Improving External Communication of Automated Vehicles Using Bayesian Optimization

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

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

Mugdha Keskar
mugdha.keskar@uni-ulm.de
Institute of Media Informatics, Ulm University
Ulm, Germany

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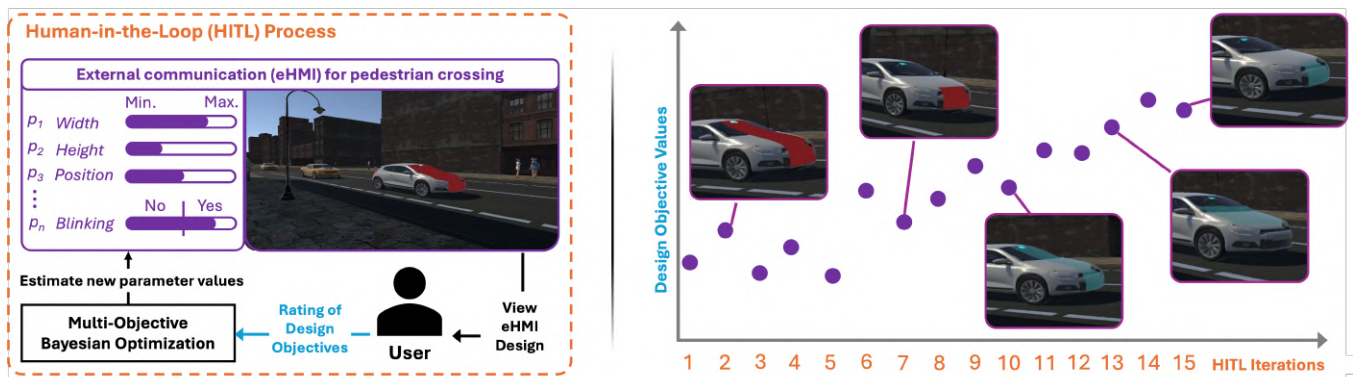Institute of Media Informatics, Ulm University
Ulm, Germany

Figure 1: Human-in-the-loop (HITL) MOBO of eHMI design in a pedestrian crossing scenario. The AV communicates its intention to stop and allow pedestrians to cross the street. The goal is to improve user ratings on multiple design objectives, such as trust, safety, acceptance, and aesthetics, while minimizing mental demand. MOBO iteratively uses design parameter values (p_1 to p_n) and user ratings of the design objectives to suggest optimized designs for each iteration.

Abstract

The absence of a human operator in automated vehicles (AVs) may require external Human-Machine Interfaces (eHMIs) to facilitate communication with other road users in uncertain scenarios, for example, regarding the right of way. Given the plethora of adjustable parameters, balancing visual and auditory elements is crucial for effective communication with other road users. With N=37 participants, this study employed multi-objective Bayesian optimization to enhance eHMI designs and improve trust, safety perception, and mental demand. By reporting the Pareto front, we identify optimal design trade-offs. This research contributes to the ongoing standardization efforts of eHMIs, supporting broader adoption.

*Both authors contributed equally to this research.

CCS Concepts

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

Keywords

External communication; Autonomous vehicles; Pedestrian Behavior; eHMI.

1 Introduction

Automated vehicles (AVs) are poised to significantly alter traffic dynamics [38] and interactions within traffic environments [22, 28, 58]. Communication tools become essential in the absence of a human operator to communicate with other road users in uncertain situations, such as determining the right of way. These tools are commonly referred to as external Human-Machine Interfaces (eHMIs).

Various modalities were proposed for eHMIs, including displays on grilles [39] and windshields [30], LED strips [39, 59], movement patterns [80], and projections [2, 64]. External devices like smartphones [45] and enhanced infrastructure were considered [72]. Additionally, various eHMI concepts and designs have been explored, such as text-based designs [12] that convey vehicle status [37] or intentions [34]. Research generally found positive effects of these eHMIs on trust, clarity, hedonic qualities, and pedestrian crossing behaviors.

Previous work evaluated a limited subset of possible eHMI designs. However, given the numerous design parameters, it is crucial to have a broader picture of the optimal eHMI. In the realm of user interfaces (UIs) for AV passengers, Normark [65] enabled passengers to manually customize the size, location, and color of icons on the dashboard, center stack, and Head-Up Display (HUD). By tailoring these interfaces to individual preferences, the user experience can be significantly enhanced, leading to increased perceived safety and trust.

Instead of relying on manual adjustments to different UI design aspects, Multi-Objective Bayesian Optimization (MOBO) offers a method to identify optimal design parameters iteratively (see Figure 1). MOBO has been successfully applied across various domains to address design optimization challenges [10, 11, 48, 50, 56]. MOBO predicts which design changes will most effectively meet desired objectives, such as enhancing passengers' trust. It manages multiple objectives by finding the best balance, represented by the Pareto front, ensuring the most effective trade-offs in UI design [61].

In consideration of these factors, this study is guided by the following research question (RQ):

> RQ1: What are the characteristics of an optimized eHMI for AVs that, among other objectives, enhance pedestrians' trust and perceived safety?

To explore this RQ, we conducted a user study with N=37 participants. We applied MOBO to optimize the eHMI's design (see Figure 1), focusing on the parameters: color, blink frequency, size, positioning, and loudness of the auditory component. These parameters were optimized based on the objectives: trust in automation, understanding, mental demand, perceived safety, acceptance, aesthetics, and the duration until starting to cross.

To explore design preferences across gender, we compared the parameters and objectives based on the participants' gender. Previous work [66] already showed that gender impacts crossing decisions. Therefore, Colley et al. [25] recommended also evaluating this factor for eHMI studies. Therefore, the second RQ was:

> RQ2: How do the parameters and objectives differ for gender?

In the virtual reality (VR) study, N=37 participants had to cross a two-lane road with traffic in both directions. The traffic was mixed, consisting of AVs and manually driven vehicles. On the near side of the road, AVs communicated their stopping using the eHMI design determined by the MOBO in each iteration.

The comparison of eHMI design parameters between female and male participants revealed no significant differences. However, qualitatively comparing the resulting parameter values on the Pareto front suggests that certain parameter ranges—such as cyan color and a 3Hz flashing animation—could serve as starting points for

all users and allow focusing on optimization of other parameters to increase efficiency. Unlike earlier LED strip-based approaches, using the entire front of the AV as an eHMI, combined with a high auditory volume, supports accessibility and multimodality. These findings offer a practical baseline to personalize and refine eHMIs for diverse users. Participants consistently rated the eHMI highly, reinforcing the need for explicit eHMIs, contrasting with studies questioning their necessity.

Contribution Statements [78]:

- **Artifact or System** We developed a VR, Unity-based simulation of a pedestrian crossing a non-signalized street, designed to optimize the eHMI through MOBO iteratively, 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=37) to investigate the impact of gender on user experience and UI design for eHMIs. Our findings show no significant differences between male and female participants. The MOBO process led to very high objective scores.

2 Related Work

We base our work on research on pedestrian-vehicle interactions, focusing on crossing behaviors in traffic and the role and influence of eHMIs in AVs on traffic interactions. Additionally, it presents BO to optimize eHMIs.

2.1 External Communication of Automated Vehicles

Current traffic interactions frequently rely on gestures and eye contact to resolve ambiguities [68]. Although explicit communication is rarely required [54], eHMIs were proposed to facilitate interactions between AVs and vulnerable road users such as pedestrians and cyclists [44]. Previous research has categorized external communication strategies by modality, message type, and vehicle location [22, 23, 28]. Colley and Rukzio [22] identified eight types of messages: Instruction, Command, Advisory, Answer, Historical, Predictive, Question, and Affective. Communication can occur at various locations on the vehicle, through personal devices, or via infrastructure like sidewalks, with key interaction points such as the windshield or bumper being particularly important [33]. Most work and the ISO technical report [1] recommend communicating the intention of the AV instead of giving advice or commands such as "Go". Therefore, our optimization focused on transmitting this information.

Effective eHMI deployment includes considering the communication relationship (ranging from one-to-one to many-to-many), ambient noise levels, and road user (e.g., pedestrian, cyclist) [22]. Research examined eHMI effectiveness across different groups, including children[13, 27], individuals with vision, mobility, or cognitive impairments [4, 24, 42], and general pedestrians [3, 31, 32, 58], and cyclists [47]. Generally, eHMIs have shown positive results. For instance, Dey et al. [30] demonstrated that distance-dependent information could significantly improve pedestrians' understanding of AV intentions and willingness to cross safely. Colley et al. [24]

found that visually impaired individuals preferred clear, speech-based communications over other forms. To address accessibility challenges, our optimization also includes speech output.

However, challenges remain, such as ensuring children accurately interpret and use eHMIs [27], concerns about overtrust [46], scalability [17, 18, 26], and exploration of the social implications of eHMIs [16, 52, 69, 70]. Furthermore, previous work indicated that instead of defining novel eHMIs, research should focus on consolidating previous approaches [32]. With this optimization of well-known parameter constraints (e.g., position), we contribute to this consolidation.

2.2 Multi-Objective Bayesian Optimization for Interface Designs

The design of UIs involves selecting parameter values such as element positioning, color, and transparency to achieve specific design objectives like trust and perceived safety. These objectives can be represented as objective functions, with their values (e.g., trust ratings) being determined by specific combinations of design parameters. Since predicting the best parameter combinations is challenging due to the unknown relationship between parameters and objective functions, this problem can be framed as an optimization task [9].

Given the impracticality of testing all potential design combinations due to the vast design space, BO offers an efficient approach to modeling the relationship between parameter combinations and objective function values. BO is particularly suitable for optimizing unknown and complex functions with limited prior data [7, 10, 55]. It balances *sampling* of new parameter spaces and the *optimization* of known promising regions, leading to optimal designs with fewer iterations.

BO has been applied successfully in various UI design contexts, such as fine-tuning animation appearances [8], customizing image aesthetics [51], optimizing font settings [48], and improving UI interactions [35]. In Human-Computer Interaction (HCI), obtaining objective function values requires user feedback often, making HITL optimization a valuable method. HITL adds a user-centered feedback loop to BO, allowing iterative refinement of design parameters based on user interactions and ratings [14, 51, 79].

