

Comparing Preferences Between Japan and Germany for External Communication of Automated Vehicles Using Bayesian Optimization

MARK COLLEY*, UCL Interaction Centre, United Kingdom

PASCAL JANSEN*, Institute of Media Informatics, Ulm University, Germany and UCL Interaction Centre, United Kingdom

XINYUE GUI, The University of Tokyo, Japan

YUAN LI, The University of Tokyo, Japan

DING XIA, The University of Tokyo, Japan

ENRICO RUKZIO, Institute of Media Informatics, Ulm University, Germany

TAKEO IGARASHI, The University of Tokyo, Japan

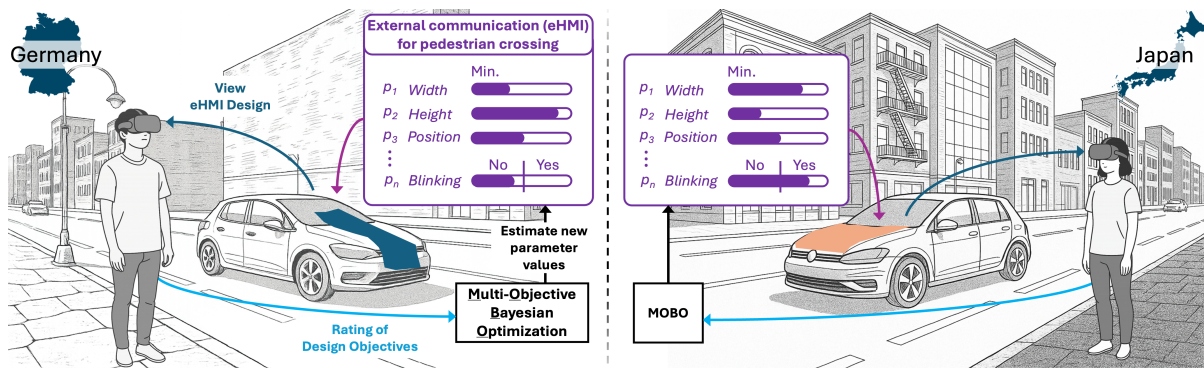


Fig. 1. Concept of comparing the preferences in eHMI design between **Germany** and **Japan** using Human-in-the-loop Multi-Objective Bayesian Optimization (MOBO). The automated vehicle's eHMI communicates its intention and allows pedestrians to cross the street (with left-hand traffic in the Japanese variant). The goal is to improve user ratings on multiple design objectives, such as trust, safety, acceptance, and aesthetics, while minimizing mental demand. MOBO iteratively adjusts design parameter values (p_1 to p_n) to suggest optimized designs that meet individual objectives.

The absence of human users in automated vehicles (AVs) could require external Human-Machine Interfaces (eHMIs) to allow for communication with other vulnerable road users in uncertain scenarios. This could be, for example, regarding the right of way. Given the plethora of adjustable parameters, balancing visual and auditory elements is crucial for effective

*Both authors contributed equally to this research.

Authors' Contact Information: [Mark Colley](mailto:m.colley@ucl.ac.uk), m.colley@ucl.ac.uk, UCL Interaction Centre, London, United Kingdom; [Pascal Jansen](mailto:pascal.jansen@uni-ulm.de), pascal.jansen@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany and UCL Interaction Centre, London, United Kingdom; [Xinyue Gui](mailto:xinyuegui@gmail.com), xinyuegui@gmail.com, The University of Tokyo, Tokyo, Japan; [Yuan Li](mailto:liyuan@keio.jp), liyuan@keio.jp, The University of Tokyo, Tokyo, Japan; [Ding Xia](mailto:dingxia1995@gmail.com), dingxia1995@gmail.com, The University of Tokyo, Tokyo, Japan; [Enrico Rukzio](mailto:enrico.rukzio@uni-ulm.de), enrico.rukzio@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; [Takeo Igarashi](mailto:takeo@acm.org), The University of Tokyo, Tokyo, Japan, takeo@acm.org.



This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2026 Copyright held by the owner/author(s).

ACM 2474-9567/2026/9-ART94

<https://doi.org/10.1145/3831988>

communication with other road users. With $N=40$ ($n=20$ in Germany and $n=20$ in Japan) participants, this study employed multi-objective Bayesian optimization to evaluate optimized eHMI designs between Japan and Germany. By comparing the Pareto front, we identify optimal design trade-offs and their differences. We also evaluate how the process is perceived between regions.

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

Additional Key Words and Phrases: External communication; Autonomous vehicles; Pedestrian Behavior; eHMI.

ACM Reference Format:

Mark Colley, Pascal Jansen, Xinyue Gui, Yuan Li, Ding Xia, Enrico Rukzio, and Takeo Igarashi. 2026. Comparing Preferences Between Japan and Germany for External Communication of Automated Vehicles Using Bayesian Optimization. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 3, Article 94 (September 2026), 29 pages. <https://doi.org/10.1145/3831988>

1 Introduction

Automated vehicles (AVs) will change traffic operations and interactions in traffic settings [27, 34, 45, 70]. Without a human driver, explicit communication is required in ambiguous situations, such as negotiating priority. External human machine interfaces (eHMIs) address this need. Proposed modalities include grille and windshield displays [37, 46], LED strips [46, 71], vehicle motion cues [98], and projections [2, 78], with extensions via smartphones [53] and infrastructure [88]. Designs span text-based messages [15] that communicate vehicle state [44] or intent [42]. Studies consistently report gains in trust, clarity, hedonic evaluation, and pedestrian crossing behavior.

Recent work applies Human-in-the-Loop Multi-Objective Bayesian Optimization, HITL MOBO, to iteratively identify optimal eHMI design parameters [23] (see Figure 1). MOBO predicts which design changes (e.g., a different color) best meet individual users' design objectives, such as increasing trust. It addresses multiple objectives by seeking *Pareto-optimal* design trade-offs, that are points on the Pareto front where no objective can improve without another worsening [74]. Compared to methods that often need large labeled datasets (e.g., supervised learning), many interactions per user (e.g., reinforcement learning), or exhaustive A/B tests, HITL MOBO learns a personalized model during interaction and chooses the next design expected to provide the most improvement towards objectives, so it reaches trade-offs in a few iterations, important when user feedback is scarce or tiring.

In interactions between vulnerable road users and AV, cultural background can shape behavior, and responses may differ across study regions that appear similar [59]. Studies report differences between Asian contexts, for example Japan and China, and Western contexts, for example Germany and the USA [50, 59, 77]. Such region- and individual-level differences in eHMI perception may lead to misinterpretation, which, at non-signalized crossings, elevates traffic risk.

However, two gaps remain. First, while prior work has shown that HITL MOBO can optimize eHMI parameters for individual users, it remains unclear whether the resulting multi-objective trade-offs are stable across study regions. Second, even when cross-regional differences in eHMI perception are reported, little is known about which design parameters account for these differences within a multi-objective design space. This limits both understanding of region-sensitive eHMI preferences and practical guidance for designing eHMIs that remain effective beyond a single study context. The overarching research gap is a lack of systematic, data-driven comparisons of multi-objective eHMI design trade-offs across culturally distinct regions.

To address this without overgeneralization, we treat the country and study region as the unit of comparison. We therefore make claims about people living in Germany and Japan in this specific scenario and sample, not about Europe or Asia as aggregates. Our primary contribution is a proof-of-concept demonstrating that HITL MOBO can be used to quantify and compare multi-objective eHMI design trade-offs across regions.

We ground this comparison in cultural constructs that are plausibly relevant for pedestrian-vehicle coordination under uncertainty. Prior work links Germany–Japan differences in risk avoidance and compliance to uncertainty avoidance and collectivism versus individualism [50, 51, 77]. We use these constructs as expectations for interpretation rather than as causal explanations. Specifically, higher uncertainty avoidance should increase preference for signals that reduce ambiguity through salience and redundancy. In our parameterization, this maps to higher alpha, larger eHMI area (width and height), and higher auditory loudness. In contrast, preferences that prioritize minimizing distraction and visual noise should map to lower blink frequency and lower auditory loudness, while keeping an unambiguous intent signal. This study is therefore guided by the following research question (RQ):

RQ1: How do Pareto-optimal eHMI designs differ between participants in Germany and Japan when optimizing trust, mental demand, visual appeal, and time to start crossing?

To explore this RQ, we conducted a user study with $N=40$ (20 German, 20 Japanese) participants. We applied MOBO to optimize the eHMI's design (see Figure 1), for the parameters color, blink frequency, size, positioning, and loudness of the auditory component. Beyond that, we were interested in regional differences in perceptions of the MOBO process. Therefore, the second RQ was:

RQ2: How does the perceived experience of the HITL MOBO process (expectation conformity, satisfaction, agency, confidence, ownership) differ between participants in Germany and Japan?

In the virtual reality (VR) study, $N=40$ participants crossed a two-lane road with traffic in both directions (on the right for Germany, left for Japan). 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.

Contribution Statement [95]:

- **Artifact or System.** Using Unity VR, we simulated a pedestrian crossing an unsignalized street and applied MOBO to iteratively optimize the eHMI, identifying optimal design parameters from user feedback on four design objectives.
- **Empirical study that tells us about how people use a system.** We conducted a between-subjects study with 40 participants, 20 in Germany and 20 in Japan, to assess regional effects on user experience and eHMI interface design.

2 Related Work

We organize related work around (1) external communication for automated vehicles, (2) HITL MOBO for interface design, and (3) prior cross-region and cultural evidence in road-crossing behavior that motivates our Japan–Germany region comparison.

2.1 External Communication of Automated Vehicles

Traffic interactions often rely on implicit cues such as gestures and eye contact [82]. eHMIs were proposed to support AV communication with pedestrians and cyclists [52]. External communication has been categorized by modality, message type, and locus [27, 29, 34]. Colley and Rukzio [27] defines eight message types: Instruction, Command, Advisory, Answer, Historical, Predictive, Question, and Affective. Key loci include the windshield and bumper areas [41]. Most work and ISO TR [1] recommend conveying AV intention rather than issuing prescriptive commands; hence, our optimization targets intention communication.

Effective deployment depends on the communication relationship, ambient noise, and road user group [27]. Studies report benefits across children [16, 33], people with vision, mobility, or cognitive impairments [5, 30, 48], general pedestrians [3, 38, 40, 70], and cyclists [56]. Distance-dependent information improves understanding

and willingness to cross [37], visually impaired participants prefer speech, which we include in our optimization [30]. Open issues remain, including child interpretation [33], overtrust [54], scalability [20, 21, 31], and social implications [19, 64, 86, 87]. Recent work advises consolidation rather than yet more novel concepts [40], our parameter-focused optimization contributes to this consolidation.

eHMIs address missing cues between driverless AVs and vulnerable road users [18, 48, 55] and also with manual drivers [22]. Colley and Rukzio [28] propose a design space with three concept dimensions, *message type*, *modality*, *locus*, and six situational dimensions, *communication relationship*, *communication partner*, *number of lines*, *acoustic noise level*, *traffic autonomy*, *weather*. Industry and research explore auditory and visual symbols on the windshield [66], surveys of industry concepts [7, 35], abstract light elements [12, 67], road projections [8, 79], tracking and feedback [39, 85], and textual or auditory communication [73].

