

Fly Away: Evaluating the Impact of Motion Fidelity on Optimized User Interface Design via Bayesian Optimization in Automated Urban Air Mobility Simulations

Luca-Maxim Meinhardt
luca.meinhardt@uni-ulm.de
Institute of Media Informatics, Ulm
University
Ulm, Germany

Clara Schramm
clara.schramm@uni-ulm.de
Institute of Media Informatics, Ulm
University
Ulm, Germany

Pascal Jansen
pascal.jansen@uni-ulm.de
Institute of Media Informatics, Ulm
University
Ulm, Germany

Mark Colley
m.colley@ucl.ac.uk
Institute of Media Informatics
Ulm, Germany
UCL Interaction Centre
London, United Kingdom

Enrico Rukzio
enrico.rukzio@uni-ulm.de
Institute of Media Informatics, Ulm
University
Ulm, Germany

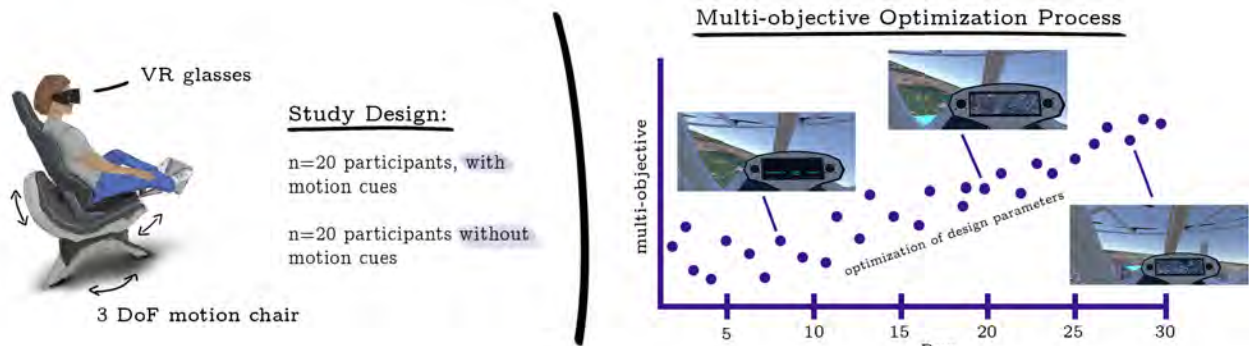


Figure 1: Overview of the user study setup and optimization process. (Left) N=40 participants were divided into two groups: n=20 experienced a simulated air taxi flight with motion cues using a 3-Degrees-of-Freedom motion chair and VR, while the other n=20 experienced the simulation in VR without motion cues. (Right) We applied Multi-Objective Bayesian Optimization to iteratively refine the air taxi’s user interface design over 30 optimization runs, focusing on six key objectives: trust, perceived safety, mental demand, understanding, acceptance, and aesthetics.

Abstract

Automated Urban Air Mobility (UAM) can improve passenger transportation and reduce congestion, but its success depends on passenger trust. While initial research addresses passengers’ information needs, questions remain about how to simulate air taxi flights and how these simulations impact users and interface requirements. We conducted a between-subjects study (N=40), examining the influence of motion fidelity in Virtual-Reality-simulated air taxi flights on user effects and interface design. Our study compared simulations with and without motion cues using a 3-Degrees-of-Freedom motion chair. Optimizing the interface design across six objectives,

such as trust and mental demand, we used multi-objective Bayesian optimization to determine the most effective design trade-offs. Our results indicate that motion fidelity decreases users’ trust, understanding, and acceptance, highlighting the need to consider motion fidelity in future UAM studies to approach realism. However, minimal evidence was found for differences or equality in the optimized interface designs, suggesting personalized interface designs.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; *Empirical studies in visualization.*



This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
<https://doi.org/10.1145/3706598.3713288>

Keywords

urban air mobility, virtual reality, motion chair, Bayesian optimization, mobility

ACM Reference Format:

Luca-Maxim Meinhardt, Clara Schramm, Pascal Jansen, Mark Colley, and Enrico Rukzio. 2025. Fly Away: Evaluating the Impact of Motion Fidelity on Optimized User Interface Design via Bayesian Optimization in Automated Urban Air Mobility Simulations. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 1, 2025, Yokohama, Japan. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3706598.3713288>

1 Introduction

Automated Urban Air Mobility (UAM) is emerging as a potential solution to alleviate current congestion issues. For instance, in cities like London, Paris, and Brussels, people experience over 130 hours of congestion annually [63]. Even with the rise of hybrid working [62], recent statistics show that traffic congestion continues to worsen [64], highlighting the need for innovative transportation alternatives like UAM. Advancements in battery technology have enabled the development of electrically powered air taxis, which produce zero local emissions. For instance, Volocopter's electric air taxi, VoloCity, has a range of 35 km, sufficient for transporting passengers between city centers and airports in most of the world's megacities [77]. These vehicles, known as electrical vertical take-off and landing (eVTOL) aircraft, are particularly beneficial due to their minimal urban space requirements. Due to these advances, the European Union Aviation Safety Agency (EASA) predicts that this new mode of transportation could become available within this decade [1]. In its early stages, UAM is expected to focus on small-scale operations, with air taxis carrying only a few passengers per trip [24, 77]. However, future visions extend beyond individual transportation, aiming to implement large-scale public transportation systems similar to buses [61]. Yet, in both visions, inexperienced, non-pilot passengers are expected to utilize these air taxis. Nonetheless, studies indicate that trust and perceived safety are critical factors influencing passengers' willingness to adopt to UAM [1, 3, 23]. To enhance passenger trust, Al Haddad et al. [3] suggested implementing new safety standards, such as in-cabin surveillance cameras. Other solutions can be found in the field of Human-Computer Interaction, as Colley and Meinhardt et al. [17] demonstrated that simple trajectory visualizations on the windshield display (WSD) of air taxis significantly increase passengers' trust and perceived safety compared to no visualization or other visualizations such as a tunnel visualization or augmented landmarks. However, their study was only conducted as a low/mid-fidelity virtual reality (VR) study, although Edwards and Price [23] recommended constructing a high-fidelity simulator to study passenger needs and the impact of rotor noise and vibration in the cabin. This raises the methodological question of how the fidelity of a simulator impacts the user studies' results when investigating passengers' information needs in the field of UAM.

While most studies in the automotive domain have been conducted in VR [14, 27], recent research is already using higher-fidelity simulations like motion chairs [54] or real vehicles [28, 39]. Comparing these different types of simulations, Yeo et al. [82] found significant differences in perceived motion fidelity and presence between VR-only setups, motion platforms, and real vehicles. Therefore, we suggest that motion fidelity may also affect users' information needs. However, to our knowledge, no studies have investigated the

effects of motion fidelity on passengers in simulated UAM scenarios. This research is important, though, as the simulation's motion fidelity may influence the results' reliability, especially when addressing the complex issues of passengers' perceived safety and trust that arise for different visualizations of automated air taxis, as indicated by Colley and Meinhardt et al. [17]. While not focusing on motion fidelity, they only investigated a single aspect of visualization design for UAM—namely trajectories. However, a complete user interface (UI) involves multiple interconnected components. Optimizing the combination of these components becomes increasingly complex when trying to enhance multiple aspects, such as perceived safety and trust, which is especially critical in UAM contexts, where passengers often feel less familiar and more vulnerable compared to traditional ground transportation. Yet, in the field of Automated Vehicles (AVs), Normark [60] allowed passengers to manually customize the size, location, and color of icons on the dashboard, center stack, and Head-Up Display (HUD). By tailoring these UIs to individual preferences, the overall user experience can be enhanced [73], fostering increased perceived safety and trust in future automated UAM.

Instead of relying on manual adjustments to various design elements within the visualizations, Multi-Objective Bayesian Optimization (MOBO) offers a more efficient solution. MOBO iteratively identifies the optimal design parameters by incorporating user feedback at each iteration, progressively refining the design to meet the users' needs. This approach has already been successfully applied in various domains to solve design optimization problems [10, 11, 37, 44, 47]. MOBO predicts which design changes will most effectively achieve the desired objectives, such as increasing passengers' trust and perceived safety. It manages multiple objectives by identifying the best trade-offs among them, known as the Pareto front, ensuring the most effective compromises for UI design [50]. In light of these considerations, this work is guided by the following two research questions (RQs):

RQ1 What are the characteristics of an optimized UI design for automated air taxis that enhance passengers' trust, perceived safety, mental demand, understanding, acceptance, and aesthetics?

RQ2 How does the motion fidelity of the simulation affect the design parameters and user's effects for a UI in automated UAM?

To address these RQs, we conducted a between-subject user study with $N=40$ participants. The participants were divided into two groups, with $n=20$ participants, each experiencing a simulation of an automated air taxi flight created in Unity. The first group experienced the simulation using a 3-degrees-of-freedom (DoF) motion chair combined with VR, while the second group experienced the simulation using VR only, without the motion chair. Additionally, we applied MOBO to the air taxi's UI design. This optimization focused on 12 design parameters, such as the length and transparency of the ego-trajectory on the WSD, a map visualization on the internal display, and the visualization of boundary boxes around other air taxis. We optimized these parameters based on six objectives: trust in automation, understanding, mental demand, perceived safety, acceptance, and aesthetics.

Our findings show that while motion fidelity had a minimal effect on most UI design parameters, we found strong to extreme evidence that it reduces users' trust, understanding, and acceptance, highlighting its importance in future UAM studies. Nonetheless, the lack of differences in immersion between the groups may derive from participants' limited experience with real air taxi flights. Additionally, the relatively high variance in design preferences on the Pareto front suggests that a one-size-fits-all approach may not be suitable for air taxi interfaces. Although objectives improved during optimization, personalized interfaces tailored to individual preferences could enhance user experiences. Eventually, we provide practical guidelines for future UAM research on automated air taxi interfaces.

Contribution Statement [80]

- **Artifact or System** We developed a VR, Unity-based simulation of an automated air taxi flight, designed to iteratively optimize the UI through MOBO, allowing the identification of optimal design parameters based on user feedback across six objectives.
- **Empirical study that tells us about how people use a system.** We conducted a between-subjects study (N=40) to investigate the impact of motion fidelity on user experience and UI design for automated air taxis. Our findings show that there is evidence that motion fidelity decreases trust, understanding, and acceptance while having minimal effect on optimized UI design, suggesting the need for tailored user interfaces in future UAM applications.

2 Related Work

This section will provide an overview of prior research on users' information needs within the context of UAM and explore potential UIs designed to visualize this information. Furthermore, it introduces Bayesian Optimization (BO) as an algorithm for optimizing these UIs.

