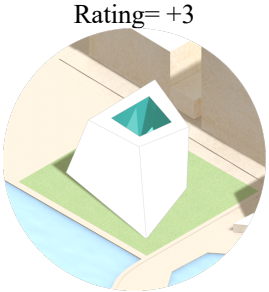
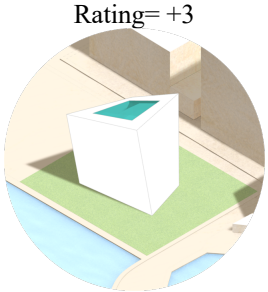
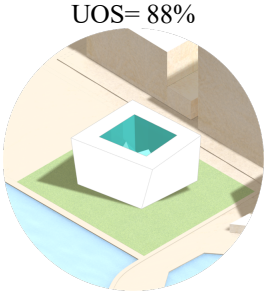
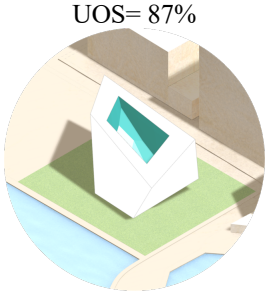
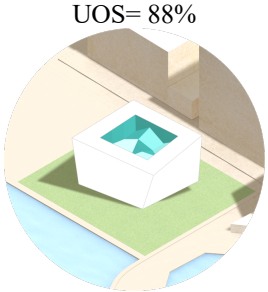
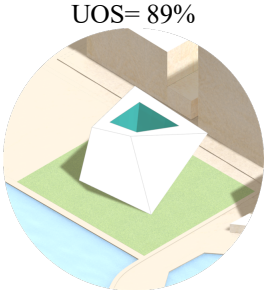
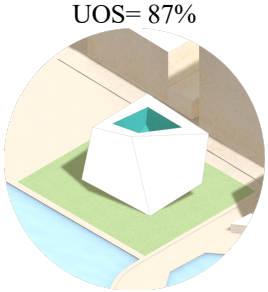
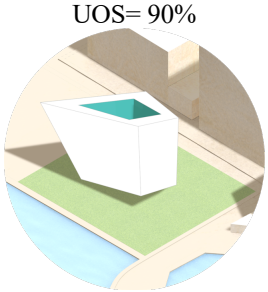
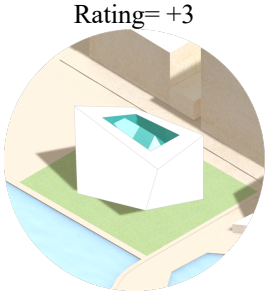
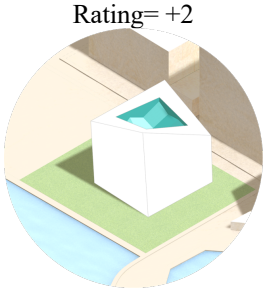
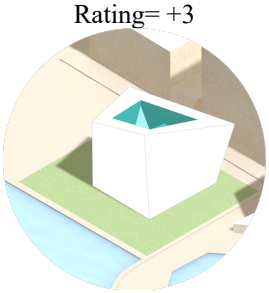
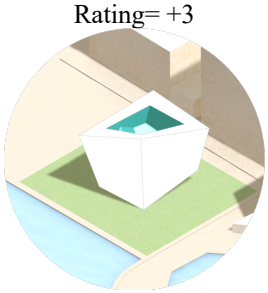
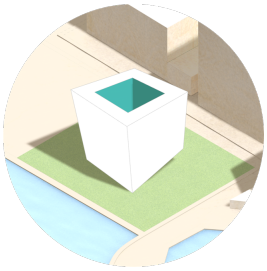
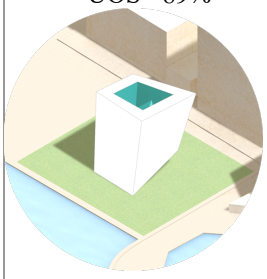
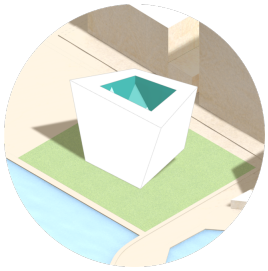
Participant	Experiment	Design Options		
				
	Experiment 1 Intervention 2	Option 1 AV= 8% NPZ= 33% UOS= 43% Rating= +3 	Option 14 AV= 11% NPZ= 27% UOS= 70% Rating= +3 	Option 12 AV= 14% NPZ= 37% UOS= 84% Rating= +3 
	Experiment 2 Intervention 1	Option 6 AV= 16% NPZ= 3% UOS= 88% 	Option 7 AV= 0 NPZ= 5% UOS= 87% 	Option 8 AV= 15% NPZ= 11% UOS= 88% 
	Experiment 2 Intervention 2	Option 11 AV= 11% NPZ= 7% UOS= 89% 	Option 12 AV= 1% NPZ= 5% UOS= 87% 	Option 13 AV= 4% NPZ= 28% UOS= 90% 