Cross-cultural evidence is limited; the symbol-based walking person eHMI affected willingness to cross and Trust in German and Chinese samples [59], polite strategies improved compliance and acceptance with cultural differences [65]; neither study was longitudinal. The need for eHMIs is debated [32], critics cite unclear targeting and attention costs [32, 93], yet willingness to cross often increases with eHMIs [4, 12]. Colley et al. [24] ran a longitudinal VR study in Germany, an online study with German and one with US participants, to examine eHMI effects across time and regions. eHMIs generally reduced mental workload and increased trust, understanding, and perceived safety. Repeated exposure strengthened these effects, yet VR and online methods yielded different absolute levels for many outcomes, so results are not interchangeable. Germans tended to wait longer and collide less, while US participants crossed earlier and collided more.

Our work adds a cultural lens to the trade-offs for eHMIs.

2.2 Multi-Objective Bayesian Optimization for Interface Designs

User interface design can be framed as an optimization problem in which design parameters (e.g., position, color, transparency) are mapped to user-centered objectives (e.g., trust, perceived safety) via an unknown, potentially noisy function, rendering exhaustive search impractical [11]. Bayesian optimization addresses this challenge by efficiently exploring such spaces with few samples, balancing exploration of under-sampled regions and exploitation of promising designs [9, 13, 68]. Prior work has applied Bayesian optimization to improve animation appearance, image aesthetics, font settings, and interactive UI behavior [10, 43, 60, 63]. In HCI, HITL variants close the optimization loop by incorporating user judgments directly into the search process [17, 63, 97]. Because interface design often involves competing objectives, MOBO has been adopted to return Pareto fronts rather than single optima, making trade-offs explicit [74]. MOBO has been used in HCI contexts ranging from multi-finger text entry to personalized explanations for machine-learning systems [14, 91].

Recent work applied HITL MOBO to automotive interfaces, where subjective perception and behavioral responses are tightly coupled and safety-critical. For example, prior work optimized eHMIs for pedestrian crossings, demonstrating that HITL MOBO can personalize eHMI parameters to improve trust and perceived safety [23]. Similarly, *OptiCarVis* [58] demonstrated that MOBO can be used to optimize in-vehicle windshield visualizations of AV functionalities (e.g., pedestrian detection) by balancing competing objectives, including understandability, distraction, and visual load. Together, these works demonstrate that HITL MOBO is viable in automotive contexts and can optimize for subjective user perceptions.

However, prior work focuses on design optimization for individuals, rather than on how optimized trade-offs differ across study regions. In particular, it remains unclear whether the Pareto-optimal design spaces identified by HITL MOBO are stable across regions with known differences in pedestrian behavior, or whether optimization converges on systematically different trade-offs even when task semantics are held constant (e.g., the same crossing task and AV stopping intent). Moreover, prior work does not examine which parts of the design space

account for such differences. Our work addresses this gap by using HITL MOBO as a comparative instrument to analyze Pareto-optimal eHMI trade-offs across regions.

2.3 Cultural Aspects in Road Crossing Behavior

Cultural background shapes pedestrian road crossing, affecting risk, timing, and the use of social cues [83]. In Nagoya and Strasbourg, larger waiting groups correlated with fewer illegal crossings. French pedestrians initiated earlier at green, Japanese pedestrians displayed higher risk awareness, crossed faster at red, and relied less on social information, consistent with prior France-Japan differences [80, 92]. Across Estonia, Greece, Kosovo, Russia, and Turkey, overall Pedestrian Behavior Scale patterns were similar, Greeks and Turks reported more transgressive behavior, Turks and Russians reported more missteps and aggression, and country context moderated how values relate to behavior [47, 90]. Cross-cultural field studies in Germany and Sri Lanka contrasted structured, rule-bound traffic with more fluid practices, using a Ghost Driver setup to examine interactions with automated and conventional vehicles [81]. Germany-Japan comparisons linked differences in risk avoidance, compliance, gap acceptance, and walking speed to Hofstede's uncertainty avoidance and collectivism versus individualism dimensions, with Japanese countries scoring higher on uncertainty and risk avoidance than Germany [50, 51, 77]. Overall, national norms, traffic organization, and rule enforcement systematically influence pedestrian crossing behavior.

Cross-cultural HCI research cautions against treating country labels as direct proxies for homogeneous or stable "cultures," highlighting risks of essentialism and over-attribution of observed differences to culture alone [57]. Related work further emphasizes methodological challenges in cross-region studies, including sampling comparability, translation quality, and the need to avoid assuming measurement equivalence across contexts [84]. Accordingly, HCI research encourages framing such comparisons in terms of study regions and interpreting differences as situated outcomes rather than as causal effects of culture [57, 84]. Following this guidance, we treat Japan and Germany as study regions with distinct traffic norms and interaction practices and use cultural constructs only as interpretive lenses grounded in prior literature, not as explanatory variables.

3 Method: Traffic Scenario and eHMI Concept

To simulate our eHMI concept (i.e., communicating AV stopping at an unsignaled pedestrian road crossing, see Figure 2), we used Unity (version 6.0). 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. [25]). VR is an appropriate choice for testing pedestrian crossing scenarios, as manipulating the eHMI design can endanger road users in the real world.

3.1 Pedestrian Crossing Scenario

The pedestrian scenario starts on 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 closest to the participant, AVs and manually driven vehicles operate at a 50:50 split. Only manually driven vehicles are in the far lane.

After 18 seconds, the next AV on the near lane will stop and let the participant cross. This duration enables the participant to experience AVs and manually driven vehicles, which represent an externally valid scenario. The manually driven vehicles on the far lane do not stop.

For external validity, we matched traffic handedness to country: right-hand traffic in the German scenario (conducted in Germany) and left-hand traffic in the Japanese scenario (conducted in Japan). To preserve internal consistency, we held the urban scene constant and selected streetscapes plausible in both contexts—areas of German and Japanese cities with a similar Western-style appearance. We intentionally avoided region-typical elements (e.g., traditional buildings) because the study did not target such environmental effects. A two-lane road layout and the vehicles used are transferable across both contexts.

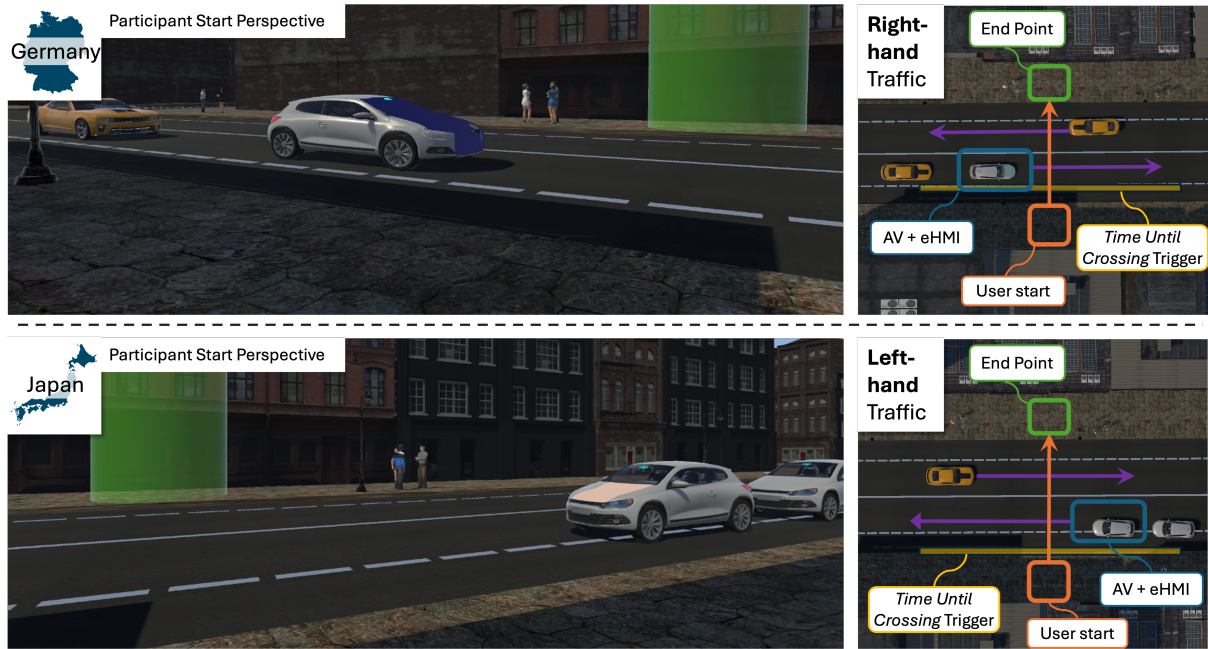


Fig. 2. Unity implementation of the crossing scenario from the participant's perspective **left** and from above **right**. The German scenario featured right-hand traffic (images adapted from Colley, Jansen et al. [23] with permission), and the Japanese had left-hand traffic.

3.2 Dynamic eHMI Design Generation

We defined the eHMI as a visual display on the hood, windshield, and grille inspired by eHMI LED light panels (e.g., see [46, 71]). This area is visible from a pedestrian's perspective for an upcoming AV. The area covers the AV's front but excludes headlights and front lights (see Figure 3).

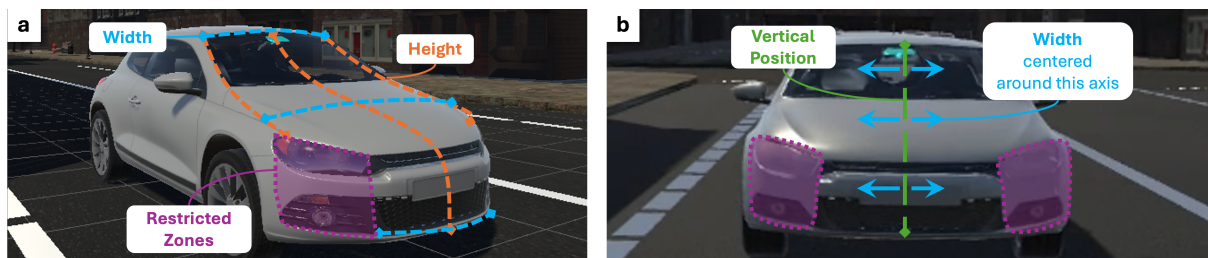


Fig. 3. The AV without an eHMI. A cyan light signal near the rearview mirror indicates that it is an AV. In (a), dashed lines show the available *width* and *height* for the eHMI mesh generation, including its lower/upper and left/right bounds. (b) Examples of possible mesh positions in allowed regions. 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). Image and eHMI generation adapted from Colley, Jansen et al. [23].

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 upper edge of the windshield. 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.