However, the design of an eHMI involves multiple objectives, necessitating the use of MOBO. MOBO optimizes several objectives simultaneously, resulting in a Pareto front representing a range of optimal designs, balancing trade-offs between conflicting objectives [61]. MOBO has been applied in diverse areas, including multi-finger text entry [74] and personalized explanations for image classifiers [11].

Despite its potential, HITL MOBO was not explored in eHMI design with its unique challenges due to subjective objectives like perceived safety and trust, considering the vulnerability of road users like pedestrians and the dynamic nature of traffic, which our work addresses.

3 External Communication Concept

To simulate our eHMI concept, we used Unity (version 2022.3). We opted for a VR simulation using the HTC VIVE Pro, a common method for immersive investigation of eHMIs (e.g., see Colley et al.

[20]). VR is an appropriate choice as testing pedestrian crossing scenarios while manipulating the eHMI design can endanger road users in the real world.

3.1 Scenario

The pedestrian scenario starts at a sidewalk. The participant needs to cross a two-lane road. Road users in this scenario are manually driven vehicles, AVs, and the participant. On the lane closer to the participant, AVs and manually driven vehicles drive at a 50:50 rate. Only manually driven vehicles are in the far lane.

After 18s, the next AV on the near lane will stop and let the participant cross. This duration lets the participant experience AVs and manually driven vehicles, representing an externally valid scenario. The manually driven vehicles on the far lane do not stop.

3.2 Generating External Communication

We defined an eHMI that employs a visual display on the hood, windshield, and grille inspired by eHMIs that use LED light panels (e.g., see [39, 59]). This area is visible from a pedestrian's perspective when driving towards the pedestrian. The area covers the AV's full width and height but excludes headlights and front lights.

The eHMI's size and position can vary horizontally and vertically, such as a narrow strip on the hood or a small display on the windshield's upper edge. However, the *optimal* position and size for pedestrian communication are unclear. A larger display may be visible from a distance but unclear in its message, while a smaller, more specific display (e.g., on the grille) may be less noticeable.

Therefore, we developed a method to generate an eHMI in Unity that can be positioned and resized dynamically across the windshield, hood, and radiator grille. The parameters and bounds of *width* and *height* for this generation are depicted in Figure 3. We always center the mesh (i.e., the *width*) horizontally on a *vertical position* axis (see Figure 3 b) for better visibility across the AV's curved surface if standing directly left/right from the AV. For instance, in Figure 2 a, the AV (partly) obstructs a small eHMI mesh on the left side of the hood.

4 Experiment

To answer RQ1 and RQ2, we conducted an experiment using HITL MOBO (see Figure 1) to optimize eHMI design in a pedestrian crossing scenario.

4.1 Bayesian Optimization: Parameterizing the eHMI and the Objective Functions

In the following, we describe the eHMI design parameters and outline our HITL MOBO setup, which iteratively adjusts these toward improving *objective function* values across iterations.

4.1.1 Design Parameters. eHMI design parameters were derived from the respective publications [17, 22, 24, 29, 32]. While RGB coloring of eHMIs should avoid unintended meaning (e.g., orange being a warning signal), we included it as a parameter to answer our RQ of the optimal eHMI. The alpha level was chosen to allow for the absence of an eHMI (i.e., if the alpha is 0.1, the eHMI is not visible). Some work has evaluated blink frequency [29], but the

Figure 2: The Unity implementation of the crossing scenario from the participant's starting point of view (a) and from above (b).
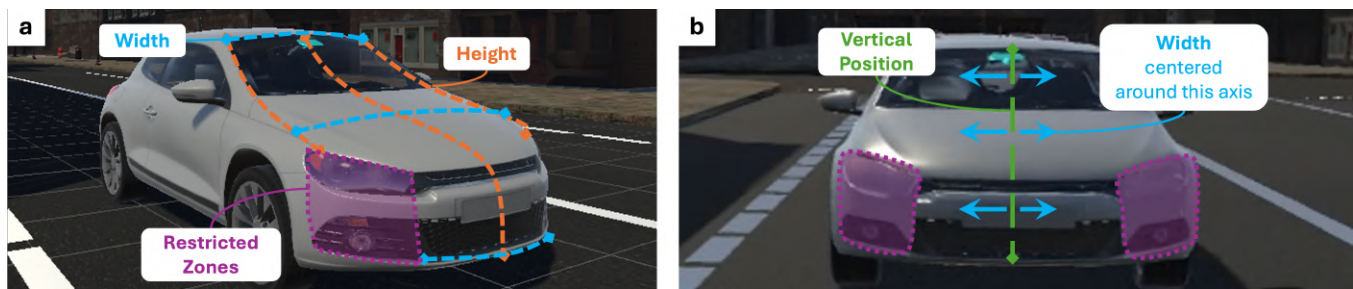

Figure 3: The white AV without an eHMI. A cyan light signal near the rearview mirror indicates that it is an AV. In (a), dashed lines show the available *width* and *height* for the eHMI mesh generation, including its lower/upper and left/right bounds. (b) Examples of possible mesh positions in allowed regions. The mesh (i.e., its *width*) is always centered horizontally along the green dashed axis to ensure visibility from both sides. The *vertical position* can be set along this axis. The eHMI does not cover the front lights (restricted zones; in purple).

evidence is minimal. The formula for blink frequency is:

$$\text{Blink frequency (Hz)} = \frac{1}{\frac{1}{\text{blinkFrequencyMOBO}} \times 0.25} \quad (1)$$

For example, a value of *0.8* equals 3.2 Hz. Finally, the size and position of the visual eHMI have to be determined. We allow any given rectangular shape along the outline of the front of the AV. Regarding auditory design, we opted for the textual, auditory message "Stopping" and varied its loudness. A minimal loudness means no auditory component. All design parameters (p_1 to p_9) are summarized in Table 1.

Other design elements, such as text-based messages, symbols, projections [17], eye-metaphors [41], biology-inspired designs [67], or hands [40, 60] are possible. However, recent works converged toward an LED stripe, and we assume that manufacturers will focus on such an aesthetic, easy-to-implement, and versatile solution.

4.1.2 Objective Functions. An objective function f maps a specific eHMI design x to a metric the optimizer aims to maximize. We focus on maximizing six subjective metrics: *perceived safety*, *trust*, *predictability*, *usefulness*, *satisfaction*, and *visual appeal*. Conversely, we aim to minimize *mental demand* as our sole subjective metric, along with one objective metric: *time to start crossing*.

Based on previous eHMI work [17], we employed the following questionnaires after every optimization run in the HITL process:

We assessed **mental demand** via the mental demand subscale of the raw NASA-TLX [43] on a 20-point scale ("How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?"; 1=*Very Low* to 20=*Very High*; lower is better). Regarding predictability and trust, we used the subscales *Predictability/Understandability* (*Predictability*) and *Trust* of the *Trust in Automation* questionnaire by Körber [49]. **Predictability** is determined via agreement on four statements (two direct: "The system state was always clear to me.", "I was able to understand why things happened."; two inverse: "The system reacts unpredictably.", "It's difficult to identify what the system will do next.") using 5-point Likert scales (1=*Strongly disagree* to 5=*Strongly agree*). **Trust** is measured via agreement on the same 5-point Likert scale on two statements ("I trust the system." and "I can rely on the system."; both times, higher is better). Participants rated their perceived **safety** using four 7-point semantic differentials from -3 (anxious/agitated/unsafe/timid) to +3 (relaxed/calm/safe/confident; higher is better) [37]. Finally, we added three single items. Two were defined with the van der Laan acceptance scale [77] in mind ("I find the visualizations of the automated vehicle **useful**", "I find the visualizations of the automated vehicle **satisfying**"). These were combined into a single "acceptance" objective. We also adapted the question regarding **visual appeal** from Colley et al. [21] ("I found the visualizations visually appealing"; on a 5-point Likert scale).
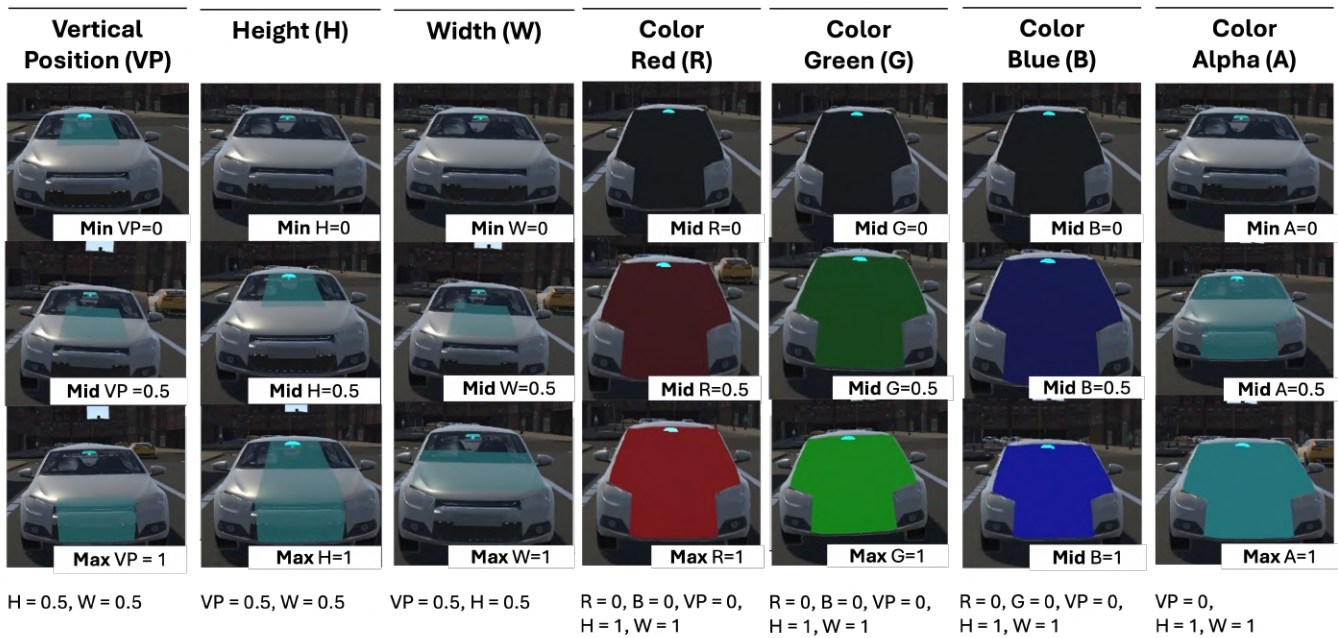
Figure 4: eHMI design parameter value ranges. At the bottom of each column, the default values for the other parameters are shown. Cyan is used as the default color to demonstrate other parameter values.