2.1 Human-Computer Interaction for Urban Air Mobility

Approach UAM with an HCI perspective, Kim et al. [41], conducted a workshop focused on User Experience in UAM with HCI experts. The insights from this workshop were further analyzed by Lim et al. [48], revealing a transition in focus from initial concerns of *safety* and *acceptance* in the first phase of UAM to later emphasizing *comfort* as potential passengers' major concerns. However, they did not propose explicit solutions for these concerns. Edwards and Price [23] conducted a similar workshop with aviation professionals, identifying six primary categories of passenger concerns: safety, noise and vibration, passenger well-being, and environmental concerns. To address these concerns, they suggested creating a high-fidelity simulator to understand passenger needs better and study the effects of rotor noise and vibration inside the cabin. Further advancing the field, Meinhardt and Colley et al. [52] held a workshop with six professional helicopter pilots to evaluate automation and visualization possibilities for future UAM piloting. Their focus included visualizing air traffic, avoiding obstacles, and map visualization to enhance situational awareness. Their findings align

with Janetzko and Kacem [35], who investigated the information needs of emergency flight pilots, noting a desire to prioritize critical information, such as collision avoidance, during these flights. Diving into the visualization of these information needs, initial research was conducted by Colley and Meinhardt et al. [17]. In a VR user study, they found that a chevron path line visualization was the most effective visualization for significantly increasing trust and perceived safety compared to other visualizations. Particularly in the presence of other air traffic, the path line increased trust and provided a more predictable view of the air taxi's future trajectory. Based on their work, Valente et al. [75] investigated the path line visualization during different flight phases (take-off, cruise, landing) and visibility conditions (daylight, night, foggy). They found that their participants' understanding decreased during take-off and landing and under poor visibility conditions when using visualizations other than the path line, suggesting that this visualization performs best in most tested conditions.

Drawing from the insights of visualization of passengers' information needs [17, 53, 75], we see that *how* flight-relevant information is displayed significantly impacts passengers' trust and perceived safety. However, these studies have primarily focused on a single visualization component—namely trajectories, whereas a complete UI involves multiple interconnected components. Optimizing the combination of these components, however, becomes increasingly complex. To address this challenge, this work employs MOBO to explore and optimize multiple parameters in the design of air taxi UIs. Therefore, the next section introduces MOBO for interface designs to relate to **RQ1** of this research.

2.2 Multi-Objective Bayesian Optimization for User Interface Designs

Designing an effective UI for automated UAM involves selecting the best combination of design parameters—such as position, transparency, and size—to achieve key objectives like enhancing trust and perceived safety. Given the complexity of modern UIs, where numerous design parameters must be optimized, testing every possible combination is infeasible.

BO offers an efficient solution by modeling the relationship between design parameters and objective functions. BO iteratively refines designs by balancing seeding (trying new designs) and exploitation (focusing on promising ones), which makes it particularly effective in optimizing UI designs with limited data and unknown outcomes [8]. For subjective design objectives, such as trust and perceived safety, it is necessary to integrate the user into the BO process. In such a Human-in-the-Loop (HITL) process (see Chan et al. [10]), the BO adjusts the design parameters in each optimization iteration based on the user's subjective evaluation of the last design. HITL BO has been applied successfully in various UI design challenges, such as defining animation parameters based on user feedback [7], optimizing font settings for faster reading [37], and improving UI interactions to reduce task completion times [21]. For UI design for automated air taxis, multiple objectives like trust, perceived safety, and mental demand must be balanced [17, 53, 75]. Hence, MOBO (e.g., see Johns et al. [36]) can be used to find the optimal design parameters that balance all objectives. In particular, MOBO generates a range of optimal designs, known as the Pareto

front, where no design can improve one objective without compromising another [50]. This approach has been applied in contexts like touchscreen keyboards [22] and haptic interfaces [31].

In the context of UAM, the UI design challenge is further complicated by the need to account for the impact of motion fidelity on user perceptions. Unlike automated driving, where motion cues occur primarily in two dimensions, UAM involves three-dimensional movement, including vertical maneuvers like takeoffs and landings. Research has shown that passengers' understanding is lower during vertical flights compared to cruising [75]. However, this research was based on video simulations that did not incorporate actual motion, raising questions about how motion fidelity might affect the study outcomes in the UAM domain. To address these concerns, we explored how different levels of motion fidelity (motion cue and no motion cues) influence UI design in UAM, as outlined in **RQ2**. The following section will give an overview of motion fidelity simulators in the field of AVs and UAM.

2.3 Motion Fidelity Simulators

Simulators used in AV studies vary widely in terms of motion fidelity, which refers to how accurately they replicate the physical movements of a vehicle. This fidelity can range from fixed-base setups with minimal or no motion cues to advanced moving-base simulators with multiple degrees of freedom (DoF) [9, 32, 57].

Low-cost, fixed-base simulators employ simple rotational or linear motion to simulate basic vehicle dynamics [6, 20, 33]. In contrast, high-end moving-base simulators offer more sophisticated motion representation, such as 6-DoF platforms that can simulate complex maneuvers including pitch, roll, and yaw [12, 16]. Recent works focused on achieving effective motion fidelity with more accessible technologies in terms of costs. For example, the *SwiVR-Car-Seat* developed by Colley et al. [16] uses a 1-DoF rotational seat to simulate longitudinal and lateral vehicle dynamics. This approach is based on the finding by Rietzler et al. [68] that angular impulses alone can effectively represent vehicle acceleration or deceleration. Building on this concept, Hock and Colley et al. [32] simulated AV motion using a remote-controlled wheelchair, leveraging its acceleration and rotation capabilities. For comparison, Yeo et al. [82] developed six simulators with varying levels of motion fidelity, ranging from monitor setups without motion cues to mixed reality environments combined with a real vehicle. Their study revealed that this actual vehicle, combined with a mixed reality headset, had the significantly highest perceived motion fidelity and perceived presence. However, this research primarily focused on the fidelity aspect without exploring how different levels of motion fidelity might influence the effects of UIs on users in these environments.

In the context of UAM, motion simulators for air taxi flights have also been developed, utilizing motion chairs [5] or CAVE virtual environments [49]. However, these studies primarily concentrate on constructing the simulators rather than examining how different levels of motion fidelity might affect user preferences for UI design during flights. To address this gap and explore the optimization of UIs in automated UAM, we conducted a user study detailed in the following section.

3 User Study

To evaluate the impact of different UI optimizations in VR environments, both with and without motion cues, we conducted a between-subject design study with $N=40$ participants. The participants were divided into two groups: one group experienced the simulation without motion cues ($n=20$), while the other group used a 3-DoF motion chair ($n=20$).

3.1 Apparatus

We developed a simulation using Unity [74] version 2022.3.19f1 to present the ego perspective of a passenger inside the Volocopter 2X flying over New York City. We utilized the Google Map Tiles API for Unity to render the environment. In the simulation, the Volocopter followed a predetermined route that included various maneuvers, such as turns in both directions and altitude changes. We incorporated other air taxis into our simulations, as their presence influences passengers' trust and perceived safety [17, 53]. Predictions for New York suggest that approximately 500 air taxis might operate at any given time [29, 65–67]. Consequently, our simulation featured 500 air taxis operating on randomly timed predefined trajectories. Forty participants experienced the simulation, split into two groups. One group of $n=20$ participants used the Vive Pro headset without a motion chair, while the other group of $n=20$ participants used the Vive Pro together with the YAW VR motion chair [70], which features 3 DoF. The chair's motion was synchronized with the movement of the virtual air taxi, providing a 1:1 translation of the air taxi's rotation (yaw, pitch, and roll). In line with the work of Colley and Meinhardt et al. [17], the air taxi was equipped with a WSD capable of visualizing the trajectory of both the ego air taxi and the other air taxis. Additionally, the WSD highlighted other air taxis using boundary boxes, a technique inspired by prior research in the automotive domain where recognized vehicles are similarly highlighted [15, 18, 81]. The air taxi also featured a display as the dashboard. It showed a map of the environment beneath the ego air taxi, including its trajectory and additional information such as current speed and altitude (see Figure 4).

3.2 Multi-Objective Bayesian Optimization

The Unity simulation was connected to a Bayesian multi-objective optimizer from the Python package BoTorch [4] (version 0.11.1). This optimizer was used to iteratively adjust 12 design parameters of the UI, such as the length of the ego trajectory and its transparency (the following section will explain the design parameters in detail). We employed the qEHVI acquisition function, representing the expected hypervolume increase. To ensure that after each run, only a single batch is selected for evaluation, we set $q=1$, in line with Chan et al. [10].

The optimization process began with a five-run sampling phase using Sobol sampling [72]. This method systematically divides the design space into evenly distributed regions and selects representative design parameter configurations for each. By collecting user ratings on these samples, the optimizer explores how the individual user perceives broader regions of the design space. This identifies the initial promising regions for further exploration. All users received the same five samples during the sampling phase to prevent bias arising from different starting points, which represents the

default BoTorch behavior. We deemed five samples appropriate based on internal tests and prior studies on air taxi UIs, which identified significant user preference differences with a few distinct designs [52, 75]. This suggests that Pareto-optimal designs can likely be discovered in a comparable number of designs' origin regions, reducing the need for extensive seeding in the first phase. In the following 25-run optimization phase, the acquisition function dynamically balanced "exploitation" (refining known *good* configurations) and "seeding" (searching for new regions of the design space). In line with previous work using MOBO [10, 69], the acquisition function governs this balance, as users unfamiliar with an optimization process might not know which balance to choose manually (This is in contrast to [83], which explores user-driven balancing). We used 1024 restart candidates and 512 Monte Carlo samples to approximate the acquisition function during the optimization process. The design parameters of the UI inside the air taxi were updated throughout this process. By having more optimization runs compared to sampling runs, we focused on fine-tuning the design parameters based on individual user feedback. This aimed to converge more efficiently to personalized designs.

3.3 Design Parameters

The 12 design parameters to be optimized were derived from related work in the field of UAM [17, 59, 75]. All parameters were mapped to a range between 0 and 1. While most parameters were on a continuous scale, three were binary, representing whether a UI element was visible. To handle these binary parameters, we set a threshold: elements with a value of less than 0.5 were considered invisible, while those with a value of 0.5 or more were considered visible. This approach was chosen because MOBO is typically more efficient when working with continuous parameters [69]. The color was deliberately excluded as a design parameter, as turquoise for all augmented objects in the WSD follows an established design principle [78], ensuring a neutral appearance. Further, the visualization range in the WSD is capped at 1 km; beyond this distance, path visualizations and boundary boxes are not displayed, as they would no longer be perceivable. In Figure 2, all parameters to be optimized are depicted, including their range between 0 to 1.