To generate an eHMI in Unity that can be positioned and resized dynamically across the windshield, hood, and radiator grille, we use the approach by [23]. The parameters and *width* and *height* boundaries 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

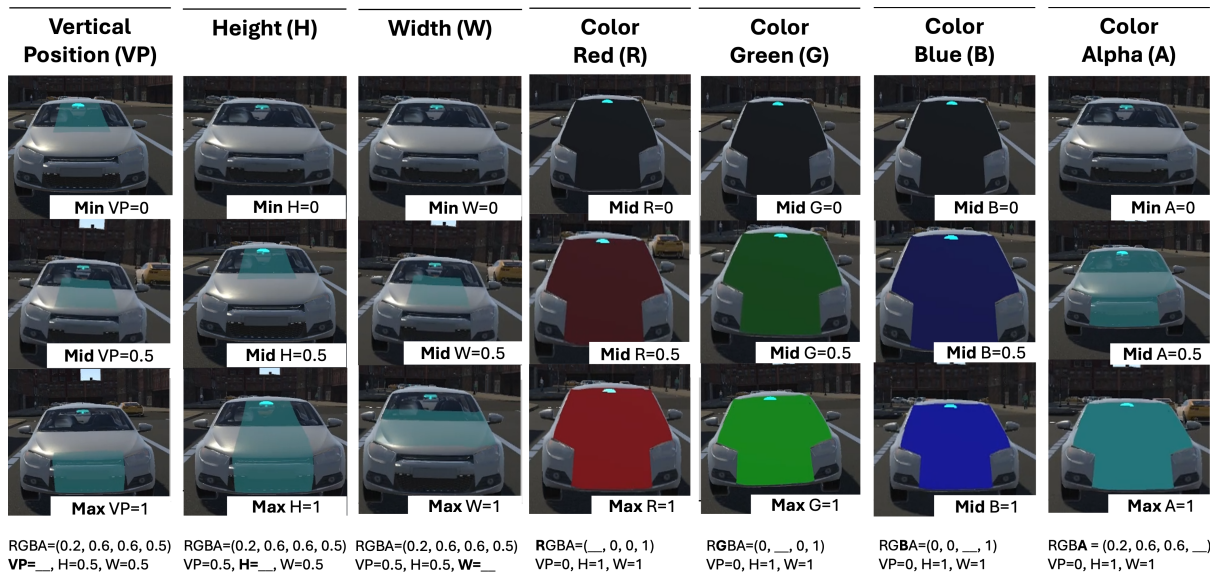


Fig. 4. eHMI design parameter value ranges. At the bottom of each column, the default values for the other parameters are shown. Cyan is used as the default color to demonstrate other parameter values. Image taken from [23].

4.1.1 Design Parameters. The eHMI design parameters were taken from Colley, Jansen et al. [23] (see Figure 4 and Table 1). While we changed the street setup to have left-hand and right-hand traffic, no other changes to the eHMI design parameters were necessary.

4.1.2 Design Objective Functions. As prior work showed that several subjective measures correlate in the given crossing scenario (e.g., perceived safety and trust) [23], we used fewer objectives to increase optimization efficiency. We focus on maximizing two subjective metrics: *trust* and *visual appeal*. Conversely, we aim to minimize *mental demand* as our sole subjective metric, along with one objective metric: *time to start crossing*.

Table 1. The 9 design parameters for the eHMI design, with ranges taken from [23]. All design parameters are modeled continuously. Example visualizations of parameter values are shown in Figure 4.

Design Parameter	Description	Reference	Range
p_1 : R	Red color channel.	[29]	[0, 1]
p_2 : G	Green color channel.	[29]	[0, 1]
p_3 : B	Blue color channel.	[29]	[0, 1]
p_4 : eHMI Alpha, α	Alpha color channel value.	[29]	[0, 1]
p_5 : Blink frequency	Blink frequency when activated. Maximum: 4Hz.	[36]	[0, 1]
p_6 : Width, W	Width. Maximum: entire width of the AV.	[29]	[0, 1]
p_7 : Height, H	Height. Maximum: entire height of the AV.	[29]	[0, 1]
p_8 : Vertical Position, VP	Position along the center AV axis (viewed from the front).	[29]	[0, 1]
p_9 : Auditory Message Loudness, l	Loudness of auditory message “Stopping”.	[30]	[0, 1]

We selected these four objectives to cover acceptance-related evaluation (trust and visual appeal), cognitive cost (mental demand), and behavior (time to start crossing) while keeping the objective set small enough for sample-efficient MOBO within a one-hour session. Prior work in the same task reports strong correlations among several subjective measures; reducing redundancy in the objective set improves optimization stability and reduces participant burden [23].

We employed the following questionnaires after every optimization run in the HITL process (as done in [23]): We assessed **mental demand** via the mental demand subscale of the raw NASA-TLX [49] 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 trust, we used the *Trust* subscale of the *Trust in Automation* questionnaire by Körber [62]. **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). We also adapted the question regarding **visual appeal** from Colley et al. [26] (“I found the visualizations visually appealing”; on a 5-point Likert scale). *Time to start crossing* is the duration after walking very close to the road since the simulation started (see Figure 2 b).

Normalization to the $[-1, 1]$ range is required because the subjective metric values span 20-, 5-, or 7-point Likert scales. We linearly transform each objective to the range $[-1, 1]$. We invert mental demand and time to start crossing so that higher values are better for all objectives.

4.1.3 Hyperparameter Setup for Bayesian Optimization. We used BoTorch [6] version 0.14.3 to implement MOBO with a multi-output Gaussian Process (GP) surrogate model and qEHVI (q-Expected Hypervolume Improvement) as the acquisition function. We chose a GP because we expected a few noisy observations and therefore benefited from a model that can both predict objective values and quantify uncertainty in unexplored regions of the design space. We used a multi-output variant to simultaneously optimize four design objectives. Based on the collected user feedback, the GP surrogate model estimates how eHMI parameter values relate to these four objectives. In other words, after each iteration, it predicts which regions of the design space are likely to contain good trade-offs, together with uncertainty about regions that have been explored less.

The acquisition function uses these predictions to decide which design to show next. It scores candidate designs according to how useful they are expected to be for improving the current set of Pareto-optimal solutions. qEHVI is designed for multi-objective optimization and therefore favors designs that are expected to improve the current Pareto front across the objectives [74]. This means that the next design is selected because it is expected to either improve the current trade-offs or reduce uncertainty in relevant parts of the design space. In this way, the acquisition function balances *exploration*, that is, testing uncertain regions of the design space, and *exploitation*,

which entails refining regions that already appear promising. We set $q = 1$ (see Chan et al. [13]), such that exactly one eHMI design was proposed and evaluated in each iteration.

The optimization process consisted of two phases: an initial *sampling* phase and a subsequent *optimization* phase. In the sampling phase, participants first rated 19 eHMI designs generated with Sobol sampling [89]. Sobol sampling spreads designs broadly across the 9-dimensional parameter space and thus provides an informative initial coverage of possible parameter combinations. This phase was necessary because the optimizer lacked user-specific data at the start of the study. To keep starting conditions comparable across participants, we used the same set of 19 sampled designs for everyone.

After these initial ratings, the procedure entered the *optimization* phase. Here, the model was first fitted to each participant's ratings from the sampling phase and then iteratively updated after every new trial. Based on the current model, qEHVI selected the next design expected to improve the Pareto front across the four objectives. This phase lasted 6 iterations. To compute the next candidate design, we used 2024 restart candidates and 512 Monte Carlo samples, following prior automotive HITL MOBO work [23, 58].

In total, the procedure comprised 25 iterations, which fit within a 60-minute session while limiting participant fatigue. The choice of 19 initial samples followed the $2n+1$ heuristic recommended in the BoTorch tutorial¹, with $n = 9$ eHMI design parameters. We chose this relatively sampling-heavy setup because we expected considerable variability across participants and study regions, and therefore prioritized broad initial coverage before adapting the search to each participant's ratings.

4.2 Additional Measurements

In addition to the optimization objectives (see Section 4.1.2), we recorded several behavioral measures throughout the VR experiment in Unity. We logged participants' movement data at 50 Hz using VR headset tracking, that is, the user's position and orientation in the virtual environment. From these data, we derived the number of collisions with vehicles, the time spent on the sidewalk, the time spent on the street, and the total trial duration.

After completing all 25 iterations, participants evaluated their experience with the HITL process and the eHMI design using five single-item self-report measures adopted from Jansen et al. [58]. All items were answered on 7-point Likert scales ranging from 1 = *Strongly Disagree* to 7 = *Strongly Agree*. The measures were: *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."), and *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.") They were then also able to provide *general* feedback: "What was your general impression of the interaction with the automated vehicles?" and *Improvement* proposals: "Do you have any improvement proposals?"

4.3 Study Procedure

First, participants were introduced to the study, its procedures, and the VR scene (see Figure 5). The introduction is in Appendix A. Participants provided informed consent, then adjusted the headset. Each trial used the same street scene and settings to avoid external influences on eHMI perception. The AV stopped after ≈ 25 s. Each trial, including the crossing, lasted ≈ 1 min. After each trial, participants completed the questionnaires described in Section 4.2 within the VR scene. The session lasted ≈ 60 min. Participants received €10 or 2000 Yen. The study was conducted in German or English in Germany, while in Japanese or Chinese in Japan. Study instructions were translated in close collaboration with native speakers.

¹https://botorch.org/docs/tutorials/multi_objective_bo; accessed 30.04.2026

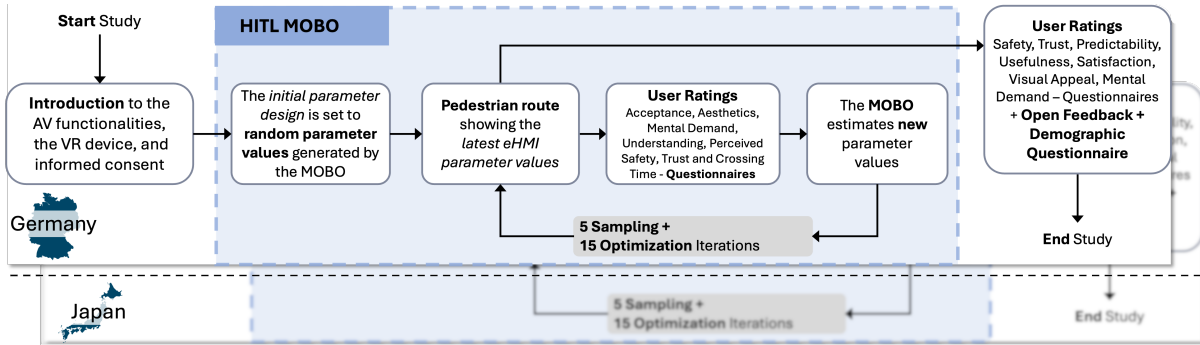


Fig. 5. Study procedure using HITL MOBO for eHMI design. There were 25 iterations (19 for sampling and 6 for optimization). Image adapted from [23] with permission. The study procedures in Germany and Japan were identical.

The experimental procedure followed our university's ethics committee guidelines 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

We report quantitative and qualitative results. We first summarize the optimization trajectories across iterations, then compare Pareto-front trade-offs between Japan and Germany (including dominance and parameter-pattern analyses). We then report process evaluations of the HITL MOBO interaction and conclude with qualitative themes from participants' open feedback to contextualize the numeric findings.

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 [74]. Using the *EMOA* [75] (Evolutionary Multiobjective Optimization Algorithms) and *moocore* [72] R packages, we determined Pareto-optimal values for each participant and focused our analysis solely on these efficient designs. R 4.6.0 and RStudio 2026.04.0 were employed. All packages were up to date in April 2026.

5.2 Participants

We report results by study region (Germany vs. Japan) as residence and test location were aligned with the traffic setup, although nationality within the Japan region varied. Across both regions, $N = 40$ participants took part ($n = 20$ living in Germany, $n = 20$ living in Japan).

In Japan, the sample comprised 7 female and 13 male participants ($M = 27.00$, $SD = 2.83$ years old); all held a college degree, including both Japanese and Chinese participants residing in Japan at the time of the study. On 5-point Likert scales ($1=Strongly Disagree - 5=Strongly Agree$), they reported interest in AVs ($M=4.55$, $SD=.76$), positive expectations that AVs would ease their lives ($M=4.70$, $SD=.57$), and skepticism that AVs would become reality by 2035 ($M=4.45$, $SD=.61$).

In Germany, there were 7 female and 13 male participants ($M = 25.40$, $SD = 3.68$ years old); 12 held a college degree, 7 a high school degree, and 1 a secondary school degree. On the same scales, they reported lower interest in AVs ($M=3.45$, $SD=1.36$), were less positive that AVs would ease their lives ($M=4.00$, $SD=1.08$), and were less skeptical that AVs would become reality by 2035 ($M=4.05$, $SD=1.23$).