Table A.2.19. Non-Expert 2T2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2T2	Experiment 1 Intervention 1	Option 7 AV= 7% NPZ= 22% UOS= 39% Rating= +2 	Option 16 AV= 2% NPZ= 34% UOS= 75% Rating= +3 	Option 5 AV= 7% NPZ= 32% UOS= 72% Rating= +3 
	Experiment 1 Intervention 2	Option 10 AV= 3% NPZ= 25% UOS= 80% Rating= +2 	Option 6 AV= 10% NPZ= 20% UOS= 77% Rating= +3 	Option 8 AV= 12% NPZ= 20% UOS= 88% Rating= +3 
	Experiment 2 Intervention 1	Option 7 AV= 0 NPZ= 4% UOS= 88% 	Option 11 AV= 19% NPZ= 2% UOS= 89% 	Option 13 AV= 3% NPZ= 6% UOS= 88% 
	Experiment 2 Intervention 2	Option 3 AV= 11% NPZ= 0 UOS= 90%	Option 4 AV= 5% NPZ= 4% UOS= 90%	Option 7 AV= 0 NPZ= 13% UOS= 84%

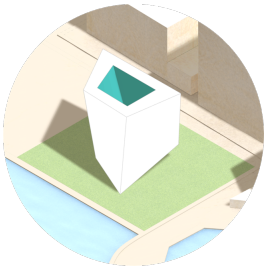
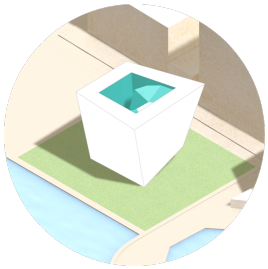
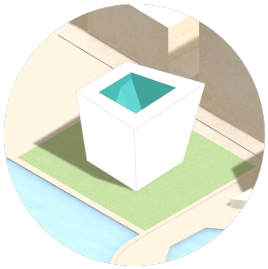
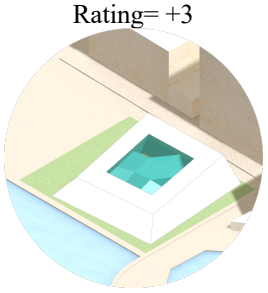
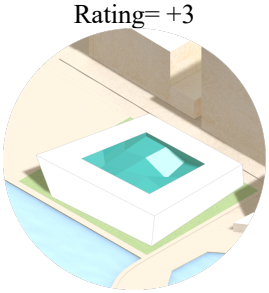
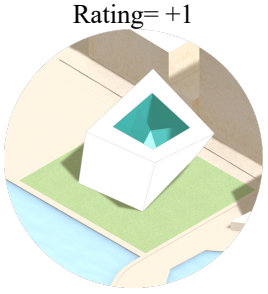
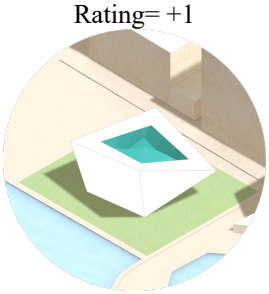
Participant	Experiment	Design Options		
				

Table A.2.20. Non-Expert 2U2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2U2	Experiment 1 Intervention 1	Option 16 AV= 17% NPZ= 25% UOS= 30% Rating= +3 	Option 0 AV= 4% NPZ= 56% UOS= 31% Rating= +3 	Option 6 AV= 24% NPZ= 24% UOS= 40% Rating= +3 
	Experiment 1 Intervention 2	Option 16 AV= 17% NPZ= 7% UOS= 81% Rating= +1 	Option 0 AV= 0 NPZ= 46% UOS= 79% Rating= +1 	Option 6 AV= 42% NPZ= 38% UOS= 89% Rating= +2 
	Experiment 2 Intervention 1	Option 12 AV= 1% NPZ= 5% UOS= 87%	Option 13 AV= 14% NPZ= 7% UOS= 88%	Option 15 AV= 10% NPZ= 8% UOS= 89%