Design Parameters:

- **Ego Trajectory Length:**
The length of the ego trajectory path in the WSD, scaled between 0m and 1km (maximum clipping distance).
- **Ego Trajectory Alpha:**
The transparency of the ego trajectory path in the WSD.
- **Ego Chevron Size:**
The size of the chevrons along the ego trajectory path, scaled between 0m and 12.5m.
- **Ego Chevron Distance:**
The distance between individual chevrons, scaled between 0m and 42m.
- **Ego Path Length in Map:**
The length of the ego path displayed on the map represents a range from 0m to 260m (260m being fully visible on the entire display).
- **Other Trajectory Length:**
The length of the trajectory paths for other air taxis, scaled between 0m and 205m, corresponding to the distance between these air taxis in the simulation.
- **Other Trajectory Alpha:**
The transparency of the trajectory paths for other air taxis in the WSD.
- **Other Chevron Size:**
The size of chevrons along the trajectory paths of other air taxis, scaled between 0m and 12.5m.
- **Other Chevron Distance:**
The distance between individual chevrons on the trajectory paths of other air taxis, scaled between 0m and 42m.
- **Map Display:**
A binary parameter indicating whether a map is shown at the display (< 0.5 : no map, ≥ 0.5 map is displayed).
- **Boundary Box:**
A binary parameter indicating whether the boundary boxes are displayed on the WSD (< 0.5 no boundary box, ≥ 0.5 : boundary boxes displayed)
- **Additional Information on Display:**
A binary parameter indicating whether additional information (altitude and speed) are shown at the display (< 0.5 no boundary box, ≥ 0.5 : boundary boxes displayed)

3.4 Objective Function

An objective function f maps a visualization design x to a subjective metric that the optimizer seeks to maximize or minimize. In our study, we focused on six subjective metrics informed by previous research [17, 19]: *perceived safety*, *trust in automation*, *understanding*, *mental demand*, *acceptance*, and *aesthetics*. Among these, *mental demand* was the only metric we aimed to minimize, while the others were to be maximized. To evaluate these metrics throughout the optimization process, we employed established questionnaires, as presented in the following:

Mental demand was assessed using the mental workload subscale of the raw NASA-TLX [30] on a 20-point Likert scale. **Understanding** was measured via the *Predictability/Understandability* subscale from the *Trust in Automation* questionnaire by Körber [42]. **Trust** was evaluated using the *Trust* subscale from the same questionnaire, both using a 5-point Likert scale. **Perceived safety** was rated by participants using four 7-point semantic differentials ranging from -3 (anxious/agitated/unsafe/timid) to +3 (relaxed/calm/safe/confident), where a higher score indicates greater perceived safety [25]. Additionally, we included three single items to assess specific aspects of the visualization design. **Acceptance** was assessed using two items inspired by the van der Laan acceptance scale [76]: "I find the visualizations of the automated vehicle **useful**" and "I find the visualizations of the automated vehicle **satisfying**." These were averaged into a single acceptance objective. **Aesthetics** was adapted from Colley et al. [17] and measured visual appeal with the statement, "I found the visualizations visually appealing," rated on a 7-point Likert scale.

To ensure consistency across the questionnaires, which employed different Likert scales, we normalized all six metrics to

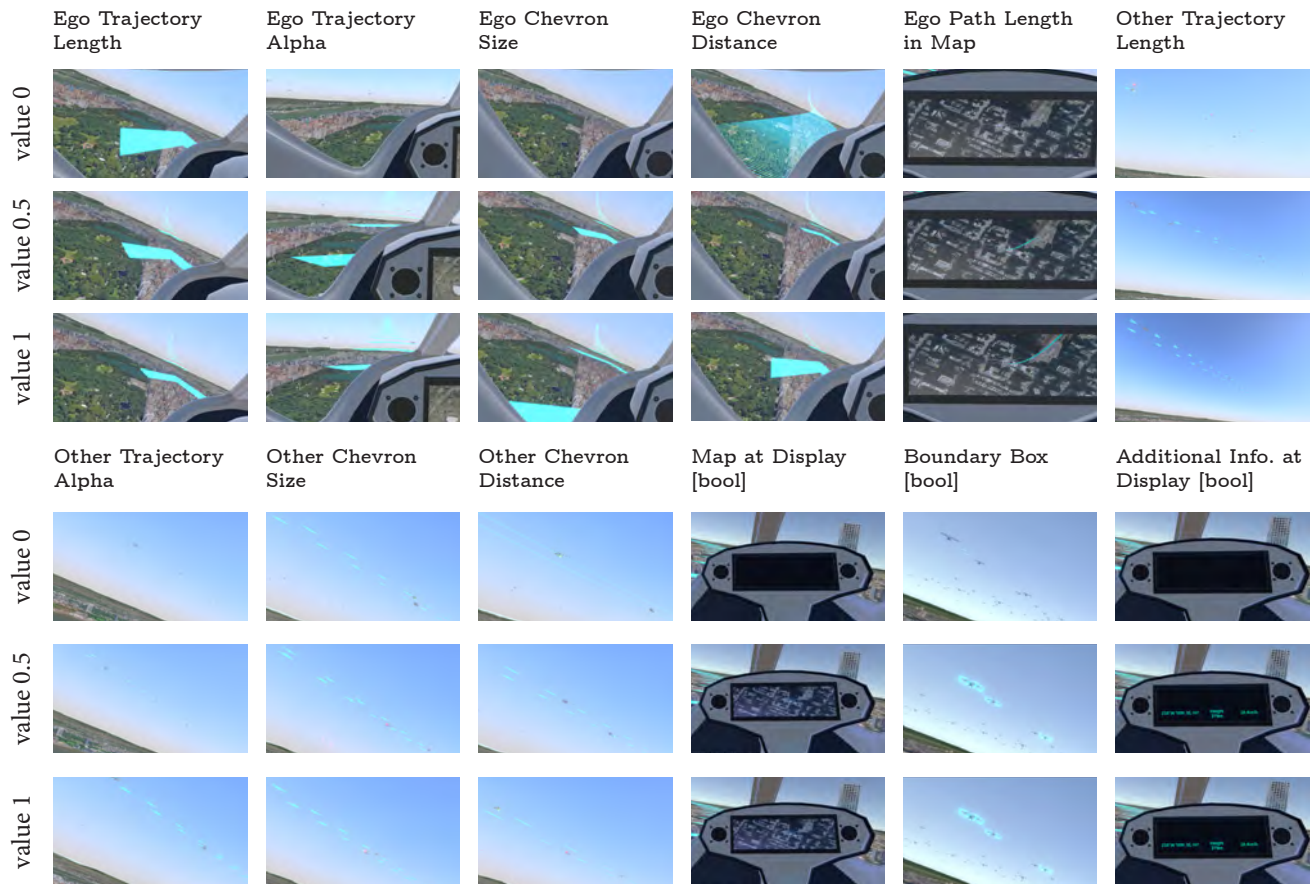


Figure 2: Design parameters to be optimized. The figure shows the values of 0, 0.5 and 1 for each design parameter

a range of $[-1, 1]$. Notably, for the mental demand metric, we inverted its orientation such that higher normalized values indicate lower perceived mental demand.

3.5 Procedure

The procedure was similar for both conditions (VR with and without a motion chair). Each session began with a brief introduction, followed by participants agreeing to a consent form. Participants then experienced 30 seconds of a simulated air taxi flight. Afterward, they rated their experience using the subjective metrics detailed in subsection 3.4. The Bayesian Optimizer used these ratings to update the design parameters and present the next iteration of the UI design. Each participant went through 30 optimization runs in total. However, if a participant gave the highest possible ratings for all objectives (or the lowest for mental demand) for a specific UI design, the study ended early, as this was considered the optimal design for that participant. After all runs, a demographic questionnaire was administered, asking questions about the participant’s age, gender, and immersion using the subscale of the Technology Usage Inventory (TUI) [43]. For their 1.5h effort, the participants were compensated with 15Euros.

3.6 Participants

As we conducted a between-subject study, the participants’ demographics are reported separately for each group.

The group with motion cues had an average age of 25.60 years ($SD=3.26$). This group consisted of 10 males, 10 females, and no non-binary participants. Among them, four participants were employed, one was self-employed, and the remaining were college students. The group without motion cues had an average age of 24.95 years ($SD=2.99$). This group included 11 males, 9 females, and no non-binary participants. All participants in this group were college students.

3.7 Results

3.7.1 Analysis. The goal of MOBO is to identify the Pareto front, which consists of all Pareto optimal points in the design space. Each point on this front (e.g., a combination of design parameters in a UI) represents a design that cannot be improved in one objective without compromising another. This set of optimal designs captures the most efficient trade-offs between conflicting design objectives [50]. Hence, we determined the Pareto optimal values using the R package *EMOA: Evolutionary Multiobjective Optimization Algorithms* [55] for each participant. For the subsequent analysis,

we exclusively used these Pareto values, filtering the combinations of design parameters and questionnaire ratings to include only those located on the Pareto front. This approach ensures we only considered the most efficient and balanced designs in our subsequent analysis. While the group experiencing motion during the study yielded $n=48$ Pareto designs, the group only experiencing the study in VR without motion yielded $n=42$ Pareto designs.

In Appendix A, we illustrate the optimization process across all participants for each objective. The results clearly show that most individual objectives improved over the course of the runs. Additionally, the aggregated and normalized scores for the combined objectives steadily increased throughout the optimization process, indicating that the iterative approach effectively enhanced the overall design outcomes. Figure 3 shows the optimization process for the combined objectives. This was done by normalizing the ratings across all six objectives to a range of $[-1, 1]$. For the mental demand, the ratings were inverted to align with the overall optimization direction (see subsection 3.4).

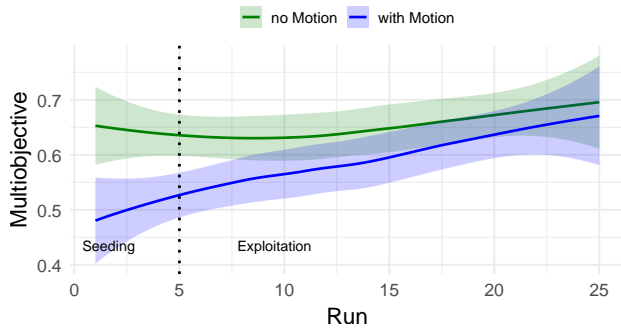


Figure 3: Optimization Process of the averaged normalized objectives during the 30 runs

To evaluate the differences between the two groups, we employed an independent Bayesian t-test. Unlike traditional null hypothesis significance testing (NHST), Bayesian analysis offers a clear advantage: it directly quantifies the evidence supporting one hypothesis over another. As noted by Williams et al. [79], “a Bayes factor analysis makes it clear when a set of observed data is more consistent with the null hypothesis than the alternative” [79, p. 1]. This approach is particularly advantageous as Bayesian methods are inherently designed to quantify evidence directly. In contrast, NHSTs are more centered on decision-making processes, such as rejecting or not rejecting a null hypothesis, rather than providing a directional measure of evidence [58]. In contrast, the Bayes factor directly indicates the strength of the evidence.

According to Lee and Wagenmakers [46], a Bayes factor less than 1 provides evidence in favor of the null hypothesis, suggesting equality between the groups. Conversely, a Bayes factor greater than 1 suggests evidence for a difference between the groups. The further the Bayes factor is from 1, the stronger the evidence supporting the respective hypothesis. If the Bayes factor is close to 1, the data do not strongly support either hypothesis.

3.7.2 Immersion. The descriptive data revealed that immersion with motion cues was rated higher ($M=21.3$, $SD=4.94$) than the simulation with only a VR setup ($M=18.5$, $SD=5.76$), with possible scores ranging from 4 to 28. A Bayesian independent t-test was conducted to examine the effect of the motion chair on immersion. However, the analysis yielded a Bayes factor of $BF = 0.998 \pm 0.01\%$, suggesting that the data provide no evidence for equality or difference between both groups. This indicates that there is no evidence regarding the effect of motion fidelity on immersion.

3.7.3 Design Parameters. The results of our analysis for each design parameter are detailed in Table 1 and plotted in Figure 5, which includes the IQRs for better representation of variability. While methods exist for visualizing multi-dimensional Pareto fronts (e.g., [2, 13]), our study involves a unique Pareto front for each participant, making it challenging to combine these individual optimal designs into a single, unified UI. Hence, we present two representative Pareto designs—one from Participant 1 (motion condition) and another from Participant 40 (no-motion condition)—in Figure 4. These examples were selected because most of their design parameter values mostly fall within the IQR of their respective groups (Participant 1: 11 out of 12 parameters; Participant 40: 9 out of 12 parameters are inside the IQR). These examples provide an illustrative comparison of one optimized design for each condition.