In Germany and Japan, participants were recruited locally via mailing lists and personal invitations. Inclusion criteria were being at least 18 years old, being able to complete the VR pedestrian-crossing task independently, understanding the English cue “Stopping,” and not reporting high susceptibility to simulator sickness in a verbal pre-screening. No participant had to abort the study, for example, due to simulator sickness, and no participant reported vision or hearing disabilities.

5.3 Questionnaire Ratings

Figure 6 shows the iteration-wise progression of the four design objective values for Germany and Japan, separated into the shared sampling phase (iterations 1–19) and the subsequent optimization phase (iterations 20–25) of the HITL MOBO process. For each iteration, the plotted points represent the group means within each region, with error bars indicating variability across participants. The dashed lines show simple linear regressions fitted separately for each region across all 25 iterations.

For trust (Figure 6a), both regions show an overall upward trend across iterations, with higher values in the final optimization iterations than at the beginning of the procedure. The German trajectory is comparatively stable across the full process, whereas the Japanese trajectory fluctuates more strongly during sampling and rises more clearly during optimization. A similar pattern is visible for aesthetics (Figure 6b): values increase in both regions, but the increase is steeper for Japan than for Germany.

For mental demand (Figure 6c), the values also show an upward trend in both regions, with the Japanese trajectory again changing more strongly across iterations than the German one. For time to start crossing (Figure 6d), the German trajectory remains almost constant across the full procedure, whereas the Japanese trajectory increases slightly over iterations. Thus, the four panels indicate different degrees of change across objectives and regions, with the clearest visible late-phase increases for trust and aesthetics in Japan.

5.4 Comparing Pareto-optimal trade-offs between Japan and Germany via the additive ϵ indicator

To compare the two Pareto sets directly, we used the directional additive ϵ indicator (see Figure 7), a standard set-based metric for multi-objective optimization [61, 99]. Here, *additive* means that the comparison is expressed as the smallest constant offset in normalized objective space needed to shift one Pareto set so that it weakly dominates the other. Thus, an additive gap measures the smallest constant value that must be added to all objectives of one Pareto set so that it is weakly dominated by another set. *Directional* means that this offset is computed separately for both directions, that is, from Japan to Germany and from Germany to Japan. As a result, $\epsilon(\text{Japan} \rightarrow \text{Germany})$ and $\epsilon(\text{Germany} \rightarrow \text{Japan})$ need not be equal. Smaller values indicate that one set is already closer to covering the trade-offs of the other, whereas larger values indicate that a larger shift is required. In this sense, the reported ϵ values are directional additive gaps: they quantify the size of the required shift from one region’s Pareto set to the other.

We computed the additive ϵ indicator for objectives normalized by their minimum and maximum values, so that all objectives were on a common scale, with lower ϵ values indicating smaller required shifts. The directional gaps were $\epsilon(\text{Japan} \rightarrow \text{Germany}) = 0.333$ and $\epsilon(\text{Germany} \rightarrow \text{Japan}) = 0.867$, yielding $\Delta = -0.534$. Thus, the Japanese Pareto set required a smaller additive shift to weakly dominate the German Pareto set than vice versa. A participant-clustered permutation test with 5000 permutations gave $p < 0.0002$; the observed Δ lay far outside the null interval $[-0.208, 0.208]$. Under the chosen normalization and ϵ -based comparison, this indicates that the Pareto set obtained in Japan was significantly closer to weakly dominating the Pareto set obtained in Germany than vice versa.

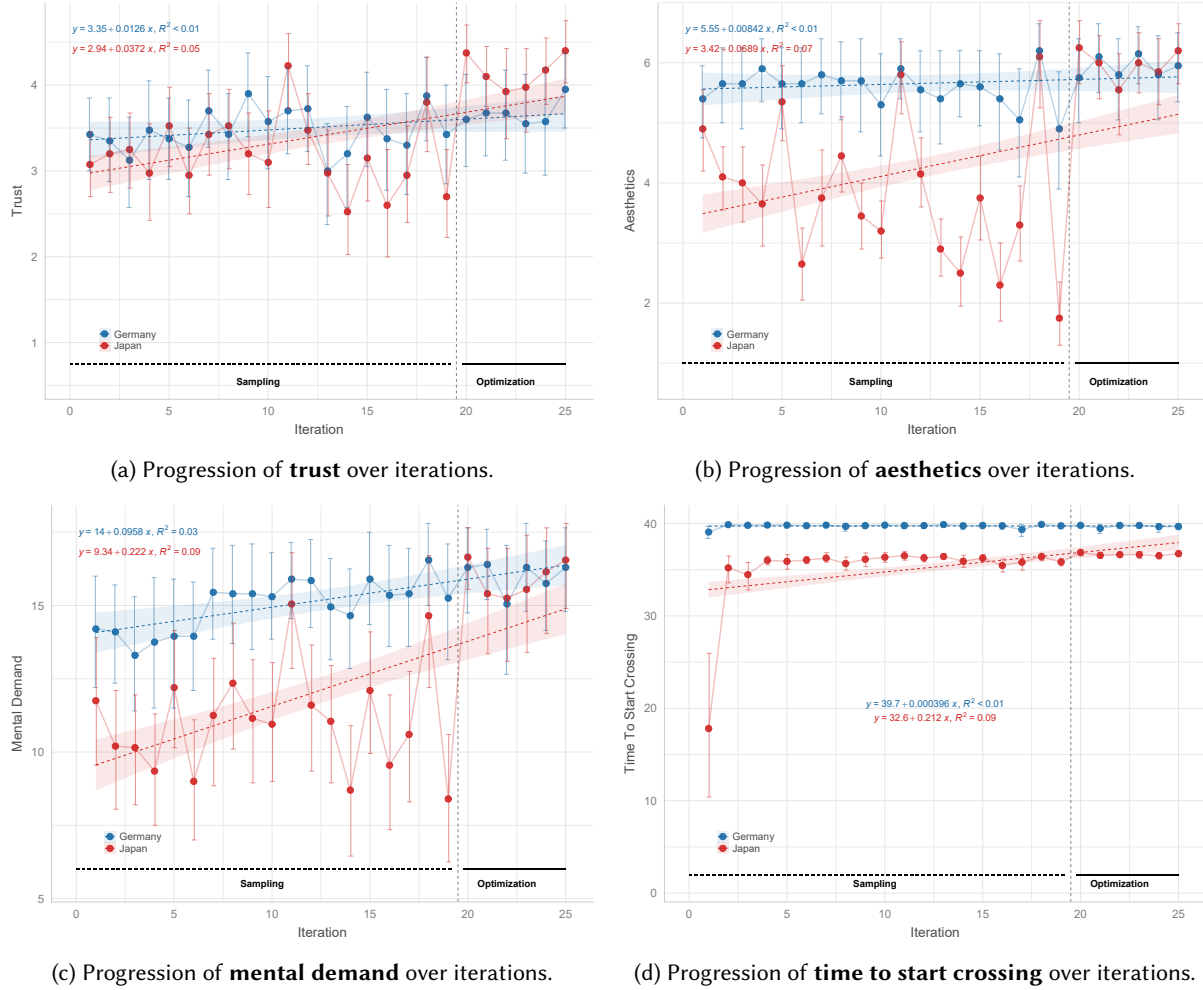


Fig. 6. Progression of the four MOBO objective values: trust, aesthetics, mental demand, and time to start crossing.

5.5 Pareto Front Design Parameter Set

Figure 8 shows the mean normalized Pareto-optimal parameter values for Germany and Japan, together with the overall spread of values across participants. All mean parameter values lie above the normalized midpoint of 0.5. In addition, the grey rectangles indicate that the overall spread of Pareto-optimal values is broader than the between-region mean differences for most parameters.

Averaging RGB values would hide important structure, so Figure 9 shows Pareto-optimal color values per participant across all 25 iterations. Overall, the figure shows recurring light blue–green (close to cyan) and magenta Pareto-optimal colors together with substantial variation across regions and participants.

During the sampling phase (iterations 1–19), all participants were exposed to the same set of sampled designs generated by the Sobol procedure (see Section 4.1.3), such that differences in Figure 9 do not reflect different sampled color sets, but differences in which of these shared sampled colors became Pareto optimal for a given

Epsilon Indicator Analysis: Regional Comparison of Pareto Fronts

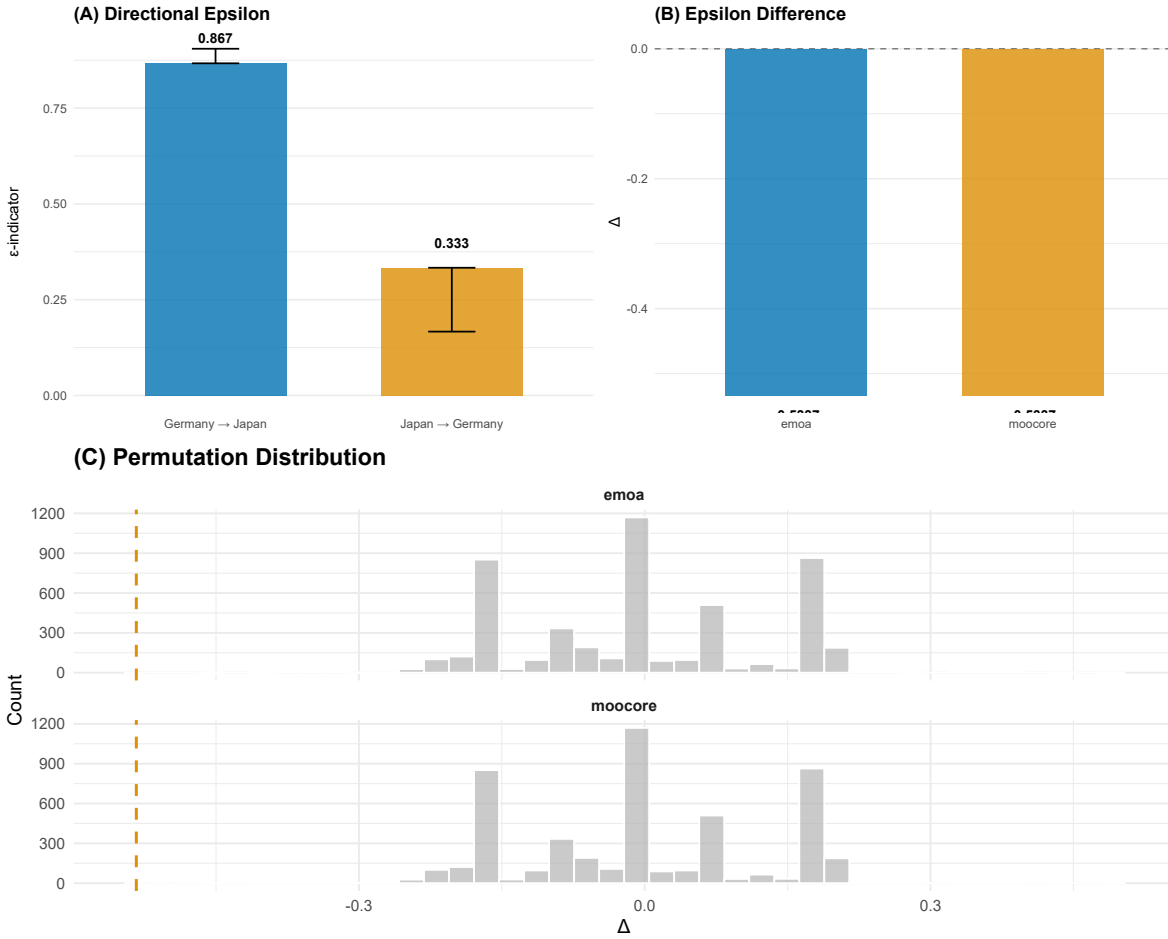


Fig. 7. Epsilon indicator analysis comparing Pareto fronts between Japanese and German participants. (A) Directional epsilon values (with 95% CI) quantify the additive distance to weak dominance. (B) The difference $\Delta = \varepsilon(\text{Japan} \rightarrow \text{Germany}) - \varepsilon(\text{Germany} \rightarrow \text{Japan})$ indicates asymmetric optimization outcomes ($***p < 0.001$ from permutation tests). (C) Null distributions from 5000 permutations indicate that the observed Δ values are statistically significant. Results from two independent implementations (emoa, moocore) demonstrate robustness. Negative Δ values indicate that the Pareto front obtained in Japan lies closer to weakly dominating the German Pareto front than vice versa across the four objectives (Trust, Mental Demand, Aesthetics, Crossing Time).