Table 1: The 9 design parameters for the eHMI design, with ranges. All design parameters are modeled continuously. Example visualizations of parameter values are shown in Figure 4.

Design Parameter	Description	Reference	Range
p_1: R	**R**ed color channel.	[23]	[0, 1]
p_2: G	**G**reen color channel.	[23]	[0, 1]
p_3: B	**B**lue color channel.	[23]	[0, 1]
p_4: eHMI Alpha, α	Alpha color channel value.	[23]	[0, 1]
p_5: Blink frequency	Blink frequency when activated. Maximum: 4Hz.	[29]	[0, 1]
p_6: Width, W	Width. Maximum: entire width of the AV.	[23]	[0, 1]
p_7: Height, H	Height. Maximum: entire height of the AV.	[23]	[0, 1]
p_8: Vertical Position, VP	Position along the center AV axis (viewed from the front).	[23]	[0, 1]
p_9: Auditory Message Loudness, l	Loudness of auditory message "Stopping".	[24]	[0, 1]

Time to start crossing is the duration after walking very close to the road since the simulation started. We used a Unity collider to mark the spot where the participant was counted as on the road. (see Figure 2 b).

Normalization into the $[-1, 1]$ range is required because the subjective metrics values have ranges based on 20-, 5-, or 7-point Likert scales. After transformation, the *mental demand* and *time to start crossing* objectives are to be maximized (a higher value means less load and earlier crossing).

4.1.3 Hyperparameter Setup for Bayesian Optimization. We used BoTorch [6] version 0.11.3 with a multi-output Gaussian Process and applied qEHVI as the acquisition function. This function represents the expected hypervolume increase, where we set $q = 1$

(see Chan et al. [10]) to ensure that after each iteration, a batch of size one is selected for evaluation. The optimization process started with a *sampling* phase of five iterations, during which we employed Sobol sampling [73] to generate initial eHMI designs. Sobol sampling systematically divides the design space into evenly distributed regions and picks a representative design from each, ensuring broad coverage. At this early stage, the MOBO algorithm has no prior user-specific data, so it needs these initial ratings to build a first understanding of the design space from each individual's perspective. To prevent bias from different starting conditions, we used the same set of five sampled designs for every participant. After gathering initial ratings, the MOBO used a 15-iteration *optimization* phase, where the optimizer iteratively balanced "exploitation" (refining

promising known designs) and "exploration" (probing new regions) for each user. To approximate the acquisition function, we used 2024 restart candidates and 512 Monte Carlo samples (see Chan et al. [10]).

4.1.4   Optimization Stopping Criterion. In internal tests, we found convergence to an optimal rating of objectives was reached relatively quickly. Therefore, we added a stopping criterion checked after every measurement: Was the perfect rating for **every** subjective metric (i.e., the **highest** rating for trust, predictability, safety, visual appeal, usefulness, satisfaction, and the **lowest** rating for mental demand; see Section 4.1.2) given for the **last** round? Participants could otherwise not opt out of optimization.

The stopping criterion **never** applied.

4.2   Additional Measurements

Besides objectives, we logged position with 50Hz, number of collisions with cars, time on the sidewalk and on the street, and total duration.

Finally, participants assessed the communication (intention to stop, intention not to stop, and timer) regarding necessity and reasonability on individual 7-point Likert scales and gave open feedback.

User Expectation Conformity: "The final design matches my imagination.", *Satisfaction*: "I'm pleased with the final design.", *Confidence*: "I believe the design is optimal for me.", *Agency*: "I felt in control of the design process." *Ownership*: "I feel the final design is mine." Regarding *Interactivity*, participants also provided feedback on desired design control levels ("... Consider aspects where you desired more or less control over the design.")

4.3   Procedure

First, participants were introduced to the study procedure and VR scene. They then signed informed consent. The introduction to the setting is in Appendix A.

They then signed informed consent and could adjust the headset. In each trial, participants crossed the same street with the same setting to guarantee no external effects altering the perception of the eHMI. An AV would stop after ≈25 seconds. With crossing the street, each trial took ≈1min. After each trial, participants answered the questionnaires described in subsection 4.2 in the VR scene. The study took ≈55min. Participants received 10€. The study was conducted in German and English.

The experimental procedure followed the guidelines of our university's ethics committee and adhered to regulations regarding handling sensitive and private data, anonymization, compensation, and risk aversion. Compliant with our university's local regulations, no additional formal ethics approval was required.

5   Results

5.1   Data Analysis

The goal of MOBO is to identify the Pareto front, which contains all Pareto optimal points in the design space. Each point represents a design that balances trade-offs between conflicting objectives [61]. Using the *EMOA* R package [62], we determined Pareto optimal values for each participant and focused our analysis solely on these

efficient designs. The female group yielded 76 Pareto designs, and the male group 90. R 4.4.2 and RStudio 2024.09.1 were employed. All packages were up-to-date in December 2024.

5.2   Participants

37 participants (Mean age = 25.6, SD = 3.5, range: [19, 34]; Gender: 45.9% women, 51.4% men, 2.70% non-binary;) took part in the study, recruited locally. 35 participants are college students, two are working, indicating that high school is their highest degree. On 5-point Likert scales (*1=Strongly Disagree — 5=Strongly Agree*), participants showed interest in AVs (*M*=4.51, *SD*=.77), were positive whether AVs would ease their lives (*M*=4.40, *SD*=.80), and were skeptical about whether they become reality by 2034 (*M*=4.30, *SD*=1.02). For gender comparison, we only used data from female and male participants due to the low number of participants with non-binary gender.

5.3   Questionnaire Ratings

We analyzed the mean ratings from the questionnaires of all participants whose design parameters were on the Pareto front. Figure 6 shows a comparison overview. Additionally, we show the value progression in Figure 12a, Figure 12b, Figure 13b, Figure 11b, Figure 13a, Figure 11a, and Figure 14, showing that the approach could optimize for the objective values over the iterations. In particular, perceived safety was increased, while mental demand and time to cross decreased. The increase in trust, predictability, and acceptance was weaker. Aesthetics remained roughly constant over the iterations.

Trust in Automation. The Bayesian analysis of Trust resulted in a $BF = 0.37 \pm 0.00\%$, suggesting inconclusive evidence for **no** difference between females (M=4.03, SD=1.05) and males (M=3.81, SD=1.12; see Figure 6a).

Predictability. For Predictability, the analysis yielded a $BF = 0.91 \pm 0.00\%$, providing inconclusive evidence for **no** differences between the two groups. Females rated Predictability higher (M=3.56, SD=0.88) than males (M=3.30, SD=0.86; see Figure 6b).

Mental Demand. Mental Demand showed a $BF = 0.23 \pm 0.02\%$, indicating moderate evidence for equality between the groups. Hence, female ratings (M=7.18, SD=2.65) than male ones (M=7.43, SD=2.08; see Figure 6c) for the designs were equal.

Perceived Safety. Perceived Safety showed a $BF = 6.00 \pm 0.00\%$, with moderate evidence favoring difference between the groups. The designs were rated lower for females (M=1.53, SD=1.50) than males (M=2.10, SD=1.14; Figure 6d).

Acceptance. Acceptance had a $BF = 0.17 \pm 0.05\%$, suggesting moderate evidence in favor of **no** difference between females (M=5.98, SD=1.00) and males (M=5.93, SD=1.25; see Figure 6e).

Aesthetics. The Bayesian analysis yielded a $BF = 0.27 \pm 0.04\%$ for the Aesthetics measure, indicating moderate evidence for **no** difference. The designs yielded a higher rating for females (M=6.09, SD=0.84) than males (M=5.92, SD=1.27; see Figure 6f).

Time To Start Crossing. The Bayesian analysis yielded a $BF = 3.50 \pm 0.01\%$ for Time to start crossing, indicating moderate evidence
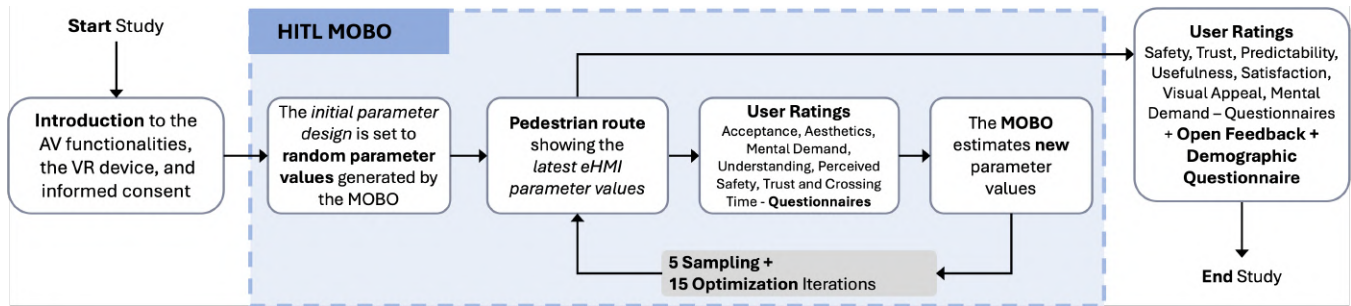
Figure 5: Study procedure using HITL MOBO for eHMI design. There were 20 (5 sampling and 15 optimization) iterations.