The parameter related to the chevron size of the other air taxis shows strong evidence for a difference between the two groups. This suggests that the optimized chevron size on the Pareto front for the “no Motion” group ($M=0.75$, $SD=0.15$) is larger than for the “with Motion” group ($M=0.59$, $SD=0.30$). There was also extreme evidence of a difference in the presence of a boundary box around the other air taxis ($M=0.45$, $SD=0.19$ for the “no Motion” group and $M=0.71$, $SD=0.25$ for the “with motion” group). This parameter is binary. Hence, as the value for “no Motion” is below the 0.5 threshold, this indicates that the optimal design for this group would not include the boundary box, while the optimal design for the “with Motion” group would include it, exceeding the threshold.

3.7.4 Questionnaire Ratings. We analyzed the mean ratings from the questionnaires of all participants whose design parameters were on the Pareto front. We then compared these ratings between the group with motion cues and the group without motion cues. Refer to Figure 6 for a comparison overview.

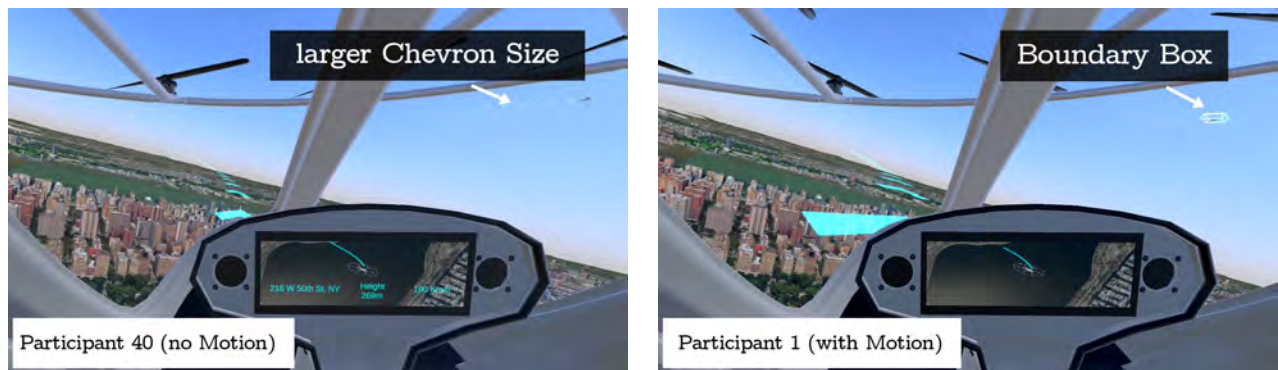
Trust in Automation. The Bayesian analysis of the trust measure resulted in a BF of $3970.14 \pm 0\%$, suggesting extreme evidence for a difference between the group without the motion cues ($M=4.14$, $SD=0.73$) and with motion cues ($M=3.21$, $SD=1.02$, see Figure 6a).

Understanding. For the understanding measure, the analysis yielded a BF of $56.85 \pm 0\%$, providing very strong evidence for the differences between the two groups. For the optimized design parameters on the Pareto front, the motion setup yielded a lower understanding ($M=3.12$, $SD=1.11$) than the setup without the motion chair ($M=3.85$, $SD=0.73$, see Figure 6b).

Mental Demand. The mental demand measure showed a BF of $0.23 \pm 0.02\%$, indicating moderate evidence for equality between the groups. Hence, the design parameters in the Pareto front in the

Table 1: Results of Bayesian Analysis for Each Design Parameter including the Medians and IQRs

Design Parameter	BF (\pm %)	no Motion Mdn (IQR)	with Motion Mdn (IQR)	Evidence
Ego Trajectory Length	0.41 \pm 0.02%	0.75 (0.55, 1.00)	0.74 (0.49, 0.99)	anecdotal equality
Ego Trajectory Alpha	0.23 \pm 0.02%	0.31 (0.17, 0.45)	0.32 (0.20, 0.43)	moderate equality
Ego Chevron Size	0.26 \pm 0.02%	0.41 (0.16, 0.66)	0.41 (0.20, 0.63)	moderate equality
Ego Chevron Distance	0.22 \pm 0.02%	0.61 (0.51, 0.82)	0.66 (0.58, 0.74)	moderate equality
Ego Path Length in Map	0.36 \pm 0.02%	0.62 (0.25, 0.98)	0.61 (0.32, 0.91)	anecdotal difference
Other Trajectory Length	0.22 \pm 0.02%	0.31 (0.14, 0.47)	0.32 (0.17, 0.48)	moderate equality
Other Trajectory Alpha	0.62 \pm 0.02%	0.56 (0.34, 0.78)	0.49 (0.36, 0.52)	anecdotal equality
Other Chevron Size	16.54 \pm 0%	0.76 (0.65, 0.86)	0.57 (0.29, 0.84)	strong difference
Other Chevron Distance	0.22 \pm 0.02%	0.42 (0.25, 0.55)	0.43 (0.38, 0.50)	moderate equality
Map at Display [bool]	0.24 \pm 0.02%	0.66 (0.53, 0.78)	0.68 (0.55, 0.76)	moderate equality
Boundary Box [bool]	32473.69 \pm 0%	0.45 (0.35, 0.63)	0.71 (0.57, 0.89)	extreme difference
Addn. Info. at Display [bool]	0.51 \pm 0.02%	0.49 (0.39, 0.58)	0.47 (0.31, 0.62)	anecdotal equality



(a) Pareto front visualization of participant 40 without Motion

(b) Pareto front visualization of participant 1 with Motion

Figure 4: Exemplary optimized design parameters of Participant 40 (no Motion) and Participant 1 (with Motion). The concrete design parameters of these participants are plotted in Figure 5. Strong to extreme evidence of differences was found for the boundary box and chevron size of other air taxis. All other differences of design parameters, such the additional information at the display are due to personal preferences but no evidence was found to support this difference between groups

setup with motion yield similar ratings ($M=4.92$, $SD=2.01$) compared to the ones without motion ($M=5.07$, $SD=3.49$, see Figure 6c).

Perceived Safety. The analysis of perceived safety resulted in a BF of $3.96 \pm 0.01\%$, with moderate evidence in favor of a difference between the groups. The Pareto front design was rated lower for the group with motion ($M=1.15$, $SD=0.95$) than the one without motion ($M=1.64$, $SD=0.84$). See Figure 6d for more details.

Acceptance. The acceptance had a BF of $67.92 \pm 0\%$, suggesting very strong evidence in favor of a difference between the group without motion ($M=4.77$, $SD=1.53$) and with motion ($M=3.65$, $SD=1.36$, see Figure 6e).

Aesthetics. The Bayesian analysis yielded a BF of $29.63 \pm 0\%$ for the aesthetics measure, indicating moderate evidence favoring the group's difference. The optimal Pareto front design parameters yield a lower rating for the group with motion ($M=3.77$, $SD=1.43$) compared to the group without motion ($M=4.79$, $SD=1.39$, see Figure 6f).

3.8 Correlation between the Objectives

MOBO identifies optimal design parameters that balance multiple objectives along the Pareto front, ensuring that no single objective can be improved without compromising others [50]. To assess whether there are trade-offs between these objectives, we calculated the correlations among all objectives (see Figure 7). This analysis helps us understand how changes in one objective might influence others. The results show that all correlations were statistically significant. In particular, trust and understanding had a strong positive correlation ($r = 0.8$), as did acceptance and aesthetics ($r = 0.71$), meaning improvements in one were often associated with improvements in the other. In contrast, the correlations between Mental Demand and Aesthetics ($r = 0.14$) and Mental Demand and Acceptance ($r = 0.08$) were weak. The remaining objective pairs demonstrated moderate positive correlations.

4 Discussion

Automated UAM is expected to reduce urban congestion, with air taxis expected to operate in major cities within the next decade [1, 77]. However, successfully adopting UAM will depend on passenger

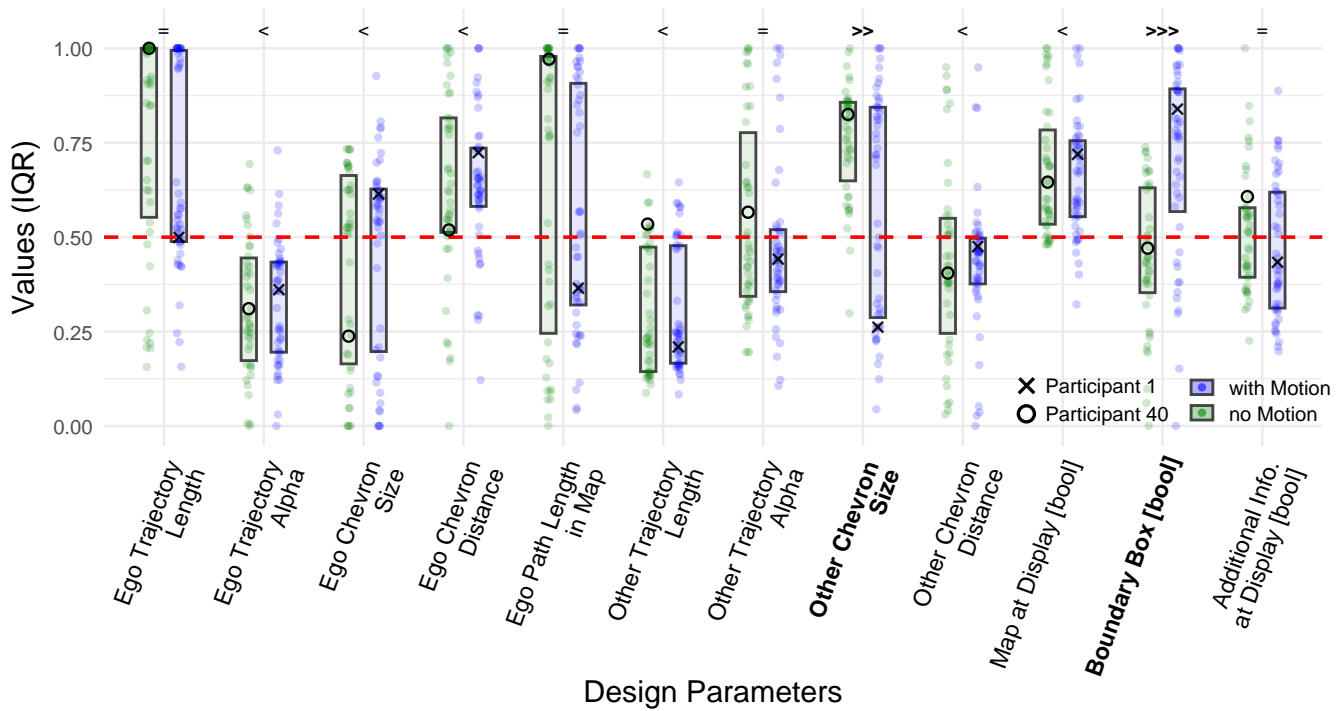


Figure 5: Comparison of the design parameters for both groups, using a Bayesian t-test. For both groups, the IQR is plotted. Further, the exemplary Pareto front designs of Participant 1 (with Motion) and Participant 40 (no Motion) are depicted (see Figure 4 for the visualization). The annotations are as follows defined by Lee and Wagenmakers [46]: "<<<<" for extreme evidence for equality ($BF < 0.01$), "<<" for strong or very strong evidence for equality ($BF < 0.1$), "<" for moderate evidence for equality ($BF < 0.3$), "=" for inconclusive or anecdotal evidence (BF between 0.3 and 3), ">" for moderate evidence for difference ($BF > 3$), ">>" for strong or very strong evidence for difference ($BF > 10$), and ">>>>" for extreme evidence for difference ($BF > 100$). The dotted line indicates the threshold for the boolean design parameters to either be smaller or greater than 0.5

trust and perceived safety in this novel mode of transport [3]. Research has shown that effective UI designs can significantly increase passengers' trust and perceived safety [17, 75]. However, to develop and research optimal UI designs for air taxis, it is crucial to use motion-fidelity simulators that accurately reflect the motion conditions of the UAM [23, 52]. While some studies have investigated the motion fidelity of simulators in the field of AVs [16, 32, 82], there remains limited understanding of the effects of motion fidelity on passengers and the UI in the field of UAM. Therefore, this research was driven by two RQs:

RQ1 What are the characteristics of an optimized UI design for automated air taxis that enhance passengers' trust, perceived safety, mental demand, understanding, acceptance, and aesthetics?