participant. Pareto-optimal colors appear for nearly all participants; specifically, 39 of 40 participants showed at least one Pareto-optimal color, indicating broad early coverage of the color space. The sampling phase also shows a wider visual spread of Pareto-optimal colors than the optimization phase. In Germany, this distribution is less uniform across participants, with high numbers of Pareto-optimal colors concentrated in only a few participant

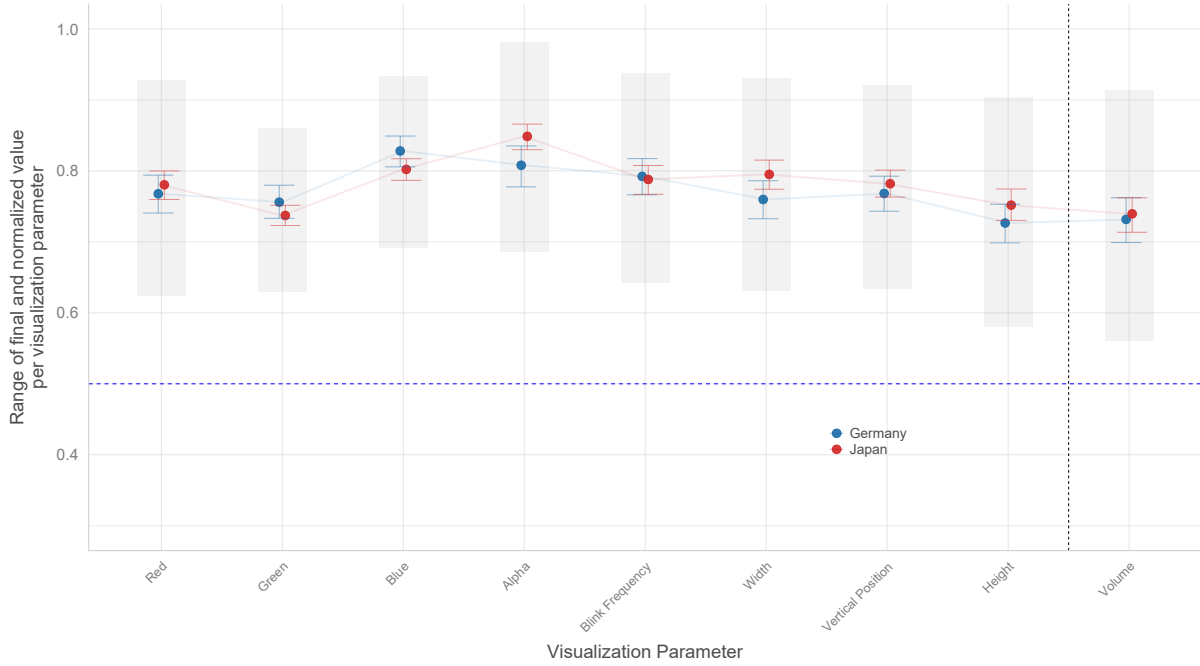


Fig. 8. Final parameter set per country. Presented are the **Pareto front** values per participant as a **mean** that is normalized. The grey rectangle shows one standard deviation from the mean of all values. The x-axis shows the ordered visualization parameters (p_1 to p_9) from left to right.

columns. In Japan, Pareto-optimal colors are distributed more evenly across participants, with more participant columns containing intermediate numbers of Pareto-optimal colors.

In the subsequent optimization phase (iterations 1–6), Pareto-optimal colors appear sparser and are more concentrated in a narrower light blue–green neighborhood, although several participants still had multiple Pareto-optimal colors. This phase also differs across regions: in Germany, Pareto-optimal colors are concentrated in fewer participants, including one particularly tall column in the light blue–green range, whereas in Japan, they are distributed across more participants, typically with shorter columns.

Figure 10 summarizes the absolute mean differences between Japan and Germany for each Pareto-optimal eHMI design parameter. The largest between-region differences are observed for the color-related parameters, with blue showing the highest difference, followed by green and red. Width also shows a comparatively large difference, whereas blink frequency, volume, vertical position, height, and alpha show smaller separations. According to the non-parametric tests, only red and green reach statistical significance.

5.6 Correlation between the Design Objectives

MOBO identifies Pareto-optimal design parameters, where any improvement in one objective requires a loss in another [74]. We therefore examined correlations among objectives for two reasons: (1) to make the *structure* of trade-offs visible (which objectives tend to move together vs. oppose each other), and (2) to check whether any objective is largely redundant (highly correlated) or decoupled (independent), which affects how practitioners should interpret multi-objective recommendations. The results show that almost all correlations were statistically



Fig. 9. Distribution of Pareto-optimal color values for each participant, divided by study region (Germany vs. Japan) and phase. Each column represents a single participant. The colored boxes show the actual eHMI colors that achieved Pareto optimality (i.e., they were not dominated by other designs). The number inside each box indicates the specific iteration in which that color was evaluated, counted from bottom to top. Missing numbers (empty spaces) represent iterations where the tested color was sub-optimal and discarded. **Key takeaway:** The varying stack heights and color palettes demonstrate highly individualized color preferences and trade-offs, with the optimization phase successfully discovering new optimal designs tailored to each user.

significant when using all data for Germany (see Figure 11a). This is not the case for the data from Japan, where the Time to start crossing is uncorrelated with any other objective (see Figure 11b).

When only taking Pareto front values into account (see Figure 12a and Figure 12b), both in Japan and Germany, no correlation between Time to start crossing and any other objective exists. Both with all data and with only Pareto values, the other objectives are positively correlated.

5.7 Design Satisfaction and Engagement

Table 2 summarizes participants' experience in the HITL MOBO process. We used Wilcoxon signed-rank tests to compare them. Expectation conformity, satisfaction, agency, and confidence were descriptively positive in both

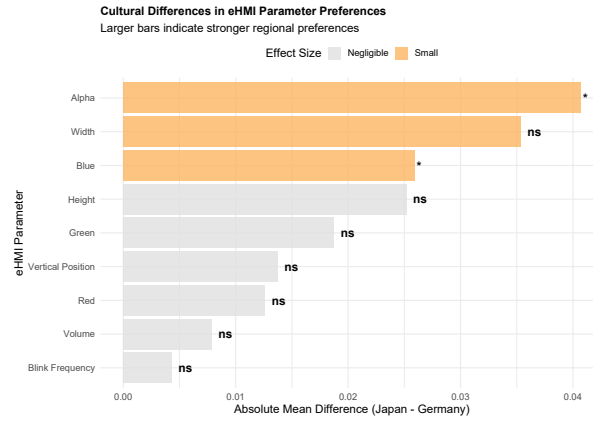


Fig. 10. Variation in eHMI parameter preferences. Each bar represents the absolute difference between mean parameter values selected by Japanese ($n=20$) and German ($n=20$) participants on their individual Pareto fronts. Parameters are ordered by effect size magnitude. Statistical significance from non-parametric tests is indicated by asterisks. Color intensity reflects Cohen's d effect size, providing a measure of practical significance beyond statistical significance. Results suggest that color-related parameters exhibit the strongest regional specificity, while certain geometric parameters (e.g., Height, Volume) show more universal preferences (ns = not significant).

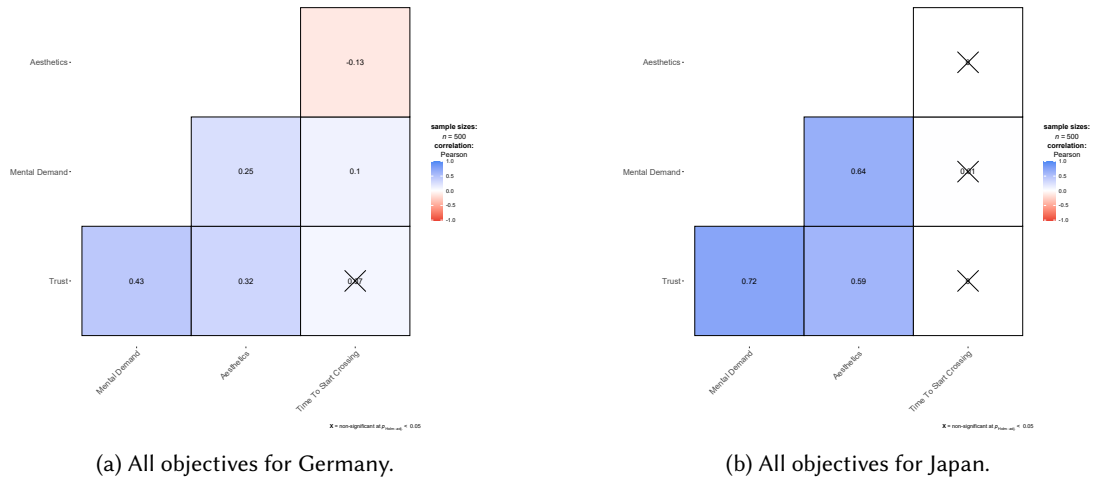


Fig. 11. Correlation heatmaps of study objectives. The images include all data points ($n = 500 = 25 * 20$, i.e., 20 iterations for 20 participants per region).

Japan and Germany. Ownership showed the lowest mean values of the five measures in both groups. However, no significant differences were found. As in previous work [23], the results indicate that participants were generally satisfied and felt in control, while ownership of the final design remained comparatively weaker.

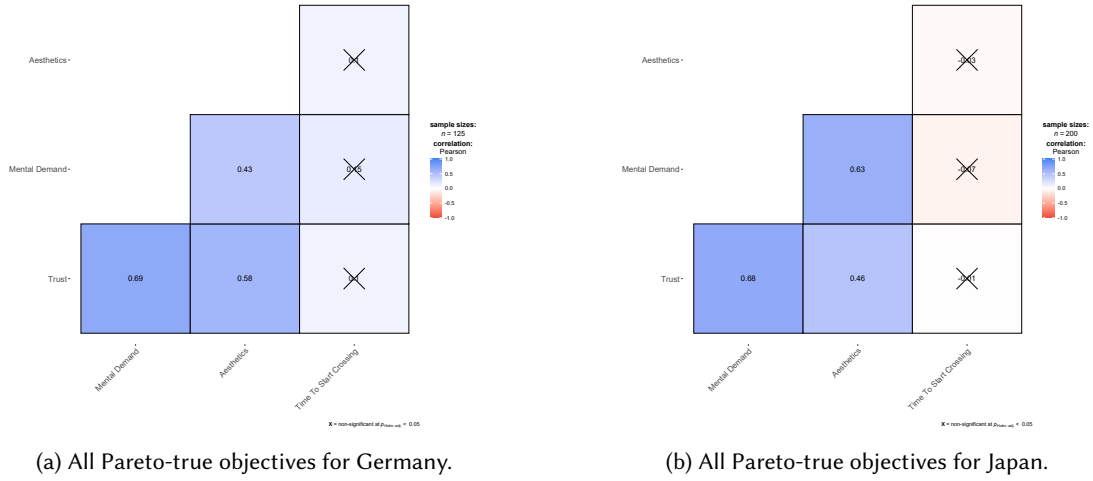


Fig. 12. Correlation heatmaps of Pareto-true study objectives per region.