(a) Trust (b) Predictability (c) Mental Demand (d) Perceived Safety (e) Acceptance (f) Aesthetics (g) Time To Start Crossing (s)
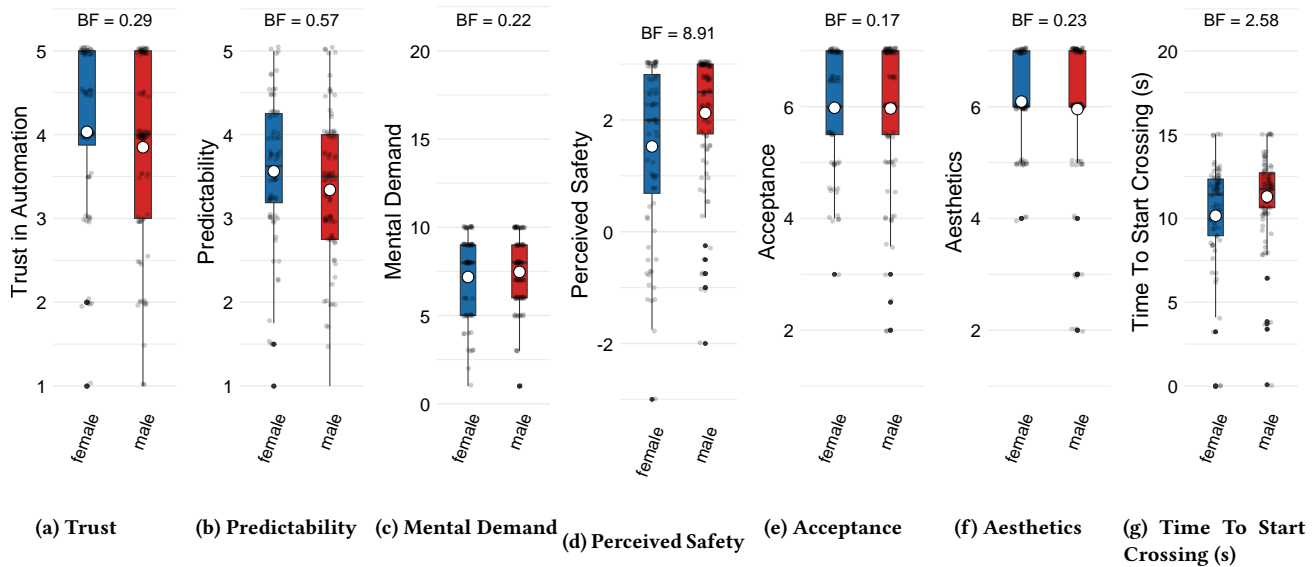
Figure 6: Rating for the subjective questionnaires comparing *female* and *male* of Pareto-optimal values. The Bayes factor shows trends towards equality (<1) and difference (>1).

favoring a difference. The designs yielded a faster start for females (M=10.2, SD=3.55) than males (M=11.4, SD=2.52; see Figure 6g).

5.4 Pareto Front Parameter Set

The results for each design parameter are detailed in Table 2. Interestingly, we found no evidence for difference for **any** parameter. Figure 7 shows this visually.

Figure 8 visualizes the final mean parameter set on the Pareto front in the scene. We use the commonly used color cyan [29] as a reference. The values between minimum, mean, and maximum do not vary widely.

Because averaging RGB values would obscure meaningful differences, we plot all resulting color values on the Pareto-front per participant across the 20 iterations in Figure 9. Each column shows a single participant's Pareto-true color designs over time. Each colored box represents one iteration where a given design was Pareto-optimal (i.e., non-dominated), and empty spaces indicate

dominated designs and, thus, not on the Pareto front. While various colors emerged, hues close to cyan were consistently present. This recurrence suggests that "standard" cyan hues remain a robust choice even in a process driven by iterative optimization. During sampling (iterations 1–5), in which all users saw the same initial five designs, 34 out of 37 participants had at least one Pareto-optimal color, suggesting that the initial sampling effectively spanned the design space. However, fewer participants achieved Pareto-optimal colors during optimization (iterations 1–15), where MOBO focused on refining promising regions. Some participants even had more than five Pareto-optimal colors in this phase, highlighting individual preferences and indicating that more iterations may be needed to identify optimal colors for certain users reliably.

5.5 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 [61]. To assess
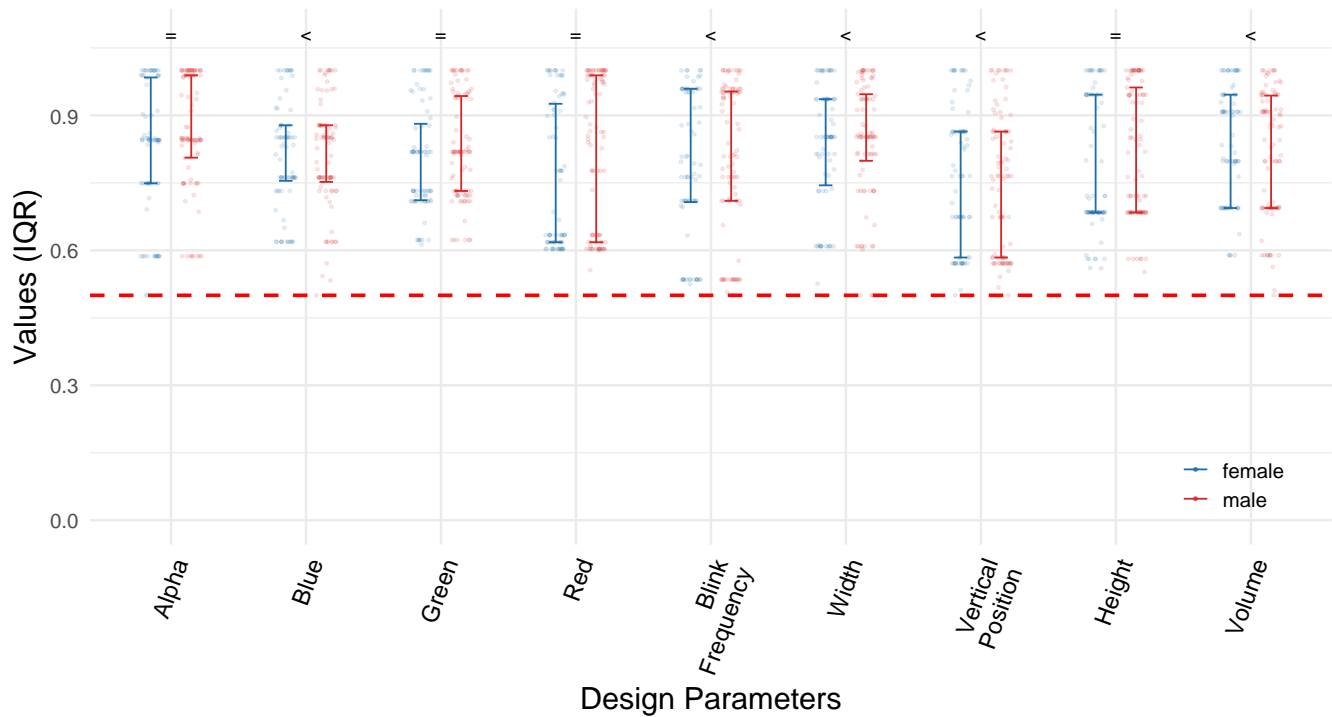
Figure 7: Comparison of the design parameters for both groups, using a Bayesian t-test. The annotations are as defined by Lee and Wagenmakers [53]: "<<<" for extreme evidence for equality (BF < 0.01), "<<" for strong or very strong evidence for equality (BF < 0.1), "<" for moderate evidence for equality (BF < 0.3), "=" for inconclusive (also called 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 is a reference line at 0.5

Table 2: Results of Bayesian Analysis for Each Design Parameter including the IQR ranges

Design Parameter	BF (± %)	Female IQR	Male IQR	Evidence
Alpha	0.71 ± 0.02%	0.75-0.98	0.81-0.99	inconclusive evid. for equality
Blue	*0.17 ± 0.05%*	*0.75-0.88*	*0.75-0.88*	*moderate evid. for equality*
Green	0.39 ± 0.03%	0.71-0.88	0.73-0.94	inconclusive evid. for equality
Red	1.40 ± 0.01%	0.62-0.93	0.62-0.99	inconclusive evid. for difference
Blink frequency	*0.19 ± 0.05%*	*0.71-0.96*	*0.71-0.95*	*moderate evid. for equality*
Width	*0.22 ± 0.05%*	*0.74-0.94*	*0.80-0.95*	*moderate evid. for equality*
Vertical position	*0.19 ± 0.05%*	*0.58-0.86*	*0.58-0.86*	*moderate evid. for equality*
Height	0.37 ± 0.03%	0.68-0.95	0.68-0.96	inconclusive evid. for equality
Volume	*0.21 ± 0.05%*	*0.69-0.95*	*0.69-0.94*	*moderate evid. for equality*

whether there are trade-offs between these objectives, we calculated the correlations among all objectives (see Figure 10). This analysis helps us understand how changes in one objective might influence others.

The results show that almost all correlations were statistically significant when using all data (see Figure 10a). In particular, Trust and Predictability had a strong positive correlation ($r = 0.65$), as did acceptance and aesthetics ($r = 0.71$), meaning improvements in one were associated with improvements in the other. In contrast, the correlations between Mental Demand and the other objectives

were negative. There were no correlations between Time to start crossing with Predictability and Perceived Safety.

When only taking Pareto front values into account (see Figure 10b), much fewer significant correlations exist. However, Trust still had a strong positive correlation with Predictability ($r = 0.54$) and Perceived Safety ($r = 0.43$), for example. Compared to when all data was used, all correlations became less strong and no correlation inverted.