RQ2 How does the motion fidelity of the simulation affect the design parameters and user's effects for a UI in automated UAM?

To explore these RQs, we conducted a between-subject study with $N=40$ participants. One group experienced a simulated air taxi flight using a 3-DoF motion chair combined with VR ($n=20$), while the other group experienced the same simulation in VR without motion cues ($n=20$). For both groups, we applied MOBO to iteratively optimize the UI designs based on participants' ratings

of trust, perceived safety, understanding, mental demand, acceptance, and aesthetics [10]. This approach allowed us to compare user ratings and the resulting optimal UI designs for both motion and no-motion conditions on the Pareto front, where improving one objective inherently requires compromising another [50].

4.1 Motion Fidelity Decreases Trust, Understanding, and Acceptance

Our analysis found no evidence of a difference or equality in immersion between participants who experienced motion cues and those who did not. In contrast, Yeo et al. [82] found that motion platforms in VR significantly increased presence. While immersion refers to the technical quality of the virtual environment, presence describes the psychological feeling of being in a (virtual) place [71]. The lack of difference in immersion in our study may be attributed to the fact that none of the participants had previously experienced a real flight in an automated air taxi, making it difficult for them to rate the immersion of the simulations accurately. This raises the question of whether valid studies can be conducted in UAM contexts without participants having firsthand experience with air taxi movements. In contexts like AVs, most users already have some experience with driving or being passengers in vehicles, which

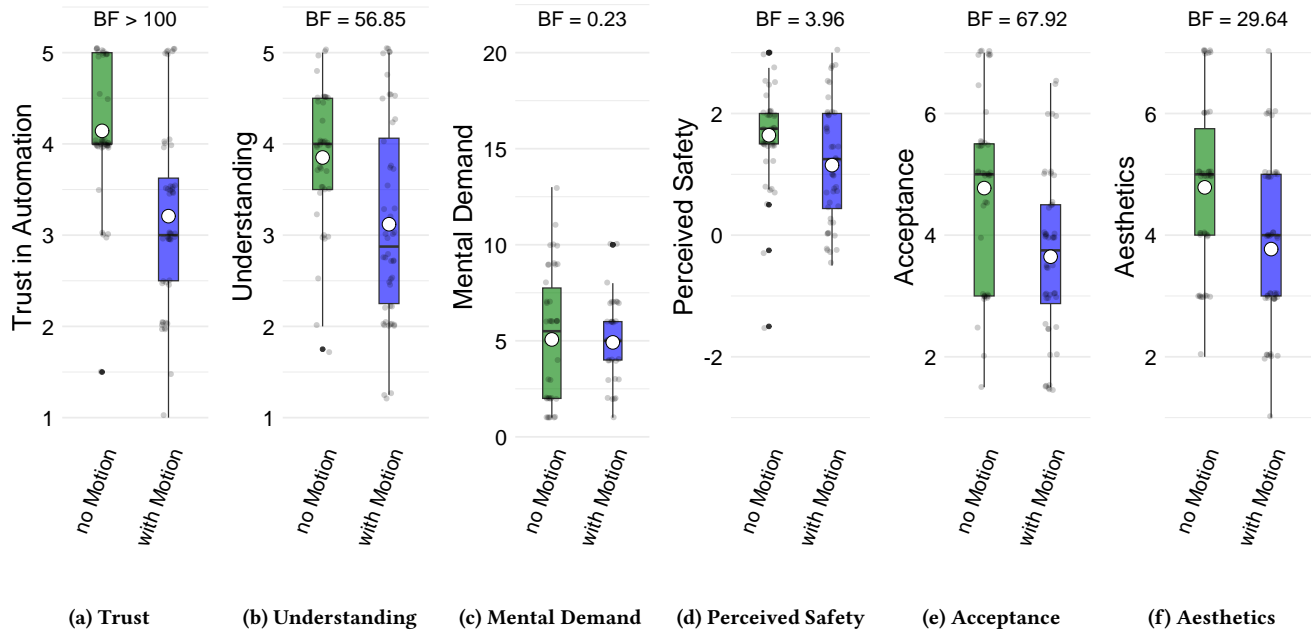


Figure 6: Rating for the subjective questionnaires comparing both groups: no Motion and with Motion of all Pareto optimal values. The Bayes factor shows the trend towards equality (<1) and difference (>1)

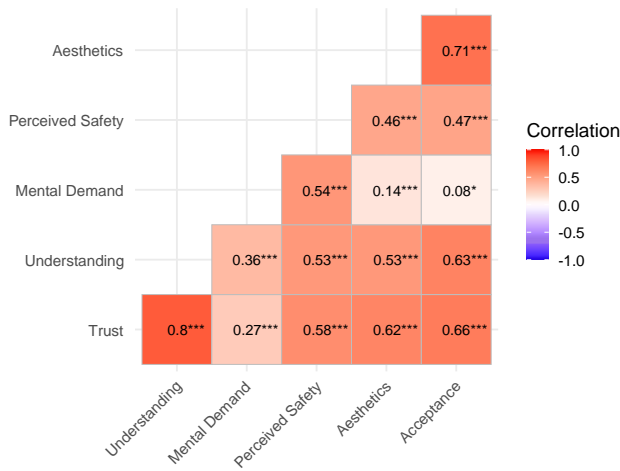


Figure 7: Correlation of all Objectives shows that most of them correlate and are not trade-offs to each other

makes it easier to simulate and measure realistic user responses. However, in UAM, there is no widespread base of experience, which can weaken the reliability of participant responses in simulations.

We, however, found strong to extreme evidence that motion fidelity negatively impacted trust, understanding, and acceptance (see subsection 3.7.4). This aligns with concerns about automated air taxis in prior (market) studies [1, 3, 23], suggesting that participants' lower ratings in the motion group may more accurately reflect how passengers will respond to future real-world

automated air taxi scenarios. While the MOBO process successfully optimized UI design to address these effects to some extent, resulting in slightly adjusted UI designs, it could not fully compensate for the negative impact of motion on trust, understanding, and acceptance. Although the MOBO process optimized UI design to mitigate these effects (see Appendix A), the optimized interface was unable to reach the level of trust, understanding, and acceptance with the no motion condition. This aligns with the theory of trust by Lee and See [45], who state that trust in automation depends on multiple factors beyond UI design, named environmental context including the user's self-confidence. Similarly, Hoff and Bashir [34] proposed that trust also depends on the pre-knowledge of the users. This suggests that while interface improvements can address some aspects of user experience, they alone cannot resolve fundamental trust deficits caused by the motion simulation.

Nevertheless, when studying user effects in the UAM context, we recommend using motion fidelity simulators to closely replicate participants' real-world ratings when actual/absolute ratings are necessary. However, since most air taxi interface design studies focus on comparing different interface designs [17, 75], they primarily require relative comparisons rather than absolute real-world ratings. For these comparative studies, simulating exact real-world user ratings during an air taxi flight is less critical, and even a monitor study such as by Valente et al. [75] might be sufficient.

4.2 Optimized Design Choices in Automated Air Taxi Interfaces

The comparison of optimized UI design parameters on the Pareto front revealed only minor evidence for differences or similarities between the motion and no-motion conditions. However, two strong/extreme evidences for differences emerged. First, there was strong evidence that the size of the chevrons representing other air taxis was smaller with motion than without. Second, there was extreme evidence suggesting that, in the motion condition, participants preferred to visualize boundary boxes around other air taxis, while in the no-motion condition, participants' ratings suggested not to display these boxes. Although these differences were subtle, as the IQR of the no-motion group overlaps with the 0.5 threshold line (see Figure 5), the preference for boundary boxes in the motion condition aligns with participants' reduced trust in this scenario. Participants experiencing motion likely felt a stronger need for visual confirmation that other air taxis were detected, reflecting the compensatory role of interface elements in addressing reduced trust in motion fidelity contexts. Further, participants' preferences for smaller chevrons in the motion group suggest that larger chevrons of other air taxis may be distracting when motion is added. However, given that simulations aim to replicate real-world conditions as closely as possible—and in reality, motion would always be present—we argue that boundary boxes should be included in all simulations, regardless of whether motion fidelity is present. Similarly, chevron sizes for other air taxis should remain small in all conditions to avoid distraction.

Beyond these specific differences, the overall impact of motion fidelity on UI design was minimal, suggesting that motion has only limited influence on most design parameters. Also, visually comparing individuals' UI designs (see Figure 4) yielded only minimal perceivable differences. We, however, observed that the starting point for all objectives was already relatively high (see Appendix A). Hence, it seems that participants found any UI combination to be sufficient as long as some form of visualization was present. This is further supported by the fact that none of the design parameters approached a value of zero, indicating that all visual elements were considered necessary to some extent. This aligns with findings from Colley and Meinhardt et al. [17], where even simple displays significantly improved users' trust ratings compared to having no visualization at all.

4.3 User Interface Optimization Process

The relatively high variation in the UI design parameters on the Pareto front indicates that a single, universally optimized design was unsuitable for the participants. While some design parameters, such as the alpha of the ego trajectory and the length of the other air taxis' trajectories, showed relatively low variance and consistent preferences among participants, other values showed high variation. This suggests that a one-size-fits-all approach to UI design may not be effective for automated air taxis. Even though most of the objectives constantly increased during the optimization process for both groups, creating personalized UI designs tailored to the individual user's needs may be more beneficial than focusing on a single optimized design. This is consistent with instrumental personalization,

as described in the framework by Fan and Poole [26] on personalization, which emphasizes adapting tools and systems to users' individual needs to improve efficiency and effectiveness, making it a relevant concept for UAM interface design. This approach has already been explored in the automotive domain, where Normark [60] found that “[...] *personalizable vehicle interfaces can improve both the usefulness and the user experience with the product*” [60, p. 744]. One could extend this concept by considering personalized flight styles, as previous research for AVs has already shown that personalizing driving styles (such as driving speed and acceleration) could increase passengers' trust and comfort [73]. Moreover, this preference for personalization extends beyond AVs. Emergency helicopter pilots have also preferred personalized UIs during their flights [35]. This suggests that such a tailored approach could be equally beneficial for inexperienced passengers in automated air taxis. However, design decisions must also consider practical constraints beyond trust and acceptance, such as cost, legal restrictions, or weight limitations of the hardware. Since UI design has a limited effect on trust compared to other influencing factors [45] (see subsection 4.1), a degree of creative freedom for designers should remain, allowing them to balance functional and aesthetic priorities. This freedom allows them to balance functional needs with aesthetic and practical considerations. One way to integrate this creative flexibility with BO is by restricting the algorithm's design space, enabling it to optimize within predefined boundaries that align with broader design priorities. For instance, Mo et al. [56] suggest using forbidden regions to exclude undesirable search-space areas before and during optimization, ensuring that results remain aligned with user needs and designers' goals.