Table 2. Design satisfaction and engagement ratings. Region differences were tested using Wilcoxon signed-rank tests.

Measure	Item	Japan M (SD)	Germany M (SD)	W	p
Expectation conformity	"The final design matches my imagination."	4.85 (2.06)	5.20 (1.70)	181.50	0.61
Satisfaction	"I'm pleased with the final design."	4.95 (2.01)	5.30 (1.45)	188.50	0.76
Agency	"I felt in control of the design process."	4.55 (1.90)	4.80 (1.28)	196.50	0.93
Confidence	"I believe the design is optimal for me."	4.60 (2.09)	5.05 (1.54)	183.00	0.65
Ownership	"I feel the final design is mine."	4.60 (2.11)	4.75 (1.37)	206.00	0.88

5.8 Open Feedback

Participants provided textual feedback at the end of the study regarding their experience with the design and interaction. Because this was not a dedicated qualitative study, we treat these comments as illustrative context rather than as a comprehensive thematic analysis.

We conducted a lightweight inductive analysis. Two authors independently reviewed all responses and identified recurring concepts. We then consolidated these into a small set of shared themes and selected representative quotes. Disagreements were resolved through discussion. No formal codebook or saturation criterion was applied.

We highlight the topics around our open questions regarding interactivity, general impressions, and areas for improvement.

Interactivity and Control. German participants emphasized control over signals and stopping behavior. They reported wanting fewer surprises and clearer mappings from feedback to the next change: "I feel less control over the crossing the road part", "I liked it more when it was not blinking when the car stopped", "I wanted more control over the shape of the indicator", "I'd like to have been able to signal the cars to stop, e.g. holding my hand out."

Japanese participants focused more on tunable visual properties and placement: "I felt that I can control the size of eHMI. desired more: the place & color of eHMI on car" "I prefer sky blue and desire more eHMI area to

cover the car,” “position and area”, “I feel I only can control the size of visualization. If possible, I would like to control the position.”

General Impressions of AV Interaction. Overall impressions were cautiously positive but split in emphasis. German participants highlighted predictability and reliability alongside occasional failures: “They were predictable, which I liked,” “Mostly reliable,” yet also “the cars hit me twice, but in other rounds they do the right things.”

Japanese participants more often framed trust as conditional and noted interactions with other traffic: “great,” “I can trust the automated vehicles if everything goes correctly,” “It works well in car-pedestrian interaction, but need more improvements on the autonomous vehicle-other traditional vehicle (driver) interaction”, “The functionality is not that mature”, “Generally I trust them, still I need double-check to determine.”

Areas for Improvement. Requests from Germany centered on workload and traffic flow: “Less cars in the screen and one car to interact with,” “Maybe the waiting time before the automated vehicle stops could be reduced?” “Maybe fewer tries, 25 for me is too many.”

Japanese participants asked for clearer semantics and additional cues: “add text,” “It would be great if the automated vehicles could hint at the incoming cars from the opposite direction,” “Maybe change the background of crossing street,” “let user choose a specific color ... instead of other random colors.”

Collectively, feedback indicates a cross-regional split between control over interaction dynamics (Germany) and richer, user-steerable visual semantics and placement (Japan). This pattern suggests potential regional differences in priorities, which may inform future design considerations.

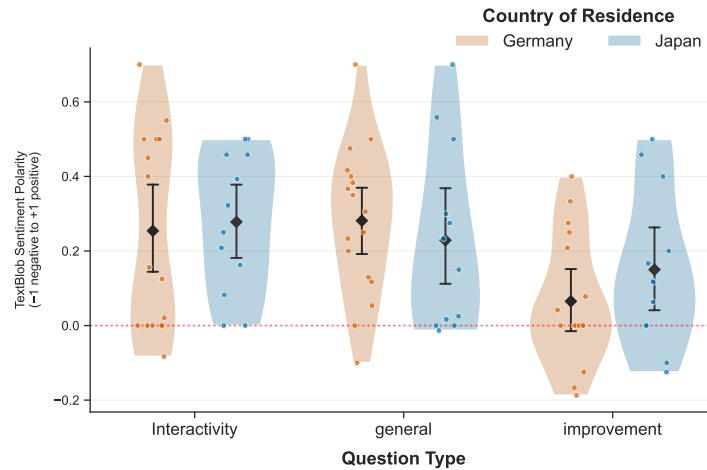


Fig. 13. TextBlob Sentiment Polarity by Question Type and Country of Residency. Post-hoc Wilcoxon signed-rank tests (Holm-Bonferroni corrected) revealed no significant pairwise differences between question types within either country.

Sentiment Analysis. Furthermore, Sentiment analysis was conducted using the [TextBlob](#) Python library, which assigns a polarity score ranging from -1 (highly negative) to +1 (highly positive). To analyze the differences in sentiment, a 2×3 Mixed Analysis of Variance (ANOVA) was conducted with Country of Residency (Germany vs. Japan) as the between-subjects factor and Question Type (Interactivity, General, Improvement) as the within-subjects factor (see [Figure 13](#)). Because sentiment scores frequently violate the assumption of normality, post-hoc pairwise comparisons between question types within each country were conducted using non-parametric

Wilcoxon signed-rank tests. To control for the family-wise error rate, p-values were adjusted using the Holm-Bonferroni correction.

The Mixed ANOVA revealed a significant main effect of Question Type on sentiment, $F(2, 54) = 5.17$, $p = .009$, partial eta-squared = 0.16. There was a non-significant main effect of Country of Residency, $F(1, 27) = 0.18$, $p = .676$, partial eta-squared = 0.01. Furthermore, the interaction between Question Type and Country of Residency was non-significant, $F(2, 54) = 0.68$, $p = .509$, partial eta-squared = 0.02.

Post-hoc Wilcoxon signed-rank tests (Holm-Bonferroni corrected) were utilized to examine differences between question types within each country. For participants residing in Germany, sentiment regarding Interactivity ($M = 0.25$, $SD = 0.26$) was significantly higher than improvement ($M = 0.07$, $SD = 0.17$), $W = N/A$, $p = .027$; sentiment regarding general ($M = 0.28$, $SD = 0.20$) was significantly higher than improvement ($M = 0.07$, $SD = 0.17$), $W = N/A$, $p = .027$. Comparisons that are not explicitly mentioned did not reach statistical significance ($p > .05$).

6 Discussion

While numerous works evaluated different eHMI designs [27, 36, 94], the question of what constitutes an *optimal* eHMI remains open. Early evidence suggests HITL MOBO can optimize eHMI parameters [23], yet the role of background in shaping both the resulting design and the design process is unclear.

This paper evaluates HITL MOBO as a method for multi-objective eHMI parameter search and for comparing Pareto trade-offs across study regions. The Germany–Japan comparison is therefore reported as site-specific evidence from one scenario and one sample, not as a general claim about European or Asian culture.

Guided by this gap, we examine cross-regional differences (Germany vs. Japan) in two respects: (RQ1) the characteristics of eHMIs optimized for pedestrians' trust, mental demand, perceived aesthetics, and time to cross; and (RQ2) how the optimization process is experienced, including satisfaction, agency, confidence, ownership, and overall user experience.

We conducted a between-subject study in Japan ($n=20$) and Germany ($n=20$) to investigate these RQs. We employed HITL MOBO to iteratively optimize the eHMI designs based on participant ratings of trust, mental demand, aesthetics, and time to start crossing. This is a design-method innovation, as it allows not only to distinguish individual designs but also to identify design trade-offs. Furthermore, this method enables us to determine whether there is a Universal Design for all participants, that is, one that all had on their individual Pareto fronts. Designs on the Pareto front mean enhancing one objective would necessitate trade-offs with others [74].

6.1 Regional Differences Regarding *Optimal* eHMI Design

Pareto sets differed by region. Using the additive ε indicator on normalized objectives (lower is better), we observed an asymmetric gap, with $\varepsilon(\text{Japan!} \rightarrow \text{!Germany}) = 0.333$ and $\varepsilon(\text{Germany!} \rightarrow \text{!Japan}) = 0.867$ ($\Delta = -0.534$). Under this normalization, the Pareto-optimal designs identified in Japan lie closer to jointly desirable trade-offs across trust, visual appeal, mental demand, and time to start crossing. Importantly, this does not mean that Japanese participants rated designs as uniformly better on any single metric. Rather, it indicates that the *shape of the trade-off space* differs by region. For example, an eHMI design that is both trusted and visually appealing without increasing mental demand, while still eliciting an early crossing response, may be attainable in one region but not in the other.

This interpretation aligns with prior work reporting cross-regional differences in pedestrian–vehicle interaction, including willingness to cross, trust, compliance, and risk avoidance under uncertainty [50, 59, 77, 83]. We extend this literature by examining these differences in a multi-objective design space rather than along isolated design objective dimensions. Using a formal dominance metric shows that regional effects persist when considering only Pareto-optimal designs and that behavioral timing ("time to start crossing") can be decoupled from subjective

evaluations such as trust or visual appeal. For instance, an eHMI design can be rated as trustworthy and aesthetically pleasing without necessarily leading to faster crossings, or vice versa. Moreover, this decoupling indicates that design objectives that are often assumed to covary in HCI (e.g., higher trust implying quicker action) do not necessarily move together on the Pareto front.

Having established that regional differences emerge at the level of multi-objective trade-offs, we next localize where these differences appear in the eHMI design space. The largest between-region separations occur in the RGB *color* parameters (see Figure 10), which in our implementation primarily modulate visual salience, that is, how strongly an already visible light signal attracts attention relative to its surroundings through hue and saturation. By contrast, parameters that mainly affect detectability—whether the signal can be perceived at all under given sensory constraints—such as eHMI *size* (e.g., height) or *opacity* (e.g., α), show smaller or non-significant differences (see Figure 8).

This result partially diverges from the construct-based expectations that motivated our experiment (see Section 2.3). In the following, we therefore use cultural constructs discussed in prior work as interpretive lenses. Prior work links Germany–Japan differences in pedestrian behavior to uncertainty avoidance and individualism–collectivism [50, 51, 77], which would predict stronger preferences for redundancy-enhancing parameters that increase detectability, such as greater opacity, larger display area, or additional auditory cues. Instead, this (*color* > *size/opacity*) pattern indicates that—assuming a constant communicated stopping intent—regional differences are already expressed through preferences for visual salience rather than changes in basic detectability.

Correlational structure also differed. Across all observations, most pairwise correlations among subjective objectives were significant in Germany, whereas in Japan, the behavioral metric *Time to start crossing* was uncorrelated with other objectives. Restricting to Pareto-true points, *Time* was uncorrelated with subjective objectives in both regions, while subjective objectives were positively correlated. Thus, on the Pareto front, behavioral timing did not track subjective ratings, consistent with multi-objective trade-offs [74].

Optimization progressed similarly across regions. Trajectories showed increasing trust and aesthetics, decreasing mental demand, and shorter *Time to start crossing* over iterations (see Figures 6a–6d), consistent with HITL MOBO’s exploration–exploitation behavior [6, 11, 13, 23]. Besides, in-loop experience did not differ significantly between regions on expectation conformity, satisfaction, agency, confidence, or ownership (Wilcoxon, $p \geq .61$); descriptive means were positive in both, aligning with prior reports that users can work productively with iterative optimization [23].