Minimum Parameter Values	**Mean** Parameter Values	**Maximum** Parameter Values

VP = 0.606, H = 0.651, W = 0.705, RGB = *Cyan*, A = 703

VP = 0.759, H = 0.802, W = 0.839, RGB = *Cyan*, A = 0.843

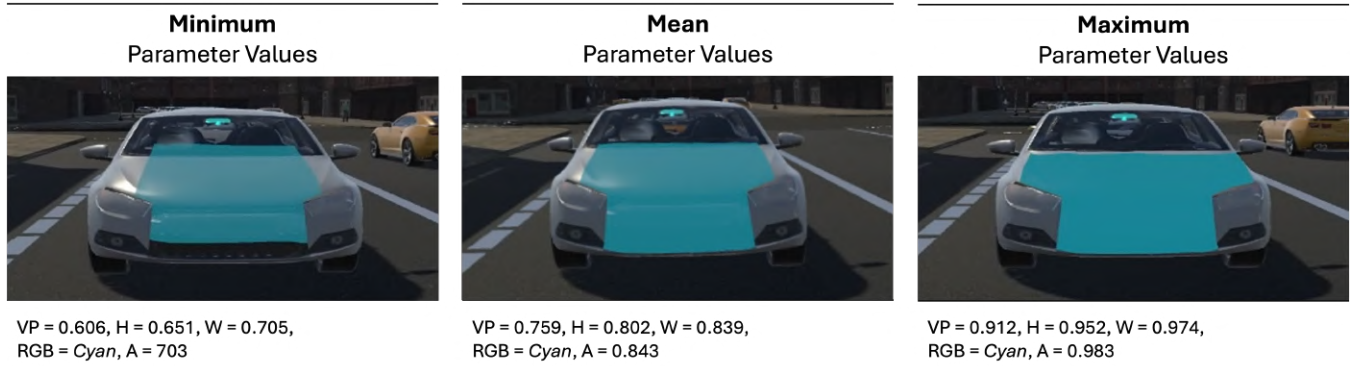VP = 0.912, H = 0.952, W = 0.974, RGB = *Cyan*, A = 0.983

Figure 8: Visualizations of the resulting design parameter value ranges on the Pareto front. Cyan was used as a default color (RGB) to demonstrate the non-color-related values: vertical position (VP), height (H), width (W), and alpha (A). Minimum/Maximum shows the lowest/highest values across participants where the design received Pareto-optimal ratings.


Figure 9: Color distribution for the Pareto front, split into sampling (1-5) and optimization (1-15) phases. The number in a box indicates the iteration in which this color was set, counting from the bottom. The columns on the x-axis represent the individual participants. An empty spot in a column means this iteration was not Pareto-true (i.e., dominated by another design).

5.6 Design Satisfaction and Engagement

Participants evaluated several design process aspects using Likert scales ranging from 1 (Strongly Disagree) to 7 (Strongly Agree).
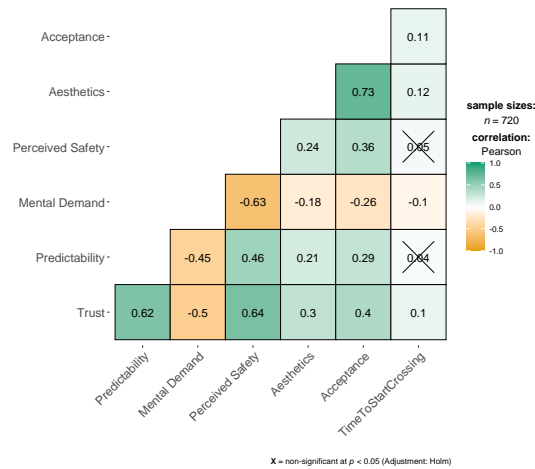
(a) All objectives.
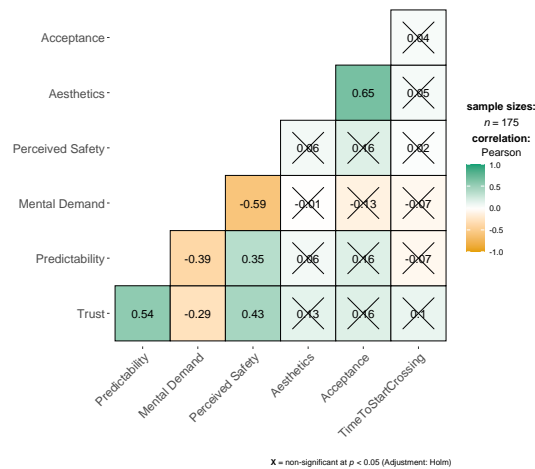
(b) Only Pareto-true objectives.

Figure 10: Correlation heatmaps of study objectives. (a) includes all data points ($n = 720 = 20 * 37$, i.e., 20 iterations for 37 participants), and (b) includes only Pareto-true data points. "x" indicates non-significant at $p < 0.05$ (Adjustment: Holm).

Firstly, participants reported that the final design matched their expectations ("The final design matches my expectation."; *M*=5.70, *SD*=1.58). This suggests that most users felt the outcome aligned with their anticipated expectations. Similarly, satisfaction ("I'm pleased with the final design.") with the final design was high (*M*=6.00, *SD*=1.41).

Control over the design process ("I felt in control of the design process.") was another positive aspect (*M*=5.81, *SD*=1.35). Confidence in the optimality of the design ("I believe the design is optimal for me.") for their needs was also high (*M*=5.97, *SD*=1.30).

However, ownership of the final design ("I feel the final design is mine."), while still positive, showed a slightly lower mean of *M*=5.30 (*SD*=1.61).

Overall, the data suggests that participants were generally pleased with the design and felt a strong sense of control and confidence in the process, but the degree of ownership felt towards the final design may warrant further exploration to ensure greater personal investment in the outcomes.

5.7 Open Feedback

Participants provided detailed feedback regarding their experience with the design and interaction during the study. Several key themes emerged, including interactivity, general impressions, and areas for improvement.

Interactivity and Control: While many participants felt the design was effective, some expressed uncertainty about their level of control over the process. One participant mentioned, *"I did not realise I was controlling the design process,"* highlighting a disconnect between the design and user awareness of control. Another echoed a similar sentiment, *"I would like more control over how the HMI looked in general. The range of adjustments was too small and easy to miss sometimes."*

General Impressions: Overall, feedback was positive, with several participants praising the design. One stated, *"Overall, the study and the design of the eHMI was effective. It was well paced, and organised."* Another appreciated the clarity of the vehicle's communication, saying, *"The announcement of stop/go was useful alongside the concept of colors. It made it easier for pedestrians to understand when to and not to cross the road."* The audio and visual cues were particularly well-received, with a participant noting, *"The sound 'stopping' is really a good reminder for me to cross the road, let me feel safe."*

Areas for Improvement: Despite positive feedback, several suggestions for improvement were made. One participant mentioned the unpredictability of the vehicle's stopping behavior: *"Not sure when the automated car decides it wants to stop. The eHMI seems to come out suddenly."* Another recommended more gradual cues, suggesting, *"A gradual light/indicator that the car is going to slow down for me—recognizes and conveys that intent."* Some participants also noted that the interaction could benefit from additional elements, such as *"more interaction elements in the scenario, or building more than one."*

A few participants commented on the unpredictability of the vehicles, stating, *"In the beginning, it was not clear for me where exactly I have to stand and wait for the vehicles to stop."* Others felt that different colors or light effects would enhance communication, with one suggesting, *"The lights were adequate. However, a bright light in the day and night would be better, like a bright green or red color."*

6 Discussion

While numerous works evaluated different designs of eHMIs [22, 29, 75], there is the open question of how an optimal eHMI should be designed.

We conducted a between-subject study to investigate these RQs with 37 participants. We employed MOBO to iteratively optimize the eHMI designs based on participant ratings of trust, perceived safety, understanding, mental demand, acceptance, aesthetics, and time to start crossing. This method enabled us to compare designs on the Pareto, where enhancing one objective would necessitate

trade-offs with others [61]. In the following, we discuss our findings regarding the optimal eHMI and the design process.

6.1 Optimized Design Choices in External Communication of Automated Vehicles

The comparison of optimized eHMI design parameters between females and males revealed **no** evidence for differences. Regarding RQ2, however, this does not necessarily imply a generally coherent Pareto front for all users. Gender is only one demographic factor, and its absence of effect in our study does not necessarily mean that a single "universal" eHMI design would work equally well for all. User characteristics are complex, influenced by factors such as age, cultural background [19], cognitive abilities, and personal experiences (e.g., see Edelmann et al. [36]).

Yet, looking more qualitatively at the final design parameters (see Figure 7), we conclude that the Pareto front incorporates a similar design. Therefore, we posit that a certain set of parameters—e.g., cyan hue, flashing animation around 3Hz, and a large illuminated area on the vehicle's front—might be appropriate for most users, at least as an initial starting design from which personalization is more efficient. We argue in line with previous work [17, 29] that cyan is an appropriate color. We modeled the preferred flashing animation [29] and found that ≈3Hz was optimal, compared to 1Hz used by Dey et al. [29]. Compared to previous designs that used LED strips [17, 29, 76], we enabled the entire AV front to be an eHMI (also see Schlackl et al. [71] or Dey et al. [30]).

Interestingly, Figure 7 shows that on the Pareto front, the volume of the auditory message was relatively high (approximately 0.8, with no value on the Pareto front below 0.5) both for female and male participants. Moderate evidence for similarity was also found here for female and male participants. This is important for accessibility considerations [24] and also replicates findings that support the positive effect of multimodality (e.g., visual and auditory signals) of eHMIs [32]. This furthermore highlights the possibility of having a "one-fits-all" eHMI [57] for some design parameters.

This work also advocates **for** an eHMI as with (at least) 20 rounds of interaction, an eHMI was still rated highly. While perceived safety most likely also increased due to increased exposure time [15], this is still a finding supporting explicit eHMIs and is in contrast to work considering explicit eHMIs not necessary [63].

6.2 Individual Optimization and the Notion of Universality

General approaches exist to focus on *group level* optimization. For example, one can use the causal tree analysis by Athey and Imbens [5] and then the derived groups as a homogenous group for optimization. However, it is unclear which variables should be used for the grouping in the context of eHMIs. We have, therefore, used gender as a known characteristic affecting the crossing decision [25]. Interestingly, we found **no** significant differences in the objectives or the design parameters. However, this does not mean that genuine individual differences are absent; rather, it indicates that our chosen demographic factor and current sample might not have captured meaningful variability.

6.3 Trade-off in the Objectives

MOBO seeks to identify optimal design parameters that balance multiple objectives along the Pareto front, ensuring that enhancing one objective does not disproportionately compromise another [61]. This approach is particularly valuable when trade-offs exist between objectives, such as spatial error versus completion time [10]. However, our analysis uncovered strong correlations among objectives, suggesting minimal conflict between some.