Moreover, the feasibility of personalization depends on the context. While personalized interfaces may work well for individual air taxis, future visions of UAM also include shared Air Metros, described as “[...] *similar to today's public transportation systems such as subways and buses [...]*” [61, p. 4]. In such scenarios, where multiple passengers share the same space, tailoring the interface to a single user is impractical. Instead, a standardized interface design that simultaneously accommodates the needs of multiple passengers would be necessary, ensuring usability and functionality for all.

4.4 Trade-off in the Objectives

MOBO aims to find optimal design parameters that balance multiple objectives along the Pareto front, ensuring that improving one objective does not excessively compromise another [50]. This is particularly useful when trade-offs between objectives exist, such as spatial error versus completion time in 3D touch interaction [10]. However, our analysis revealed significant correlations among all six objectives, indicating that they are not in conflict with each other. This finding suggests that the set of objectives may be redundant in this study. Although selected based on prior UAM studies [17, 53, 75], future research could streamline them by focusing on the most distinct ones. Since trust and understanding, as well as acceptance and aesthetics, are strongly correlated (see Figure 7), choosing one from each pair would simplify future studies. Specifically, we recommend prioritizing acceptance over aesthetics, as acceptance relies on a validated scale [76], whereas the aesthetics

metric, as introduced in [17], was self-created without undergoing validation. Similarly, we propose focusing on trust directly rather than understanding, as understanding is only an underlying dimension of trust in automation [42].

Mental demand, with weak correlations to other objectives, should be retained as it seems to capture a distinct user effect. Lastly, perceived safety could be treated as a supplementary objective, as it is only moderately correlated to trust, mental demand, or acceptance. While these are preliminary observations, future research should further explore and validate the potential reduction and refinement of objectives in UAM studies to ensure more efficient design optimization. For example, relevant psychological models, such as *trust in automation* by Körber [42], suggest that understanding is an underlying dimension of trust, and therefore, they assess these dimensions separately in their questionnaire. While this distinction is important for in-depth psychological analysis, we argue that in the context of the MOBO process, due to the strong correlation between trust and understanding, separating them may not be more effective for optimization purposes.

4.5 Practical Guidelines for Interface Designs in Automated Urban Air Mobility

The findings from our study offer several practical insights for future research regarding future UIs for automated air taxis. Below, we provide key guidelines:

- **Use Motion Cues for Realism, But VR Alone Suffices for UI Comparisons:** When designing UAM interfaces, including motion fidelity simulations (e.g., using a motion chair) is crucial when requiring closer-to real-world user ratings. However, since motion fidelity did not have a strong impact on the UI design, it seems that only VR or even monitors without motion cues is sufficient for studies that compare UI designs.
- **Focus on Personalized Interfaces for Automated Air Taxis:** Due to the high variability in user preferences for the optimized UI, a one-size-fits-all approach for air taxi interfaces may not be effective. Instead, focus on developing personalized interface solutions that allow users to personalize elements based on their individual information needs, such as visualization of the ego trajectories.
- **Prioritize Providing Visual Information Over Fine-Tuning Details:** The high starting point for most objectives during the optimization process, especially in the seeding phase, suggests that any form of visualization in the interface is beneficial to contribute to passenger's trust, understanding and acceptance. Hence, rather than focusing on minor design adjustments, such as the exact length of the ego trajectory, ensuring the interface provides sufficient information to foster trust is more effective.

4.6 Limitations and Future Work

While we compared the effects on user's effect and optimal UI design for 3-DoF motion and without motion, we did not include more simulators into this study as done by Yeo et al. [82]. Hence, while we can derive our findings towards motion fidelity due to a 3-DoF motion chair, we cannot tell if other simulators might have yielded other results. Further, we did not assess simulator sickness.

Yet, previous work has shown that adding motion fidelity to simulations can reduce simulator sickness by better aligning physical and visual perceptions of motion in flight simulators [38, 40]. Conversely, mismatches between physical motion and visual cues can increase simulator sickness [51]. In our study, the movements of the 3-DoF motion chair were precisely matched to the air taxi's simulated motions, which likely minimized simulator sickness in the motion condition compared to the no-motion condition. Although simulator sickness was not the main focus of this study, evaluating it in future research could help clarify whether variations in UI evaluations are influenced by the presence of motion cues.

Additionally, we focused on subjective ratings as the primary objective for optimizing the UI design parameters. However, similar to the approach by Chan et al. [10], incorporating objective measures such as real-time physiological data (e.g., heart rate or galvanic skin response) could offer a more holistic view of user experience. This integration would allow future studies to create more comprehensive optimization criteria for air taxi interfaces by combining both subjective feedback and objective physiological responses. Additionally, factors beyond user-centered metrics should be considered when optimizing UI designs for air taxis. For instance, hardware weight limitations or compliance with legal regulations may also influence design decisions. These practical considerations might conflict with user-preferred designs, highlighting the need for balanced trade-offs in future research. Nonetheless, the subjective ratings in this study were answered on validated questionnaires (trust [42], understanding [42], perceived safety [25], mental demand [30]) or inspired by related work (acceptance [76], aesthetics [17]). While valid, future work should explore the relationship between these dimensions and more targeted subaspects of trust. It is also important to consider the potential redundancy of some questionnaire items due to strong correlations between certain measures (see subsection 4.4).

5 Conclusion

In this study, we explored the impact of motion fidelity in VR-simulated air taxi flights on passengers' effects and their interface design preferences using the MOBO approach. We conducted a between-subjects study with $N=40$ participants, divided into two groups of $n=20$, where one group experienced the simulation with motion cues using a 3-DoF motion chair, while the other group experienced it without motion cues, just in VR.

Our findings reveal that while motion fidelity had minimal influence on most UI design parameters, there was evidence that it reduced users' trust, understanding, and acceptance, highlighting its importance in future UAM studies. However, the lack of differences in immersion may be attributed to participants' lack of experience with real air taxi flights, which makes it difficult for them to rate the immersion of the simulations accurately. Moreover, the relatively high variance in the optimal design preferences on the Pareto front suggests that a one-size-fits-all approach may not be suitable for UI design in automated UAM contexts. Hence, although we were able to show that most of the objectives constantly increased during the optimization process, personalized user interfaces tailored to individual preferences may provide better user

experiences. Finally, due to the strong correlation between objectives like trust/understanding and acceptance/aesthetics, future studies may simplify the evaluation by focusing on fewer metrics. Optimizing one often improves others, making it unnecessary to evaluate all independently in the MOBO process.

Acknowledgments

We would like to thank Johannes Schöning for his valuable mental support during the *CHITogther 2024* in St. Gallen

References

- [1] European Union Aviation Safety Agency. 2021. *Study on the societal acceptance of Urban Air Mobility in Europe*. European Union Aviation Safety Agency. Retrieved March 11, 2022 from <https://www.easa.europa.eu/sites/default/files/dfu/uam-full-report.pdf>
- [2] Gautam Agrawal, Christina Bloebaum, Kemper Lewis, Kevin Chugh, Chen-Hung Huang, and Sumeet Parashar. 2004. Intuitive Visualization of Pareto Frontier for Multiobjective Optimization in n-Dimensional Performance Space. In *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*. American Institute of Aeronautics and Astronautics, Albany, New York. doi:10.2514/6.2004-4434
- [3] Christelle Al Haddad, Emmanouil Chaniotakis, Anna Straubinger, Kay Plötner, and Constantinos Antoniou. 2020. Factors affecting the adoption and use of urban air mobility. *Transportation Research Part A: Policy and Practice* 132 (2020), 696–712. doi:10.1016/j.tra.2019.12.020
- [4] Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. 2020. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Advances in Neural Information Processing Systems* 33. <http://arxiv.org/abs/1910.06403>
- [5] Keegan Bergman. 2018. Mixed-Reality and Traditional Flight Simulators for Urban Air Mobility Vehicles. In *Proceedings of the Wisconsin Space Conference*.
- [6] Guillaume Bouyer, Amine Chellali, and Anatole Lécuyer. 2017. Inducing Self-Motion Sensations in Driving Simulators Using Force-Feedback and Haptic Motion. In *2017 IEEE Virtual Reality (VR)*. 84–90. doi:10.1109/VR.2017.7892234
- [7] Eric Brochu, Tyson Brochu, and Nando De Freitas. 2010. A Bayesian interactive optimization approach to procedural animation design. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM, New York, NY, USA, 103–112.
- [8] Eric Brochu, Vlad M. Cora, and Nando de Freitas. 2010. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. <http://arxiv.org/abs/1012.2599> arXiv:1012.2599 [cs].
- [9] Lucas Bruck, Bruce Haycock, and Ali Emadi. 2021. A Review of Driving Simulation Technology and Applications. *IEEE Open Journal of Vehicular Technology* 2 (2021), 1–16. doi:10.1109/OJVT.2020.3036582
- [10] Liwei Chan, Yi-Chi Liao, George B Mo, John J Dudley, Chun-Lien Cheng, Per Ola Kristensson, and Antti Oulasvirta. 2022. Investigating Positive and Negative Qualities of Human-in-the-Loop Optimization for Designing Interaction Techniques. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 112, 14 pages. doi:10.1145/3491102.3501850
- [11] Suyog Chandramouli, Yifan Zhu, and Antti Oulasvirta. 2023. Interactive Personalization of Classifiers for Explainability Using Multi-Objective Bayesian Optimization. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (Limassol, Cyprus) (UMAP '23). Association for Computing Machinery, New York, NY, USA, 34–45. doi:10.1145/3565472.3592956
- [12] L D Chen, Yiannis Papelis, Ginger Watson, and Dario Solis. 2001. NADS AT THE UNIVERSITY OF IOWA: A TOOL FOR DRIVING SAFETY RESEARCH. (2001), 14.
- [13] Po-Wen Chiu and Christina Bloebaum. 2008. Hyper-Radial Visualization (HRV) for Decision-Making in Multi-Objective Optimization. In *46th AIAA Aerospace Sciences Meeting and Exhibit*. American Institute of Aeronautics and Astronautics, Reno, Nevada. doi:10.2514/6.2008-907
- [14] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–11. doi:10.1145/3411764.3445351
- [15] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 155, 11 pages. <https://doi.org/10.1145/3411764.3445351>
- [16] Mark Colley, Pascal Jansen, Enrico Rukzio, and Jan Gugenheimer. 2021. SwiVR-Car-Seat: Exploring Vehicle Motion Effects on Interaction Quality in Virtual Reality Automated Driving Using a Motorized Swivel Seat. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–26.
- [17] Mark Colley, Luca-Maxim Meinhardt, Alexander Fassbender, Michael Rietzler, and Enrico Rukzio. 2023. Come Fly With Me - Investigating the Effects of Path Visualizations in Automated Urban Air Mobility. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, IMWUT 7, 2 (2023), 52. <https://doi.org/10.1145/3596249>
- [18] Mark Colley, Max Rädler, Jonas Glimmann, and Enrico Rukzio. 2022. Effects of Scene Detection, Scene Prediction, and Maneuver Planning Visualizations on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 49 (jul 2022), 21 pages. doi:10.1145/3534609
- [19] Mark Colley, Max Rädler, Jonas Glimmann, and Enrico Rukzio. 2022. Effects of Scene Detection, Scene Prediction, and Maneuver Planning Visualizations on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 49 (jul 2022), 21 pages. doi:10.1145/3534609
- [20] Fabien Danieau, Julien Fleureau, Philippe Guillotel, Nicolas Mollet, Anatole Lécuyer, and Marc Christie. 2012. HapSeat: Producing Motion Sensation with Multiple Force-Feedback Devices Embedded in a Seat. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology - VRST '12*. ACM Press, Toronto, Ontario, Canada, 69. doi:10.1145/2407336.2407350
- [21] John J. Dudley, Jason T. Jacques, and Per Ola Kristensson. 2019. Crowdsourcing Interface Feature Design with Bayesian Optimization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300482
- [22] Mark Dunlop and John Levine. 2012. Multidimensional Pareto Optimization of Touchscreen Keyboards for Speed, Familiarity and Improved Spell Checking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI '12). Association for Computing Machinery, New York, NY, USA, 2669–2678. doi:10.1145/2207676.2208659
- [23] Thomas Edwards and George Price. 2020. eVTOL Passenger Acceptance. 27 pages. <https://ntrs.nasa.gov/citations/20200000532> Contractor Report.
- [24] EHang. 2020. *The future of Transportation: White Paper on Urban Air Mobility Systems*. EHang. <https://www.ehang.com/app/en/EHang%20White%20Paper%20on%20Urban%20Air%20Mobility%20Systems.pdf>
- [25] Stefanie M. Faas, Andrea C. Kao, and Martin Baumann. 2020. A Longitudinal Video Study on Communicating Status and Intent for Self-Driving Vehicle – Pedestrian Interaction. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376484
- [26] Haiyan Fan and Marshall Scott Poole. 2006. What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. *Journal of Organizational Computing and Electronic Commerce* 16, 3-4 (Jan. 2006), 179–202. doi:10.1080/10919392.2006.9681199
- [27] Loren Puchalla Fiore, Lane Phillips, Gerd Bruder, Victoria Interrante, and Frank Steinicke. 2012. Redirected Steering for Virtual Self-Motion Control with a Motorized Electric Wheelchair. doi:10.2312/EGVE/JVRC12/045-048 Artwork Size: 4 pages ISBN: 9783905674408 ISSN: 1727-530X Pages: 4 pages Publication Title: Joint Virtual Reality Conference of ICAT - EGVE - EuroVR.
- [28] Lukas A. Flohr, Joseph Sebastian Valiyaveetil, Antonio Krüger, and Dieter P. Wallach. 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 2123–2137. doi:10.1145/3563657.3596051
- [29] Julien Haan, Laurie A. Garrow, Aude Marzuoli, Satadru Roy, and Michel Bierlaire. 2021. Are commuter air taxis coming to your city? A ranking of 40 cities in the United States. *Transportation Research Part C: Emerging Technologies* 132 (2021), 103392. doi:10.1016/j.trc.2021.103392
- [30] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, Amsterdam, The Netherlands, 139–183.
- [31] Vincent Hayward, Jehangir Choksi, Gonzalo Lanvin, and Christophe Ramstein. 1994. Design and multi-objective optimization of a linkage for a haptic interface. *Advances in robot kinematics and computational geometry* (1994), 359–368.
- [32] Philipp Hock, Mark Colley, Ali Askari, Tobias Wagner, Martin Baumann, and Enrico Rukzio. 2022. Introducing VAMPIRE – Using Kinaesthetic Feedback in Virtual Reality for Automated Driving Experiments. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Seoul Republic of Korea, 204–214. doi:10.1145/3543174.3545252
- [33] Philipp Hock, Johannes Kraus, Marcel Walch, Nina Lang, and Martin Baumann. 2016. Elaborating Feedback Strategies for Maintaining Automation in Highly