This quantitative parity is strongly corroborated by our sentiment analysis of the open-ended feedback. The lack of a main effect for study region in the sentiment scores confirms that participants in both Germany and Japan shared a similarly positive baseline experience with the HITL-MOBO process and the AV interactions. Interestingly, while our qualitative thematic analysis revealed that the focus of the feedback diverged—with German participants emphasizing control and predictability, and Japanese participants prioritizing visual semantics—the underlying emotional valence remained consistent. The only significant variation in sentiment was driven by question type (specifically in Germany, where asking for “improvements” naturally elicited more critical, problem-focused language than general impressions). Taken together, these findings suggest that while regional backgrounds may shape which eHMI parameters users prioritize (as seen in the Pareto fronts), they do not negatively impact the universal acceptance and usability of the optimization process itself.

Sample and protocol factors require interpretation. Groups differed in age and education (Japan: higher mean age, all with college degrees; Germany: lower mean age, mixed education), baseline attitudes toward AVs differed descriptively, and scenarios used left- vs. right-hand traffic. These are potential covariates, and we highlight that we make no causal claims based on them.

6.2 Towards Optimal eHMIs for Multi-Pedestrian Encounters

Urban crossings rarely involve a single pedestrian. Co-located groups mix cultural conventions and heterogeneous preferences, so an eHMI that works for one person may confuse another. Our HITL MOBO pipeline optimizes for one user at a time and may not transfer directly to new users. A practical next step is *group-based* MOBO that learns group preferences from aggregated feedback (e.g., [69]). Future work should prototype crowd-robust eHMIs with redundancy across modalities (color, motion, text/icon), multi-view layouts for varied vantage points, and fallbacks to traffic norms when consensus is low. Study designs should move from single-user loops to micro-groups and compare individual and group outcomes. In deployment, group-optimal eHMIs could be computed offline, with on-vehicle detection assigning an observed group to an available design. A key challenge is inferring the “right” group when information on nearby pedestrians is scarce. Vehicle-to-X approaches that query user information from their smartphones merit investigation.

6.3 Design Implications for eHMI Practitioners

We derive practical implications by translating Pareto-optimal parameter patterns into defaults and adjustment priorities for eHMI implementations under similar constraints (unsignalized crossing; light-based panel with optional audio). The evidence for each implication is grounded in (i) parameter distributions on the Pareto front (Figure 8 and Figure 9) and (ii) region differences in Pareto-front parameters (Figure 10).

Implication 1: Constrain the eHMI color palette and personalize within that neighborhood. Figure 9 shows that Pareto-optimal solutions concentrate in a cyan-like region across participants, despite broad sampling of the RGB space early in the session. This recurrence suggests that, within our design and scenario constraints, cyan-like hues are a robust “safe default” for signaling intent. We therefore recommend selecting a constrained palette centered on this region and supporting personalization within it (e.g., small shifts in hue/saturation) rather than switching between semantically different colors.

The rationale for prioritizing color for locale-specific adjustment comes from Figure 10: the largest between-region differences on the Pareto front occur in the color channels (p_1 – p_3), while several non-color parameters show smaller or non-significant differences. Taken together, these results indicate that color is both consistently present in Pareto-optimal designs and comparatively more sensitive to site, making it a high-leverage parameter for design adjustment.

Implication 2: Increase eHMI detectability primarily via opacity and area before adding temporal dynamics. The Pareto-front parameter distributions in Figure 8 indicate that successful eHMI designs typically rely on static salience controls such as opacity (p_4 , α) and visible area (p_6/p_7 , width/height), together with a stable placement (p_8), rather than requiring aggressive temporal modulation. Based on this pattern, we recommend an adjustment order that first increases α and/or area to improve detectability (e.g., for distance, glare, or visual clutter). Blink frequency (p_5) should be treated as a secondary adjustment when static salience is insufficient, because it trades improved noticeability against potential distraction.

Implication 3: Use audio to add redundancy when visual changes are constrained. Audio provides an additional channel that can raise eHMI detectability without changing the visual appearance or the semantics of the signal. In our parameterization, this corresponds to increasing loudness (p_9). The practical interpretation is to treat audio as an auxiliary redundancy mechanism: increase p_9 when visual salience cannot be increased further (e.g., aesthetic constraints or regulatory limits), rather than using audio as the primary carrier of intent.

6.4 Limitations and Future Work

We treated subjective ratings as the primary objectives for UI parameter optimization and included one behavioral objective: time to start crossing. The value of adding additional objective signals, such as physiological measures,

remains open. Ratings came from validated questionnaires or instruments grounded in prior work. More generally, cross-regional comparisons are difficult, as translated scales may lack measurement invariance, road rules and enforcement differ, color meanings and signage conventions vary, pedestrian norms and exposure to AVs are not comparable, and recruitment often yields mismatched samples.

This comparison is further confounded by site-specific factors. Japan used left-hand traffic, and Germany used right-hand traffic, which can change scanning behavior and attention allocation. Besides, the sample was relatively young, highly educated, and strongly interested in AVs, which limits generalizability to broader pedestrian populations. In particular, the findings may not transfer to older adults, people with lower educational attainment, people with lower familiarity with digital systems, or pedestrians with weaker interest in or more skeptical attitudes toward automated driving. The study should therefore be interpreted as evidence for this specific sample and scenario rather than as a population-level account of pedestrian preferences in Germany and Japan. We therefore interpret observed differences as region differences in this sample and scenario and do not attribute them uniquely to culture. Future work may hold traffic handedness constant across regions or add a condition that decouples handedness from country.

The study lacked a non-optimization control condition, limiting the extent to which effects can be attributed specifically to optimization rather than repeated exposure to the VR scenario. This limitation should be interpreted carefully. A core mechanism of HITL MOBO is repeated human feedback to refine the search toward user-relevant trade-offs over time [13, 76]. In that sense, participants are intended to familiarize themselves with the design, compare alternatives, and calibrate their preferences through repeated exposure to candidate designs. At the same time, repeated exposure may also increase familiarity with the VR scenario, the crossing task, or the rating procedure in ways that are not specific to the optimized designs. The present study, therefore, cannot fully separate intended preference refinement from more general scenario familiarization. Accordingly, our results should not be read as a causal estimate of the optimizer's effect on objective trajectories, but rather as evidence that, under the same HITL MOBO procedure, the resulting Pareto trade-offs and parameter patterns differed between the two regional samples. Future work should therefore strengthen causal interpretation by comparing HITL MOBO against non-adaptive baselines, such as random sampling, a predefined sequence of eHMI designs, or a separate first-exposure evaluation of the final optimized designs by new participants. Such baselines would help determine whether similar changes emerge without adaptive model-based selection, whether effects are partly driven by repeated exposure or by order, and whether optimized designs remain effective for non-participants in the HITL process.

The auditory message content was not localized across study contexts. While prior work has shown that auditory eHMI effects can vary with pedestrian characteristics and environmental conditions, such as hearing-related accessibility needs and background noise [96], the role of language localization in spoken eHMI messages remains underexplored. Future cross-region experiments should therefore localize spoken auditory messages to the study language and explicitly disentangle effects of auditory presence from effects of linguistic comprehension, for example, by comparing localized speech, non-localized speech, and non-verbal auditory cues. This limitation may have been attenuated in our relatively young and educated sample (Japan: on average 27 years old, all with a college degree; Germany: on average 25.4 years old, at least a secondary school degree), for whom comprehension barriers for a single English cue ("Stopping") were likely less pronounced.

In addition, future work should investigate eHMI designs across conditions, day time versus night time, varied weather, following Colley and Rukzio [28], and broaden the design space to include text, symbols, and dynamic animations. Prior studies indicate a strong potential for LED strip-based eHMIs due to ease of integration and aesthetic appeal. Text can be inaccessible to children or non-native speakers, and symbols can be ambiguous. We used a flashing animation because earlier work reported a preference for this pattern [36]. Cross-region experiments should use identical scenarios, calibrated luminance across lighting conditions, and locale-specific piloting to control for infrastructure, enforcement, crowd behavior, and language-related comprehension effects.

Finally, while VR provides a safe and highly controlled environment for evaluating pedestrian-vehicle interactions, it inherently limits ecological validity. Because participants know they are not in physical danger, they may behave more boldly or initiate crossings more quickly in VR than in the real world. Future work should complement VR experiments with real-world testing to verify whether these behavioral metrics hold under true physical risk. For instance, evaluating optimized eHMI designs on actual vehicles in controlled test-track settings. Furthermore, our simulated traffic scenario was intentionally simplified to establish a baseline, where the far lane contained only manually driven, non-stopping vehicles. Real-world urban environments are far more complex. Future studies must investigate multi-lane AV traffic to understand how pedestrians process potentially simultaneous, and perhaps conflicting, eHMI signals from multiple automated vehicles.

7 Conclusion

We ran a cross-regional HITL MOBO study of AV external communication in VR ($N=40$; Germany $n=20$, Japan $n=20$). The optimizer adjusted nine eHMI parameters to balance trust, visual appeal, mental demand, and time to start crossing. Optimization behaved as intended: trust and visual appeal increased over iterations, mental demand and time to start crossing decreased. On the Pareto front, objectives aligned in both regions, while time to start crossing was independent of them; across all observations in Japan, time to start crossing was independent of every objective. A dominance analysis with the additive ε indicator showed $\varepsilon(\text{Japan} \rightarrow \text{Germany})=0.333$ and $\varepsilon(\text{Germany} \rightarrow \text{Japan})=0.867$; $\Delta=-0.534$ was significant in a participant-clustered permutation test (5,000 permutations, $p<0.0002$). Process evaluations were positive in both regions, with no significant differences in expectation, satisfaction, agency, confidence, or ownership. These results support parameter search with HITL MOBO for eHMI design and reveal measurable regional differences in the joint objective space. Limitations include a single crossing scenario, a fixed objective set, and VR with left-hand traffic in Japan and right-hand traffic in Germany. Future work should expand the objectives, vary the scenarios and contexts, and test transfer to on-road studies.

Open Science

The Unity application will be made available upon request, and the collected (anonymized) data is available via <https://anonymous.4open.science/r/eHMI-Japan-Germany-HITL-MOBO-2026-5528/README.md> (currently anonymized).

Acknowledgments

We thank all study participants.