This observation implies potential redundancy within the current set of objectives. Although these were selected based on previous eHMI studies [24, 26, 29, 30], future research could streamline the objectives by prioritizing those that are most distinct and impactful. For example, trust and predictability or acceptance and aesthetics show strong correlations (see Figure 10), indicating that selecting one from each pair could simplify future studies. Conversely, mental demand, which exhibits only a weak correlation with Perceived safety when looking at Pareto optimal values (see Figure 10b), should be preserved as it captures a unique user experience.

Although these are preliminary insights, future research should further investigate and validate the potential for reducing objectives in eHMI studies.

6.4 Practical Guidelines for eHMI Designs

Our study's findings provide valuable practical insights for future research on eHMIs. Here are the two key guidelines:

- **Multimodality is Key:** When designing eHMIs, accessibility of traffic is improved when employing multimodality [24] and also wanted by people without disabilities.
- **Starting Points for Efficient Personalization:** While personalization and continuous optimization remain necessary to account for individual differences (apart from gender), our findings suggest certain parameter ranges as effective starting points. These could enable the optimization to focus on fewer other parameters to yield eHMI designs that quickly align with diverse user needs.

6.5 Limitations and Future Work

We prioritized subjective ratings as main objectives for optimizing the UI design parameters. However, we included one objective value: Time to start crossing. It remains to be discussed whether more objective data, such as objective physiological responses, is useful. Nonetheless, the subjective ratings were collected using validated questionnaires or inspired by related research. While valid, future research should investigate the relationship between these dimensions and more specific subcomponents of trust. Additionally, it is crucial to consider the potential redundancy of some questionnaire items due to high correlations between certain measures (see Section 6.3).

Our analysis provided inconclusive evidence on gender differences, limiting the strength of our claims.

Future work should expand the age range and incorporate more diverse participant pools. Although we included German and Indian participants, future investigations should broaden cultural contexts [36]. Future work would also benefit from testing the

eHMI designs under various environmental conditions, such as daytime versus nighttime or weather scenarios (see Colley and Rukzio [22]), to better assess their effectiveness in real-world applications. Additionally, the scope of the eHMI designs could be expanded by incorporating other design elements, such as text-based messages, symbols, or dynamic animations. Prior work has, however, shown a high potential for LED stripes in eHMI designs due to their easy integration and aesthetically pleasing design. Text, for example, is inaccessible to children or foreigners, and symbols can be unclear. We used the flashing animation as prior work showed a preference for this pattern [29].

7 Conclusion

This study explored the eHMI design preferences using the MOBO approach. We conducted a study with N=37 participants in VR. Our findings indicate no differences between female and male participants. While this does not yet claim that a universal design for all users exists, assessing the design parameters on the Pareto front, found that some design parameter value ranges seem feasible for future work to focus on. Therefore, this work supports the standardization efforts necessary in the automotive domain.

Open Science

We make the Bayesian optimizer (see https://github.com/Pascal-Jansen/Bayesian-Optimization-for-Unity), the Unity application upon request, and the collected (anonymized) data available (see https://github.com/M-Colley/ehmi-optimization-chi25-data).

Acknowledgments

References

[1] ISO/TR 23049:2018. 2018. *Road Vehicles: Ergonomic Aspects of External Visual Communication from Automated Vehicles to Other Road Users.* Standard. International Organization for Standardization.

[2] Claudia Ackermann, Matthias Beggiato, Sarah Schubert, and Josef F Krems. 2019. An experimental study to investigate design and assessment criteria: what is important for communication between pedestrians and automated vehicles? *Applied ergonomics* 75 (2019), 272–282.

[3] Sander Ackermans, Debargha Dey, Peter Ruijten, Raymond H. Cuijpers, and Bastian Pfleging. 2020. The Effects of Explicit Intention Communication, Conspicuous Sensors, and Pedestrian Attitude in Interactions with Automated Vehicles. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). ACM, New York, NY, USA, 1–14. doi:10.1145/3313831.3376197

[4] Ashratuz Zavin Asha, Christopher Smith, Georgina Freeman, Sean Crump, Sowmya Somanath, Lora Oehlberg, and Ehud Sharlin. 2021. Co-Designing Interactions between Pedestrians in Wheelchairs and Autonomous Vehicles. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) (DIS '21). ACM, New York, NY, USA, 339–351. doi:10.1145/3461778.3462068

[5] Susan Athey and Guido Imbens. 2016. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences* 113, 27 (2016), 7353–7360.

[6] 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

[7] Ali Borji and Laurent Itti. 2013. Bayesian optimization explains human active search. *Advances in neural information processing systems* 26 (2013).

[8] 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.

[9] Eric Brochu, Vlad M. Cora, and Nando de Freitas. 2010. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User

[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). ACM, 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). ACM, New York, NY, USA, 34–45. doi:10.1145/3565472.3592956

[12] Chia-Ming Chang, Koki Toda, Takeo Igarashi, Masahiro Miyata, and Yasuhiro Kobayashi. 2018. A Video-Based Study Comparing Communication Modalities between an Autonomous Car and a Pedestrian. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON, Canada) (AutomotiveUI '18). ACM, New York, NY, USA, 104–109. doi:10.1145/3239092.3265950

[13] Vicky Charisi, Azra Habibovic, Jonas Andersson, Jamy Li, and Vanessa Evers. 2017. Children's Views on Identification and Intention Communication of Self-Driving Vehicles. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) (IDC '17). ACM, New York, NY, USA, 399–404. doi:10.1145/3078072.3084300

[14] Chia-Hsing Chiu, Yuki Koyama, Yu-Chi Lai, Takeo Igarashi, and Yonghao Yue. 2020. Human-in-the-Loop Differential Subspace Search in High-Dimensional Latent Space. *ACM Trans. Graph.* 39, 4, Article 85 (aug 2020), 15 pages. doi:10.1145/3386569.3392409

[15] Mark Colley, Elvedin Bajrovic, and Enrico Rukzio. 2022. Effects of Pedestrian Behavior, Time Pressure, and Repeated Exposure on Crossing Decisions in Front of Automated Vehicles Equipped with External Communication. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). ACM, New York, NY, USA, Article 367, 11 pages. doi:10.1145/3491102.3517571

[16] Mark Colley, Jan Henry Belz, and Enrico Rukzio. 2021. Investigating the Effects of Feedback Communication of Autonomous Vehicles. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* ACM, New York, NY, USA, 263–273. https://doi.org/10.1145/3409118.3475133

[17] Mark Colley, Julian Britten, and Enrico Rukzio. 2023. Scalability in External Communication of Automated Vehicles: Evaluation and Recommendations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–26.

[18] Mark Colley, Julian Czymmeck, Mustafa Kücükkocak, Pascal Jansen, and Enrico Rukzio. 2024. PedSUMO: Simulacra of Automated Vehicle-Pedestrian Interaction Using SUMO To Study Large-Scale Effects. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) (HRI '24). ACM, New York, NY, USA, 890–895. doi:10.1145/3610977.3637478

[19] Mark Colley, Daniel Kornmüller, Debargha Dey, Wendy Ju, and Enrico Rukzio. 2024. Longitudinal Effects of External Communication of Automated Vehicles in the USA and Germany: A Comparative Study in Virtual Reality and Via a Browser. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 176 (Nov. 2024), 33 pages. doi:10.1145/3699778

[20] Mark Colley, Surong Li, and Enrico Rukzio. 2021. Increasing Pedestrian Safety Using External Communication of Autonomous Vehicles for Signalling Hazards. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction.* ACM, New York, NY, USA, Article 20, 10 pages. https://doi.org/10.1145/3447526.3472024

[21] 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.* 7, 2, Article 52 (jun 2023), 23 pages. doi:10.1145/3596249

[22] Mark Colley and Enrico Rukzio. 2020. A Design Space for External Communication of Autonomous Vehicles. In *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20).* ACM, New York, NY, USA, doi:10.1145/3409120.3410646

[23] Mark Colley and Enrico Rukzio. 2020. Towards a Design Space for External Communication of Autonomous Vehicles. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, Hawaii USA) (CHI '20). ACM, New York, NY, USA. doi:10.1145/3334480.3382844

[24] Mark Colley, Marcel Walch, Jan Gugenheimer, Ali Askari, and Enrico Rukzio. 2020. Towards Inclusive External Communication of Autonomous Vehicles for Pedestrians with Vision Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). ACM, New York, NY, USA, 1–14. doi:10.1145/3313831.3376472

[25] Mark Colley, Marcel Walch, and Enrico Rukzio. 2019. For a Better (Simulated) World: Considerations for VR in External Communication Research. In *Proceedings of the 11th International Conference on Automotive User Interfaces and*

Modeling and Hierarchical Reinforcement Learning. http://arxiv.org/abs/1012.2599 arXiv:1012.2599 [cs].

Interactive Vehicular Applications: Adjunct Proceedings (Utrecht, Netherlands) (*AutomotiveUI '19*). ACM, New York, NY, USA, 442–449. doi:10.1145/3349263.3351523

[26] Mark Colley, Marcel Walch, and Enrico Rukzio. 2020. Unveiling the Lack of Scalability in Research on External Communication of Autonomous Vehicles. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, Hawaii USA) (*CHI '20*). ACM, New York, NY, USA. doi:10.1145/3334480.3382865

[27] Shuchisnigdha Deb, Daniel W. Carruth, Muztaba Fuad, Laura M. Stanley, and Darren Frey. 2020. Comparison of Child and Adult Pedestrian Perspectives of External Features on Autonomous Vehicles Using Virtual Reality Experiment. In *Advances in Human Factors of Transportation*, Neville Stanton (Ed.). Springer International Publishing, Cham, 145–156.

[28] Debargha Dey, Azra Habibovic, Andreas Löcken, Philipp Wintersberger, Bastian Pfleging, Andreas Riener, Marieke Martens, and Jacques Terken. 2020. Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. *Transportation Research Interdisciplinary Perspectives* 7 (2020), 100174.

[29] Debargha Dey, Azra Habibovic, Bastian Pfleging, Marieke Martens, and Jacques Terken. 2020. Color and Animation Preferences for a Light Band eHMI in Interactions between Automated Vehicles and Pedestrians. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13.