- Automated Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Ann Arbor MI USA, 105–112. doi:10.1145/3003715.3005414
- [34] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.
- [35] Dominik Janetzko and Bacem Kacem. 2024. What Do You Need? Information Requirements and Task Analysis of (Future) Advanced Air Mobility Pilots in the Emergency Medical Service. *Aerospace* 11, 3 (Feb. 2024), 197. doi:10.3390/aerospace11030197
- [36] Christoph Albert Johns, João Marcelo Evangelista Belo, Anna Maria Feit, Clemens Nylandstedt Klokmoose, and Ken Pfeuffer. 2023. Towards Flexible and Robust User Interface Adaptions With Multiple Objectives. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–17. doi:10.1145/3586183.3606799
- [37] Florian Kadner, Yannik Keller, and Constantin Rothkopf. 2021. AdaptiFont: Increasing Individuals' Reading Speed with a Generative Font Model and Bayesian Optimization. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 585, 11 pages. doi:10.1145/3411764.3445140
- [38] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (July 1993), 203–220. doi:10.1207/s15327108ijap0303_3
- [39] Gwangbin Kim, Dohyeon Yeo, Taewoo Jo, Daniela Rus, and SeungJun Kim. 2023. What and When to Explain?: On-road Evaluation of Explanations in Highly Automated Vehicles. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (Sept. 2023), 1–26. doi:10.1145/3610886
- [40] Jiwon Kim, Jihong Hwang, and Taezoon Park. 2020. Effect of Motion Cues on Simulator Sickness in a Flight Simulator. In *Virtual, Augmented and Mixed Reality. Design and Interaction*, Jessie Y. C. Chen and Gino Fragoneri (Eds.). Vol. 12190. Springer International Publishing, Cham, 493–506. doi:10.1007/978-3-030-49695-1_33 Series Title: Lecture Notes in Computer Science.
- [41] Young Woo Kim, Cherin Lim, Seul Chan Lee, Sol Hee Yoon, and Yong Gu Ji. 2021. The 1st Workshop on User Experience in Urban Air Mobility: Design Considerations and Issues. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Leeds, United Kingdom) (*AutomotiveUI '21 Adjunct*). Association for Computing Machinery, New York, NY, USA, 175–177. doi:10.1145/3473682.3477440
- [42] Moritz Körber. 2019. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Sebastiano Bagnara, Riccardo Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita (Eds.). Springer International Publishing, Cham, 13–30.
- [43] Oswald Kothgassner, A Felnhofer, N Hauk, E Kastenhofer, J Gomm, and I Krysprin-Exner. 2013. Technology Usage Inventory. https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische%20programme/programmdokumente/tui_manual.pdf. *Manual. Wien: ICARUS* 17, 04 (2013), 90. [Online; accessed: 05-JULY-2020].
- [44] Yuki Koyama and Masataka Goto. 2022. BO as Assistant: Using Bayesian Optimization for Asynchronously Generating Design Suggestions. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (*UIST '22*). Association for Computing Machinery, New York, NY, USA, Article 77, 14 pages. doi:10.1145/3526113.3545664
- [45] J. D. Lee and K. A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (Jan. 2004), 50–80. doi:10.1518/hfes.46.1.50_30392
- [46] M.D. Lee and E.J. Wagenmakers. 2013. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press. <https://books.google.de/books?id=50tkAgAAQBAJ>
- [47] Yi-Chi Liao, John J Dudley, George B Mo, Chun-Lien Cheng, Liwei Chan, Antti Oulasvirta, and Per Ola Kristensson. 2023. Interaction Design With Multi-objective Bayesian Optimization. *IEEE Pervasive Computing* 22, 1 (2023), 29–38.
- [48] Cherin Lim, Young Woo Kim, Yong Gu Ji, Solhee Yoon, and Seul Chan Lee. 2022. Is This Flight Headed Downtown? : User Experience Considerations for Urban Air Mobility. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, Simone Barbosa, Cliff Lampe, Caroline Appert, and David A. Shamma (Eds.). ACM, New York, NY, USA, 1–7. doi:10.1145/3491101.3519852
- [49] Panadda Marayong, Praveen Shankar, Jessica Wei, Hanson Nguyen, Thomas Z. Strybel, and Vernol Battiste. 2020. Urban Air Mobility System Testbed using CAVE Virtual Reality Environment. In *2020 IEEE Aerospace Conference*. IEEE, Big Sky, MT, USA, 1–7. doi:10.1109/AERO47225.2020.9172534
- [50] R.T. Marler and J.S. Arora. 2004. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization* 26, 6 (April 2004), 369–395. doi:10.1007/s00158-003-0368-6
- [51] Michael Mcauley, Lawrence Hettinger, Thomas Sharkey, and John Sinacori. 1990. The effects of simulator visual-motion asynchrony on simulator induced sickness. In *Flight Simulation Technologies Conference and Exhibit*. American Institute of Aeronautics and Astronautics, Dayton, OH, U.S.A. doi:10.2514/6.1990-3172
- [52] Luca-Maxim Meinhardt, Mark Colley, Alexander Faßbender, Michael Rietzler, and Enrico Rukzio. 2023. Up, Up and Away - Investigating Information Needs for Helicopter Pilots in Future Urban Air Mobility. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI EA '23*). Association for Computing Machinery, New York, NY, USA, Article 304, 6 pages. doi:10.1145/3544549.3585643
- [53] Luca-Maxim Meinhardt, Mark Colley, Marjana Tahmid, Max Rädler, and Enrico Rukzio. 2024. Wind of Change: Investigating Information Visualizations for Passengers and Residents' Perception of Automated Urban Air Mobility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 4 (Nov. 2024), 1–31. doi:10.1145/3699753
- [54] Luca-Maxim Meinhardt, Maximilian Rück, Julian Zähnle, Maryam Elhaidary, Mark Colley, Michael Rietzler, and Enrico Rukzio. 2024. Hey, What's Going On? Conveying Traffic Information to People with Visual Impairments in Highly Automated Vehicles: Introducing OnBoard. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 2, Article 67, 24 pages. doi:10.1145/3659618
- [55] Olaf Mersmann. 2024. *emoa: Evolutionary Multiobjective Optimization Algorithms*. doi:10.32614/CRAN.package.emoa R package version 0.5-3.
- [56] George Mo, John Dudley, Liwei Chan, Yi-Chi Liao, Antti Oulasvirta, and Per Ola Kristensson. 2024. Cooperative Multi-Objective Bayesian Design Optimization. *ACM Transactions on Interactive Intelligent Systems* 14, 2 (June 2024), 1–28. doi:10.1145/3657643
- [57] Navid Mohajer, Hamid Abdi, Kyle Nelson, and Saeid Nahavandi. 2015. Vehicle Motion Simulators, a Key Step towards Road Vehicle Dynamics Improvement. *Vehicle System Dynamics* 53 (May 2015), 1–23. doi:10.1080/00423114.2015.1039551
- [58] Richard D. Morey, Jan-Willem Romeijn, and Jeffrey N. Rouder. 2016. The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology* 72 (June 2016), 6–18. doi:10.1016/j.jmp.2015.11.001
- [59] Lenart Motnikar, Marlen Jachek, Lukas Kröninger, Michael Gafert, Alexander G. Mirnig, Viktoria Sandor, Samuel Lesak, Arno Fallast, and Peter Fröhlich. 2023. Unravelling User Perspectives on Automated Urban Air Mobility: A Virtual Reality Approach. In *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Ingolstadt Germany, 317–319. doi:10.1145/3581961.3609850
- [60] Carl Jörgen Normark. 2015. Design and Evaluation of a Touch-Based Personalizable In-Vehicle User Interface. *International Journal of Human-Computer Interaction* 31, 11 (2015), 731–745. doi:10.1080/10447318.2015.1045240 arXiv:https://doi.org/10.1080/10447318.2015.1045240
- [61] Henry Pak, Lukas Asmer, Petra Kokus, Bianca I. Schuchardt, Albert End, Frank Meller, Karolin Schweiger, Christoph Torens, Carolina Barzantny, Dennis Becker, Johannes Maria Ernst, Florian Jäger, Tim Laudien, Nabih Naeem, Anne Papenfuß, Jan Pertz, Prajwal Shiva Prakasha, Patrick Ratel, Fabian Reimer, Patrick Sieb, Chen Zhu, Rabeb Abdellaoui, Richard-Gregor Becker, Oliver Bertram, Aditya Devta, Thomas Gerz, Roman Jaksche, Andreas König, Helge Lenz, Isabel C. Metz, Fares Naser, Lukas Schalk, Sebastian Schier-Morgenthal, Maria Stolz, Majed Swaid, Andreas Volkert, and Kristin Wendt. 2024. Can Urban Air Mobility become reality? Opportunities and challenges of UAM as innovative mode of transport and DLR contribution to ongoing research. *CEAS Aeronautical Journal* (May 2024). doi:10.1007/s13272-024-00733-x
- [62] Kim Parker. 2023. *About a third of U.S. workers who can work from home now do so all the time*. Retrieved Nov 2, 2023 from <https://www.pewresearch.org/short-reads/2023/03/30/about-a-third-of-us-workers-who-can-work-from-home-do-so-all-the-time/>
- [63] Bob Pishue. 2021. 2021 INRIX global traffic scorecard.
- [64] Bob Pishue. 2023. 2023 INRIX global traffic scorecard.
- [65] A. Pukhova, C. Llorca, A. Moreno, C. Staves, Q. Zhang, and R. Moeckel. 2021. Flying taxis revived: Can Urban air mobility reduce road congestion? *Journal of Urban Mobility* 1 (2021), 100002. doi:10.1016/j.urmb.2021.100002
- [66] Suchithra Rajendran and Jake Shulman. 2020. Study of emerging air taxi network operation using discrete-event systems simulation approach. *Journal of Air Transport Management* 87 (2020), 101857. doi:10.1016/j.jairtraman.2020.101857
- [67] Suchithra Rajendran and Joshua Zack. 2019. Insights on strategic air taxi network infrastructure locations using an iterative constrained clustering approach. *Transportation Research Part E: Logistics and Transportation Review* 128 (2019), 470–505. doi:10.1016/j.tre.2019.06.003
- [68] Michael Rietzler, Teresa Hirzle, Jan Gugenheimer, Julian Frommel, Thomas Dreja, and Enrico Rukzio. 2018. VRSpinning: Exploring the Design Space of a 1D Rotation Platform to Increase the Perception of Self-Motion in VR. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. Association for Computing Machinery, New York, NY, USA, 99–108. doi:10.1145/3196709.3196755
- [69] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. 2016. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* 104, 1 (Jan. 2016), 148–175. doi:10.1109/JPROC.2015.2494218
- [70] YAW Motion Simulator. 2023. *Virtual Reality Motion Simulator - Yaw VR*. YAW Motion Simulator. Retrieved Apr 19, 2023 from <https://www.yawvr.com/>