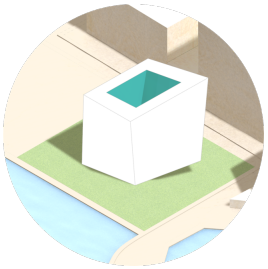
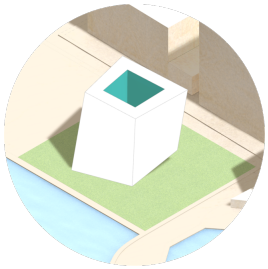
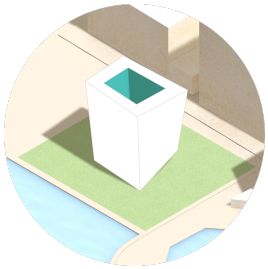
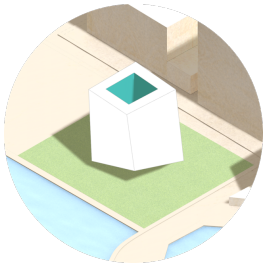
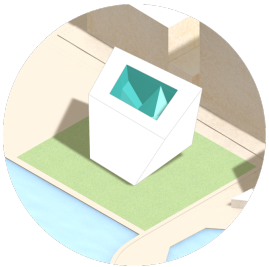
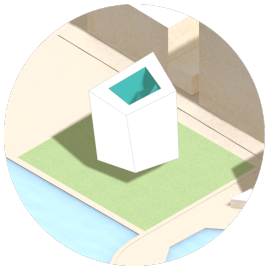
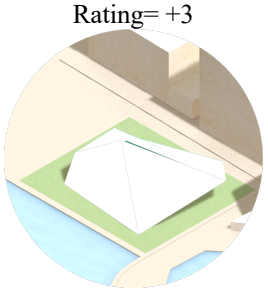
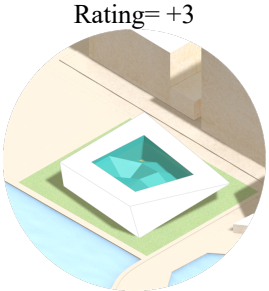
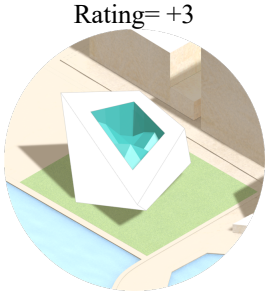
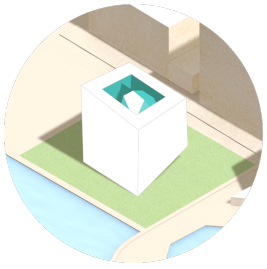
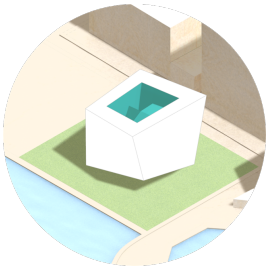
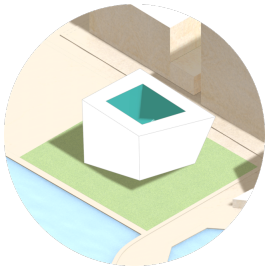
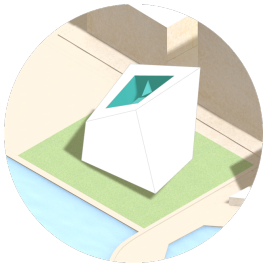
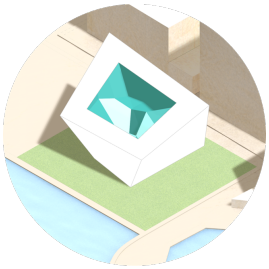
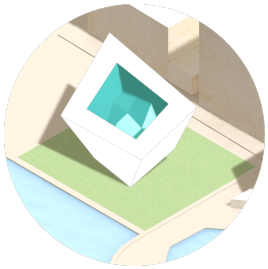
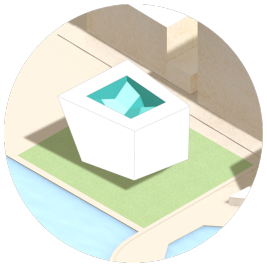
Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 2	Option 0 AV= 24% NPZ= 12% UOS= 89% 	Option 13 AV= 9% NPZ= 7% UOS= 88% 	Option 15 AV= 23% NPZ= 16% UOS= 91% 

Table A.2.21. Non-Expert 2V2 Experiment 1 and 2 Results.

Participant	Experiment	Design Options		
2V2	Experiment 1 Intervention 1	Option 6 AV= 3% NPZ= 19% UOS= 66% Rating= +3 	Option 8 AV= 28% NPZ= 18% UOS= 55% Rating= +3 	Option 18 AV= 24% NPZ= 31% UOS= 87% Rating= +3 
	Experiment 1 Intervention 2	Option 8 AV= 1% NPZ= 1% UOS= 84% Rating= +2	Option 6 AV= 2% NPZ= 15% UOS= 85% Rating= +2	Option 15 AV= 0 NPZ= 25% UOS= 84% Rating= +2

Participant	Experiment	Design Options		
				
	Experiment 2 Intervention 1	Option 3 AV= 0 NPZ= 4% UOS= 80% 	Option 5 AV= 5% NPZ= 3% UOS= 81% 	Option 17 AV= 12% NPZ= 6% UOS= 90% 
	Experiment 2 Intervention 2	Option 13 AV= 2% NPZ= 20% UOS= 84% 	Option 14 AV= 3% NPZ= 7% UOS= 86% 	Option 15 AV= 7% NPZ= 2% UOS= 84% 