[30] Debargha Dey, Kai Holländer, Melanie Berger, Berry Eggen, Marieke Martens, Bastian Pfleging, and Jacques Terken. 2020. Distance-Dependent EHMIs for the Interaction Between Automated Vehicles and Pedestrians. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) (*AutomotiveUI '20*). ACM, New York, NY, USA, 192–204. doi:10.1145/3409120.3410642

[31] Debargha Dey, Marieke Martens, Chao Wang, Felix Ros, and Jacques Terken. 2018. Interface Concepts for Intent Communication from Autonomous Vehicles to Vulnerable Road Users. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON, Canada) (*AutomotiveUI '18*). ACM, New York, NY, USA, 82–86. doi:10.1145/3239092.3265946

[32] Debargha Dey, Toros Ufuk Senan, Bart Hengeveld, Mark Colley, Azra Habibovic, and Wendy Ju. 2024. Multi-Modal eHMIs: The Relative Impact of Light and Sound in AV-Pedestrian Interaction. In *CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–16. doi:10.1145/3613904.3642031

[33] Debargha Dey, Francesco Walker, Marieke Martens, and Jacques Terken. 2019. Gaze Patterns in Pedestrian Interaction with Vehicles: Towards Effective Design of External Human-Machine Interfaces for Automated Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) (*AutomotiveUI '19*). ACM, New York, NY, USA, 369–378. doi:10.1145/3342197.3344523

[34] André Dietrich, Michael Tondera, and Klaus Bengler. 2020. Automated vehicles in urban traffic: The effects of kinematics and eHMI on pedestrian crossing behaviour. *Advances in transportation studies* 2020 (2020), 73–84.

[35] 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*). ACM, New York, NY, USA, 1–12. doi:10.1145/3290605.3300482

[36] Aaron Edelmann, Stefan Stümper, and Tibor Petzoldt. 2021. Cross-cultural differences in the acceptance of decisions of automated vehicles. *Applied ergonomics* 92 (2021), 103346.

[37] 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*). ACM, New York, NY, USA, 1–14. doi:10.1145/3313831.3376484

[38] Daniel J Fagnant and Kara Kockelman. 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice* 77 (2015), 167–181.

[39] Evelyn Florentine, Mark Adam Ang, Scott Drew Pendleton, Hans Andersen, and Marcelo H. Ang. 2016. Pedestrian Notification Methods in Autonomous Vehicles for Multi-Class Mobility-on-Demand Service. In *Proceedings of the Fourth International Conference on Human Agent Interaction* (Biopolis, Singapore) (*HAI '16*). ACM, New York, NY, USA, 387–392. doi:10.1145/2974804.2974833

[40] Xinyue Gui, Mikiya Kusunoki, Bofei Huang, Stela Hanbyeol Seo, Chia-Ming Chang, Haoran Xie, Manabu Tsukada, and Takeo Igarashi. 2024. Shrinkable Arm-based eHMI on Autonomous Delivery Vehicle for Effective Communication with Other Road Users. In *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, NY, New York, USA, 305–316.

[41] Xinyue Gui, Koki Toda, Stela Hanbyeol Seo, Felix Martin Eckert, Chia-Ming Chang, Xiang'Anthony Chen, and Takeo Igarashi. 2023. A field study on pedestrians' thoughts toward a car with gazing eyes. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, NY, New York, USA, 1–7.

[42] Mathias Haimerl, Mark Colley, and Andreas Riener. 2022. Evaluation of Common External Communication Concepts of Automated Vehicles for People With Intellectual Disabilities. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 182 (Sept. 2022), 19 pages. doi:10.1145/3546717

[43] 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.

[44] Kai Holländer, Mark Colley, Enrico Rukzio, and Andreas Butz. 2021. A Taxonomy of Vulnerable Road Users for HCI Based On A Systematic Literature Review. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 158, 13 pages. https://doi.org/10.1145/3411764.3445480

[45] Kai Holländer, Andy Krüger, and Andreas Butz. 2020. Save the Smombies: App-Assisted Street Crossing. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) (*MobileHCI '20*). ACM, New York, NY, USA, Article 22, 11 pages. doi:10.1145/3379503.3403547

[46] Kai Holländer, Philipp Wintersberger, and Andreas Butz. 2019. Overtrust in External Cues of Automated Vehicles: An Experimental Investigation. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) (*AutomotiveUI '19*). ACM, New York, NY, USA, 211–221. doi:10.1145/3342197.3344528

[47] Ming Hou, Karthik Mahadevan, Sowmya Somanath, Ehud Sharlin, and Lora Oehlberg. 2020. Autonomous Vehicle-Cyclist Interaction: Peril and Promise. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). ACM, New York, NY, USA, 1–12. doi:10.1145/3313831.3376884

[48] 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*). ACM, New York, NY, USA, Article 585, 11 pages. doi:10.1145/3411764.3445140

[49] 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.

[50] 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*). ACM, New York, NY, USA, Article 77, 14 pages. doi:10.1145/3526113.3545664

[51] Yuki Koyama, Issei Sato, and Masataka Goto. 2020. Sequential Gallery for Interactive Visual Design Optimization. *ACM Trans. Graph.* 39, 4, Article 88 (aug 2020), 12 pages. doi:10.1145/3386569.3392444

[52] Mirjam Lanzer, Franziska Babel, Fei Yan, Bihan Zhang, Fang You, Jianmin Wang, and Martin Baumann. 2020. Designing Communication Strategies of Autonomous Vehicles with Pedestrians: An Intercultural Study. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) (*AutomotiveUI '20*). ACM, New York, NY, USA, 122–131. doi:10.1145/3409120.3410653

[53] M.D. Lee and E.J. Wagenmakers. 2013. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press. https://books.google.de/books?id=50tkAgAAQBAJ

[54] Yee Mun Lee, Ruth Madigan, Oscar Giles, Laura Garach-Morcillo, Gustav Markkula, Charles Fox, Fanta Camara, Markus Rothmueller, Signe Alexandra Vendelbo-Larsen, Pernille Holm Rasmussen, et al. 2021. Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. *Cognition, Technology & Work* 23 (2021), 367–380.

[55] Yi-Chi Liao. 2023. *Human-in-the-Loop Design Optimization*. Ph. D. Dissertation. Aalto University. https://urn.fi/URN:ISBN:978-952-64-1580-2

[56] Yi-Chi Liao, John J Dudley, George B Mo, Chun-Lien Cheng, Liwei Chan, Antti Oulasvirta, and Per Ola Kristensson. 2023. Interaction Design With Multiobjective Bayesian Optimization. *IEEE Pervasive Computing* 22, 1 (2023), 29–38.

[57] Andreas Löcken, Mark Colley, Andrii Matviienko, Kai Holländer, Debargha Dey, Azra Habibovic, Andrew L Kun, Susanne Boll, and Andreas Riener. 2020. We-CARe: Workshop on Inclusive Communication between Automated Vehicles and Vulnerable Road Users. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, USA, 1–5.

[58] Andreas Löcken, Carmen Golling, and Andreas Riener. 2019. How Should Automated Vehicles Interact with Pedestrians? A Comparative Analysis of Interaction Concepts in Virtual Reality. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) (*AutomotiveUI '19*). ACM, New York, NY, USA, 262–274. doi:10.1145/3342197.3344544

[59] Victor Malmsten Lundgren, Azra Habibovic, Jonas Andersson, Tobias Lagström, Maria Nilsson, Anna Sirkka, Johan Fagerlönn, Rikard Fredriksson, Claes Edgren, Stas Krupenia, and Dennis Saluäär. 2017. Will There Be New Communication Needs When Introducing Automated Vehicles to the Urban Context?. In *Advances*

in Human Aspects of Transportation, Neville A. Stanton, Steven Landry, Giuseppe Di Bucchianico, and Andrea Vallicelli (Eds.). Springer International Publishing, Cham, 485–497.

[60] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, 1–12. doi:10.1145/3173574.3174003

[61] 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

[62] Olaf Mersmann. 2024. *emoa: Evolutionary Multiobjective Optimization Algorithms.* doi:10.32614/CRAN.package.emoa R package version 0.5-3.

[63] Dylan Moore, Rebecca Currano, G. Ella Strack, and David Sirkin. 2019. The Case for Implicit External Human-Machine Interfaces for Autonomous Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. ACM, New York, NY, USA, 295–307. doi:10.1145/3342197.3345320

[64] Trung Thanh Nguyen, Kai Holländer, Marius Hoggenmueller, Callum Parker, and Martin Tomitsch. 2019. Designing for Projection-Based Communication between Autonomous Vehicles and Pedestrians. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. ACM, New York, NY, USA, 284–294. doi:10.1145/3342197.3344543

[65] 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

[66] Eryn O'Dowd and Thomas V Pollet. 2018. Gender differences in use of a pedestrian crossing: an observational study in newcastle upon tyne. *Letters on Evolutionary Behavioral Science* 9, 1 (2018), 1–4.

[67] Max Oudshoorn, Joost de Winter, Pavlo Bazilinskyy, and Dimitra Dodou. 2021. Bio-inspired intent communication for automated vehicles. *Transportation research part F: traffic psychology and behaviour* 80 (2021), 127–140.

[68] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles* 3, 1 (2017), 61–70.

[69] Shadan Sadeghian, Marc Hassenzahl, and Kai Eckoldt. 2020. An Exploration of Prosocial Aspects of Communication Cues between Automated Vehicles and Pedestrians. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. ACM, New York, NY, USA, 205–211. doi:10.1145/3409120.3410657

[70] Hatice Sahin, Heiko Mueller, Shadan Sadeghian, Debargha Dey, Andreas Löcken, Andrii Matviienko, Mark Colley, Azra Habibovic, and Philipp Wintersberger. 2021. Workshop on Prosocial Behavior in Future Mixed Traffic. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* ACM, New York, NY, USA, 167–170. https://doi.org/10.1145/3473682.3477438

[71] Dominik Schlackl, Klemens Weigl, and Andreas Riener. 2020. eHMI visualization on the entire car body: results of a comparative evaluation of concepts for the communication between AVs and manual drivers. In *Proceedings of Mensch und Computer 2020.* ACM, New York, NY, USA, 79–83.