- [71] Mel Slater and Sylvia Wilbur. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 6, 6 (Dec. 1997), 603–616. doi:10.1162/pres.1997.6.6.603
- [72] I. M. Sobol. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *U. S. S. R. Comput. Math. and Math. Phys.* 7 (1967), 86–112.
- [73] Xu Sun, Jingpeng Li, Pinyan Tang, Siyuan Zhou, Xiangjun Peng, Hao Nan Li, and Qingfeng Wang. 2020. Exploring Personalised Autonomous Vehicles to Influence User Trust. *Cognitive Computation* 12, 6 (Nov. 2020), 1170–1186. doi:10.1007/s12559-020-09757-x
- [74] Unity Technologies. 2022. *Unity*. Unity Technologies. <https://unity.com/>
- [75] Lorenzo Valente, Filippo Gabriele Praticó, Marco Nobile, and Fabrizio Lamberti. 2024. Towards Adaptive AR Interfaces for Passengers of Autonomous Urban Air Mobility Vehicles: Analyzing the Impact of Flight Phases and Visibility Conditions on User Experience Through Simulation. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Orlando, FL, USA, 351–359. doi:10.1109/VRW62533.2024.00070
- [76] Jinke D Van Der Laan, Adriaan Heino, and Dick De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5, 1 (1997), 1–10.
- [77] Volocopter GmbH. 2021. *The roadmap to scalable urban air mobility*. Volocopter GmbH. Retrieved June 3, 2023 from <https://static1.squarespace.com/static/5d27bb3e330ac30001dc14fd/t/6165c265c89c61365aa517f5/1634058869091/Volocopter+v2+Roadmap+to+Scalable+UAM.pdf>
- [78] Annette Werner. 2018. New Colours for Autonomous Driving: An Evaluation of Chromaticities for the External Lighting Equipment of Autonomous Vehicles. doi:10.25538/tct.v0i1.692 Number: 1.
- [79] Matt N Williams, Rasmus A Bååth, and Michael C Philipp. 2017. Using Bayes Factors to Test Hypotheses in Developmental Research. *Research in Human Development* 14, 4 (Oct. 2017), 321–337. doi:10.1080/15427609.2017.1370964
- [80] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (apr 2016), 38–44. doi:10.1145/2907069
- [81] Marcel Woide, Mark Colley, Nicole Damm, and Martin Baumann. 2022. Effect of System Capability Verification on Conflict, Trust, and Behavior in Automated Vehicles. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) (*AutomotiveUI '22*). Association for Computing Machinery, New York, NY, USA, 119–130. doi:10.1145/3543174.3545253
- [82] Dohyeon Yeo, Gwangbin Kim, and Seungjun Kim. 2020. Toward Immersive Self-Driving Simulations: Reports from a User Study across Six Platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. doi:10.1145/3313831.3376787
- [83] Yijun Zhou, Yuki Koyama, Masataka Goto, and Takeo Igarashi. 2021. Interactive Exploration-Exploitation Balancing for Generative Melody Composition. In *Proceedings of the 26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) (*IUI '21*). Association for Computing Machinery, New York, NY, USA, 43–47. doi:10.1145/3397481.3450663

A Optimization Process

Figure 8 illustrates the optimization process across the individual objectives during the five seeding runs and 25 exploitation runs. Instead of optimizing the parameters to an individual optimum, we used MOBO to converge towards a Pareto front. This approach ensures that the UI design achieves a balanced optimization where no single objective can be improved without compromising another. We employed the Locally Estimated Scatterplot Smoothing (LOESS) method for data smoothing in the plots. Further, we set the span parameter to 1, which means that each point on the smoothed curve is influenced by the nearest 100% of the data points. This choice ensures that the curve accurately represents the overall trend, effectively minimizing the impact of short-term fluctuations and noise.

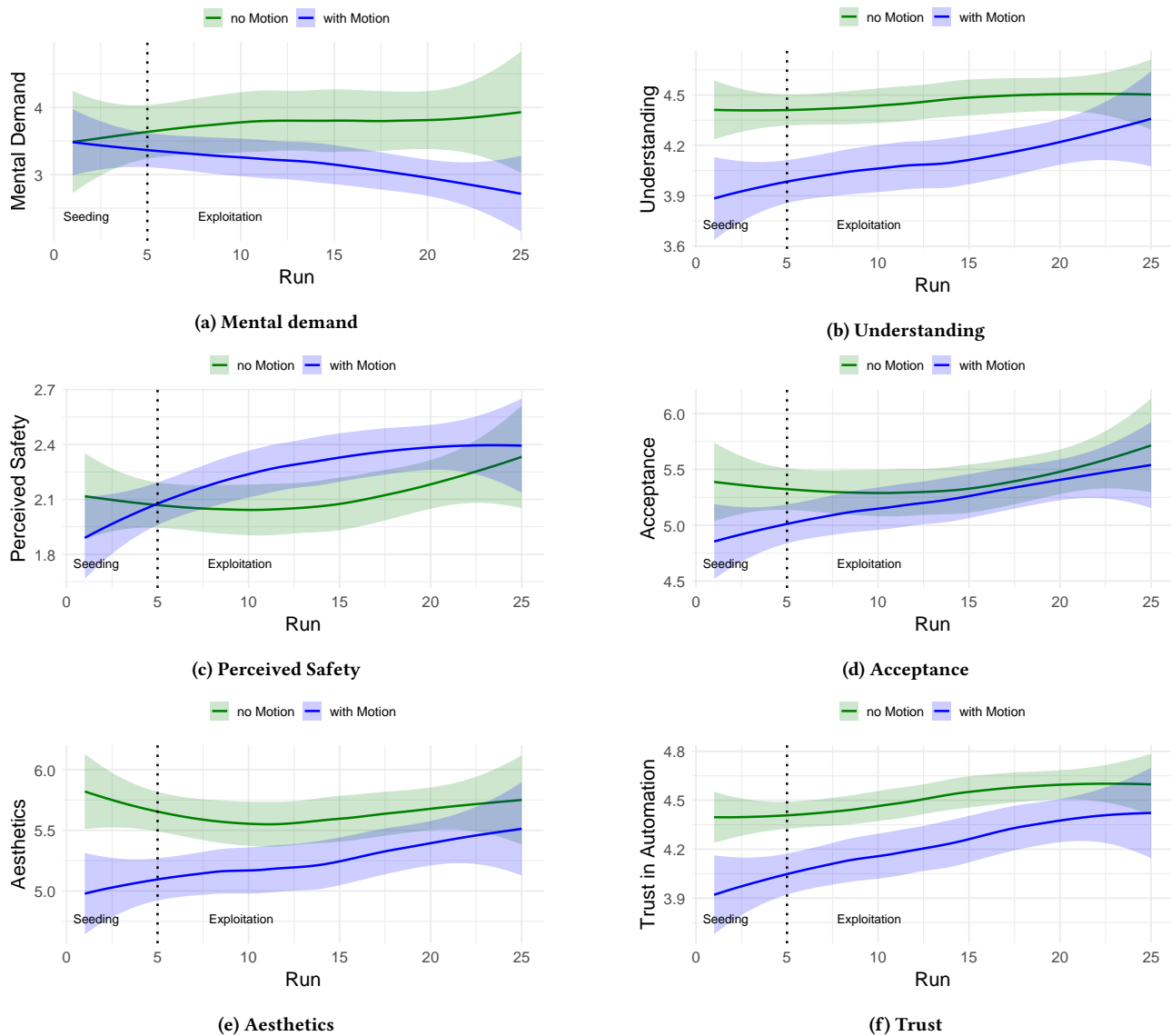


Figure 8: Optimization process of each objective during the 30 runs