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] 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*. ACM, Honolulu HI USA, 1–14. doi:10.1145/3313831.3376197

- [5] 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
- [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*. Curran Associates, New York, NY, USA. <http://arxiv.org/abs/1910.06403>
- [7] Pavlo Bazilinskyy, Dimitra Dodou, and Joost De Winter. 2019. Survey on eHMI concepts: The effect of text, color, and perspective. *Transportation research part F: traffic psychology and behaviour* 67 (2019), 175–194. Publisher: Elsevier.
- [8] Mercedes Benz. 2022. The Mercedes-Benz F 015 Luxury in Motion. <https://www.mercedes-benz.com/en/innovation/autonomous/research-vehicle-f-015-luxury-in-motion/>
- [9] Ali Borji and Laurent Itti. 2013. Bayesian optimization explains human active search. *Advances in neural information processing systems* 26.
- [10] 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.
- [11] 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. arXiv:1012.2599 [cs]. <http://arxiv.org/abs/1012.2599>
- [12] Marc-Philipp Bockle, Anna Pernestål Brenden, Maria Klingegård, Azra Habibovic, and Martijn Bout. 2017. SAV2P: Exploring the Impact of an Interface for Shared Automated Vehicles on Pedestrians' Experience. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. ACM, Oldenburg Germany, 136–140. doi:10.1145/3131726.3131765
- [13] 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
- [14] 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
- [15] 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
- [16] 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
- [17] 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
- [18] 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 *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–11. doi:10.1145/3491102.3517571
- [19] 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>
- [20] 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.
- [21] Mark Colley, Julian Czymmek, Mustafa Küçükocak, 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
- [22] Mark Colley, Tim Fabian, and Enrico Rukzio. 2022. Investigating the Effects of External Communication and Automation Behavior on Manual Drivers at Intersections. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (Sept. 2022), 1–16. doi:10.1145/3546711
- [23] Mark Colley, Pascal Jansen, Mugdha Keskar, and Enrico Rukzio. 2025. Improving External Communication of Automated Vehicles Using Bayesian Optimization. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 80, 16 pages. doi:10.1145/3706598.3714187
- [24] 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

- [25] 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>
- [26] 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](https://doi.org/10.1145/3596249)
- [27] 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](https://doi.org/10.1145/3409120.3410646)
- [28] Mark Colley and Enrico Rukzio. 2020. A Design Space for External Communication of Autonomous Vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Virtual Event DC USA, 212–222. [doi:10.1145/3409120.3410646](https://doi.org/10.1145/3409120.3410646)
- [29] 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](https://doi.org/10.1145/3334480.3382844)
- [30] 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](https://doi.org/10.1145/3313831.3376472)
- [31] 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](https://doi.org/10.1145/3334480.3382865)
- [32] Joost de Winter and Dimitra Dodou. 2022. External human-machine interfaces: Gimmick or necessity? *Transportation Research Interdisciplinary Perspectives* 15 (2022), 100643. [doi:10.1016/j.trip.2022.100643](https://doi.org/10.1016/j.trip.2022.100643)
- [33] 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.
- [34] 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.
- [35] 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 (Sept. 2020), 100174. [doi:10.1016/j.trip.2020.100174](https://doi.org/10.1016/j.trip.2020.100174)
- [36] 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.
- [37] Debargha Dey, Kai Holländer, Melanie Berger, Berry Eggen, Marieke Martens, Bastian Pfleging, and Jacques Terken. 2020. Distance-Dependent EHMI 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](https://doi.org/10.1145/3409120.3410642)
- [38] 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](https://doi.org/10.1145/3239092.3265946)
- [39] 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*. ACM, Toronto ON Canada, 82–86. [doi:10.1145/3239092.3265946](https://doi.org/10.1145/3239092.3265946)
- [40] 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](https://doi.org/10.1145/3613904.3642031)
- [41] 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](https://doi.org/10.1145/3342197.3344523)

- [42] 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.
- [43] 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
- [44] 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
- [45] 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.
- [46] 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
- [47] Marie-Axelle Granié, Marjorie Pannetier, and Ludivine Guého. 2013. Developing a self-reporting method to measure pedestrian behaviors at all ages. *Accident Analysis & Prevention* 50 (2013), 830–839. doi:10.1016/j.aap.2012.07.009
- [48] 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
- [49] 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.
- [50] Lorena Hell, Janis Sprenger, Matthias Klusch, Yoshiyuki Kobayashi, and Christian Muller. 2021. Pedestrian Behavior in Japan and Germany: A Review. In *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, Nagoya, Japan, 1529–1536. doi:10.1109/IV48863.2021.9575809
- [51] Geert Hofstede. 2011. Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture* 2, 1 (Dec. 2011). doi:10.9707/2307-0919.1014
- [52] 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>
- [53] 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
- [54] 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
- [55] 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, Yokohama Japan, 1–13. doi:10.1145/3411764.3445480
- [56] 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
- [57] Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E Grinter. 2010. Postcolonial computing: a lens on design and development. In *Proceedings of the SIGCHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1311–1320.
- [58] Pascal Jansen, Mark Colley, Svenja Krauß, Daniel Hirschle, and Enrico Rukzio. 2025. OptiCarVis: Improving Automated Vehicle Functionality Visualizations Using Bayesian Optimization to Enhance User Experience. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 85, 23 pages. doi:10.1145/3706598.3713514
- [59] Philip Joisten, Ziyu Liu, Nina Theobald, Andreas Webler, and Bettina Abendroth. 2021. Communication of Automated Vehicles and Pedestrian Groups: An Intercultural Study on Pedestrians' Street Crossing Decisions. In *Mensch und Computer 2021*. ACM, Ingolstadt Germany, 49–53. doi:10.1145/3473856.3474004
- [60] 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
- [61] Joshua D Knowles, Lothar Thiele, and Eckart Zitzler. 2006. A tutorial on the performance assessment of stochastic multiobjective optimizers. *Tik report* 214 (2006).

- [62] 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.
- [63] 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
- [64] 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
- [65] 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*. ACM, Virtual Event DC USA, 122–131. doi:10.1145/3409120.3410653
- [66] Yang Li, Hao Cheng, Zhe Zeng, Hailong Liu, and Monika Sester. 2021. Autonomous Vehicles Drive into Shared Spaces: eHMI Design Concept Focusing on Vulnerable Road Users. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE, Indianapolis, IN, USA, 1729–1736. doi:10.1109/ITSC48978.2021.9564515
- [67] Yeti Li, Murat Dikmen, Thana G. Hussein, Yahui Wang, and Catherine Burns. 2018. To Cross or Not to Cross: Urgency-Based External Warning Displays on Autonomous Vehicles to Improve Pedestrian Crossing Safety. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, Toronto ON Canada, 188–197. doi:10.1145/3239060.3239082
- [68] 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>
- [69] Yi-Chi Liao, George B. Mo, John J. Dudley, Chun-Lien Cheng, Liwei Chan, Per Ola Kristensson, and Antti Oulasvirta. 2024. Practical approaches to group-level multi-objective Bayesian optimization in interaction technique design. *Collective Intelligence* 3, 1 (2024), 1–19. doi:10.1177/26339137241241313
- [70] 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
- [71] Victor Malmsten Lundgren, Azra Habibovic, Jonas Andersson, Tobias Lagström, Maria Nilsson, Anna Sirkka, Johan Fagerlön, 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.
- [72] Manuel López-Ibáñez. 2025. *moocore: Core Mathematical Functions for Multi-Objective Optimization*. R package version 0.1.8.900. <https://multi-objective.github.io/moocore/r/>
- [73] 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*. ACM, Montreal QC Canada, 1–12. doi:10.1145/3173574.3174003
- [74] 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
- [75] Olaf Mersmann. 2024. *emoa: Evolutionary Multiobjective Optimization Algorithms*. R package version 0.5-3. doi:10.32614/CRAN.package.emoa
- [76] George Mo, John Dudley, Liwei Chan, Yi-Chi Liao, Antti Oulasvirta, and Per Ola Kristensson. 2024. Cooperative Multi-Objective Bayesian Design Optimization. *ACM Trans. Interact. Intell. Syst.* 14, 2, Article 13 (June 2024), 28 pages. doi:10.1145/3657643
- [77] R. Bruce Money and John C. Crofts. 2003. The effect of uncertainty avoidance on information search, planning, and purchases of international travel vacations. *Tourism Management* 24, 2 (April 2003), 191–202. doi:10.1016/S0261-5177(02)00057-2
- [78] 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
- [79] 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*. ACM, Utrecht Netherlands, 284–294. doi:10.1145/3342197.3344543
- [80] Marie Pelé, Caroline Bellut, Elise Debergue, Charlotte Gauvin, Anne Jeanneret, Thibault Leclerc, Lucie Nicolas, Florence Pontier, Diorne Zausa, and Cédric Sueur. 2017. Cultural influence of social information use in pedestrian road-crossing behaviours. *Royal Society Open Science* 4, 2 (Feb. 2017), 160739. doi:10.1098/rsos.160739

- [81] Champika Ranasinghe, Kai Holländer, Rebecca Currano, David Sirkin, Dylan Moore, Stefan Schneegass, and Wendy Ju. 2020. Autonomous Vehicle-Pedestrian Interaction Across Cultures: Towards Designing Better External Human Machine Interfaces (eHMI). In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–8. doi:10.1145/3334480.3382957
- [82] 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.
- [83] Amir Rasouli and John K. Tsotsos. 2020. Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE Transactions on Intelligent Transportation Systems* 21, 3 (March 2020), 900–918. doi:10.1109/TITS.2019.2901817
- [84] Katharina Reinecke and Abraham Bernstein. 2013. Knowing what a user likes: A design science approach to interfaces that automatically adapt to culture. *Mis Quarterly* (2013), 427–453.
- [85] Markus Rothmüller, Pernille Holm Rasmussen, and Signe Alexandra Vendelbo-Larsen. 2018. Designing for Interactions with Automated Vehicles: Ethnography at the Boundary of Quantitative-Data-Driven Disciplines. *Ethnographic Praxis in Industry Conference Proceedings* 2018, 1 (Oct. 2018), 482–517. doi:10.1111/1559-8918.2018.01219
- [86] 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
- [87] Hatice Sahin, Heiko Mueller, Shadan Sadeghian, Debargha Dey, Andreas Löcken, Andrii Matvienko, 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
- [88] 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.
- [89] 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.
- [90] Gaye Solmaz, Derya Azık, Gizem Findık, Yeşim Üzümcüoğlu, Özlem Ersan, Bilgesu Kaçan, Türker Özkan, Timo Lajunen, Bahar Öz, Anton Pashkevich, Maria Pashkevich, Vassiliki Danelli-Mylona, Dimitra Georgogianni, Ema Berisha Krasniqi, Muhamed Krasniqi, Evangelos Makris, Ksenia Shubenkova, and Gentianë Xheladini. 2020. Cross-cultural differences in pedestrian behaviors in relation to values: A comparison of five countries. *Accident Analysis & Prevention* 138 (April 2020), 105459. doi:10.1016/j.aap.2020.105459
- [91] 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
- [92] Cédric Sueur, Barbara Class, Charlene Hamm, Xavier Meyer, and Marie Pelé. 2013. Different risk thresholds in pedestrian road crossing behaviour: A comparison of French and Japanese approaches. *Accident Analysis & Prevention* 58 (Sept. 2013), 59–63. doi:10.1016/j.aap.2013.04.027
- [93] Wilbert Tabone, Joost de Winter, Claudia Ackermann, Jonas Bärgman, Martin Baumann, Shuchisnigdha Deb, Colleen Emmenegger, Azra Habibovic, Marjan Hagenzieker, P.A. Hancock, Riender Happee, Josef Krems, John D. Lee, Marieke Martens, Natasha Merat, Don Norman, Thomas B. Sheridan, and Neville A. Stanton. 2021. Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives* 9 (March 2021), 100293. doi:10.1016/j.trip.2020.100293
- [94] Tram Thi Minh Tran, Callum Parker, Marius Hoggemüller, 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.
- [95] 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
- [96] Wenge Xu, Foroogh Hajiseyedjavadi, Debargha Dey, Tram Thi Minh Tran, and Mark Colley. 2026. Exploring the Impacts of Background Noise on Auditory Stimuli of Audio-Visual eHMIs for Hearing, Deaf, and Hard-of-Hearing People. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (*CHI '26*). Association for Computing Machinery, New York, NY, USA, Article 351, 16 pages. doi:10.1145/3772318.3791557
- [97] 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
- [98] 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
- [99] Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M Fonseca, and Viviane Grunert Da Fonseca. 2003. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on evolutionary computation* 7, 2 (2003), 117–132.

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.