[72] Andreas Sieß, Kathleen Hübel, Daniel Hepperle, Andreas Dronov, Christian Hufnagel, Julia Aktun, and Matthias Wölfel. 2015. Hybrid City Lighting-Improving Pedestrians' Safety through Proactive Street Lighting. In *2015 International Conference on Cyberworlds (CW)*. IEEE, New York, NY, USA, 46–49.

[73] 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.

[74] Srinath Sridhar, Anna Maria Feit, Christian Theobalt, and Antti Oulasvirta. 2015. Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3643–3652. doi:10.1145/2702123.2702136

[75] Tram Thi Minh Tran, Callum Parker, Marius Hoggenmüller, Yiyuan Wang, and Martin Tomitsch. 2024. Exploring the Impact of Interconnected External Interfaces in Autonomous Vehicles on Pedestrian Safety and Experience. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 1–17.

[76] Tram Thi Minh Tran, Callum Parker, Xinyan Yu, Debargha Dey, Marieke Martens, Pavlo Bazilinskyy, and Martin Tomitsch. 2024. Evaluating Autonomous Vehicle External Communication Using a Multi-Pedestrian VR Simulator. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 130 (sep 2024), 26 pages. doi:10.1145/3678506

[77] 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.

[78] 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

[79] Mingyuan Zhong, Gang Li, and Yang Li. 2021. Spacewalker: Rapid UI Design Exploration Using Lightweight Markup Enhancement and Crowd Genetic Programming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. ACM, New York, NY, USA, Article 315, 11 pages. doi:10.1145/3411764.3445326

[80] Raphael Zimmermann and Reto Wettach. 2017. First Step into Visceral Interaction with Autonomous Vehicles. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Oldenburg, Germany) *(AutomotiveUI '17)*. ACM, New York, NY, USA, 58–64. doi:10.1145/3122986.3122988

A Procedure — Introduction

You will be crossing a road in a virtual reality (VR) environment. You will encounter automated and manually driven vehicles. Automated vehicles are identified by a strip on the upper part of the windshield. If you feel uncomfortable at any time, you can stop the study without giving any reasons and without any disadvantages. In this case, simply remove the VR goggles and inform the study leader.

You will be shown different scenarios, one after the other. The aim of all scenarios is to reach the checkpoint (blue) on the opposite side of the road. If an automated vehicle recognizes that you want to cross the road, it will communicate with you, depending on the passage. This communication will change over the course of the study.

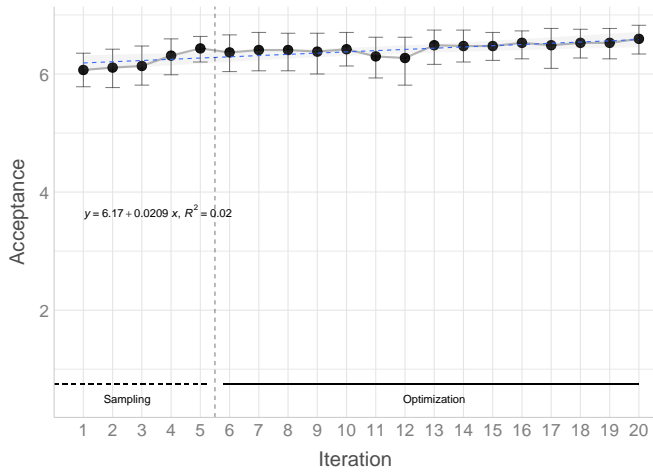## B Results

B.1 Objective Values over Iterations
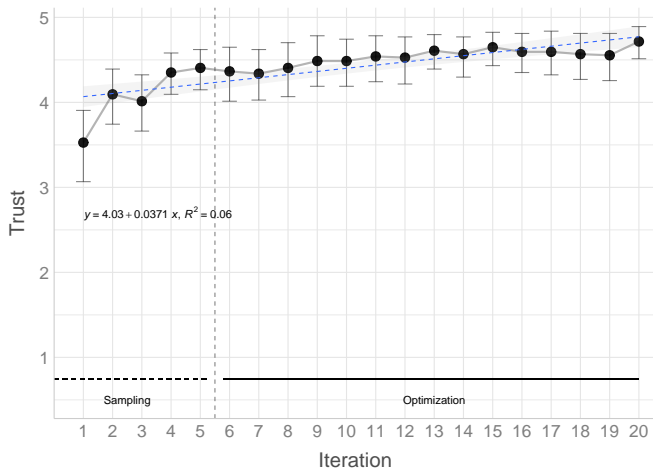
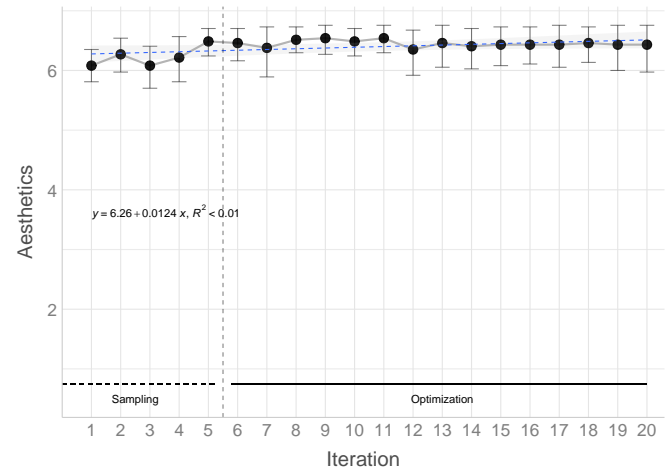(a) Progression of trust values over MOBO iterations.
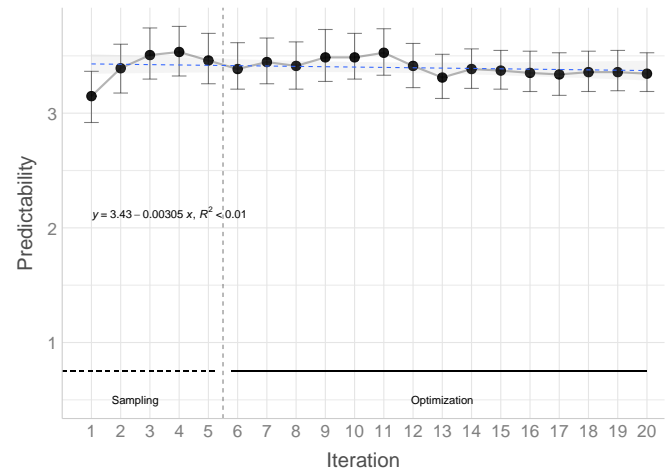
(b) Progression of predictability values over MOBO iterations.

Figure 11: Value progression of trust and predictability.
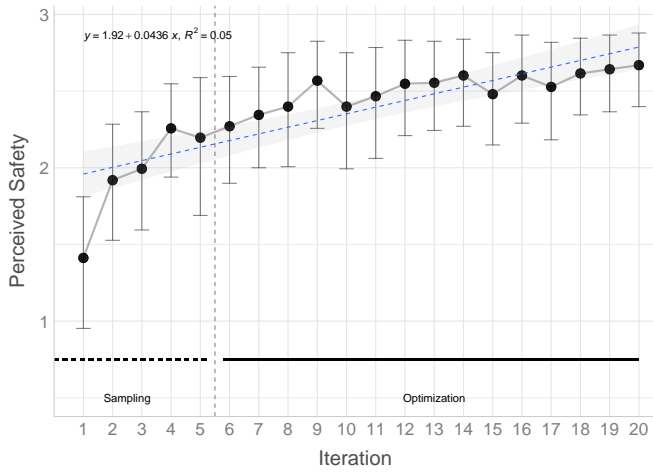

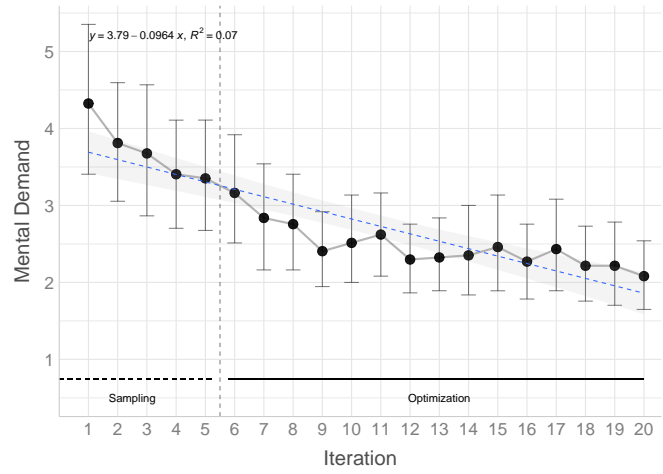(a) Progression of acceptance over MOBO iterations.

(b) Progression of aesthetics over MOBO iterations.

Figure 12: Value progression of acceptance and aesthetics.

(a) Progression of perceived safety over MOBO iterations.

(b) Progression of mental demand over MOBO iterations.

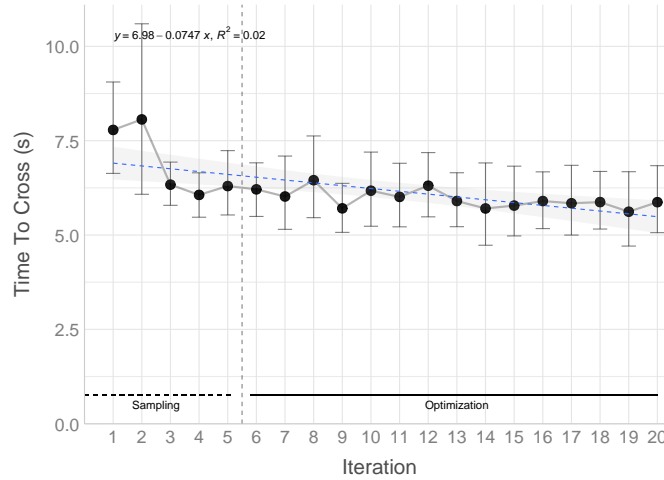**Figure 13: Value progression of perceived safety and mental demand.**

Figure 14: Progression of Time to start crossing over MOBO iterations.