Modeling Human-Drone Proxemics in an Augmented Reality Environment

Master's Thesis in Computer Science and Engineering

Vilmer Malm & Hugo Raimer

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2026

Master's Thesis 2026

Modeling Human-Drone Proxemics in an Augmented Reality Environment

Vilmer Malm & Hugo Raimer

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2026

Modeling Human-Drone Proxemics in an Augmented Reality Environment
Vilmer Malm & Hugo Raimer

Supervisor: Dr. Mohammad Obaid, Department of Computer Science and Engineering
External Supervisor: Dr. Fanta Camara, Institute for Safe Autonomy, University of York
Examiner: Prof. Morten Fjeld, Department of Computer Science and Engineering

Master's Thesis 2026
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2026

Vilmer Malm and Hugo Raimer
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

As drones increasingly move into human-occupied spaces, they must be able to navigate in ways that are not only safe but also socially acceptable. This thesis investigates human-drone proxemics in an augmented reality (AR) environment, with the aim of modeling the interpersonal distances people prefer when approached by a social drone. Building on prior two-dimensional rubber-sheet models from human-robot proxemics, this work extends the approach into a three-dimensional "rubber-bubble" model suitable for drones. A user study was conducted in Sweden and Vietnam ($N = 69$), where participants interacted with a virtual drone approaching from 21 different angles in an area in front of them. Participants indicated when the drone reached their preferred stopping distance, producing a dataset used to evaluate approach-angle effects, demographic variables as predictors, cultural differences, and various degrees of personalization. The methodology combined statistical testing, polynomial rubber-bubble models, linear mixed models, and Bayesian modeling for an exhaustive and exploratory analysis. The results show that rubber-bubble models can represent individualized proxemic preferences within the observed frontal zone, suggesting usefulness for autonomous drone navigation. Gender and culture both showed potential in predicting stopping distances. In summary, this work contributes with an extension of 2D rubber-sheet to 3D rubber-bubble modeling, a thorough analysis of demographics and culture as potential predictors for human-drone proxemics, and an open-source AR environment for safe and repeatable human-drone research.

Keywords: Human-Drone Interaction, Human-Drone Proxemics, Social Drones, Augmented Reality, Proxemic Modeling, Autonomous Drone Navigation, Rubber-Bubble Model, Three-Dimensional Proxemics

Acknowledgements

We would like to sincerely thank Dr. Mohammad Obaid and Dr. Fanta Camara for their guidance, support and encouragement throughout this thesis.

We would also like to thank our examiner, Prof. Morten Fjeld, for his feedback and support during this work.

We are very grateful to VNUHCM - University of Science and Dr. Khanh-Duy Le for welcoming and accommodating us during our time in Vietnam and for enabling us to conduct the study there. We would also like to thank Duy's students Diep Phi Vu Vo and Dinh-Thuan Duong-Le for their help and support throughout our stay.

We also want to extend our gratitude to Prof. Richard Torkar for his help in the Bayesian modeling and analysis.

A special thank you to Javier Jiménez Ruescas for helping us with the Unity scene and providing the drone model and other assets that sped up this study significantly.

We are also grateful to Fuse and T-Rax for letting us use their space during our study in Sweden, and especially to Malin Anker for all her help and coordination throughout the process.

A big thank you, as well, to Moa Jönsson for assisting with graphical design for illustrations used in the thesis.

Last, but not least, we would of course like to thank all participants who took part in the study and contributed their time and data to this research.

Vilmer Malm and Hugo Raimer, Gothenburg, June 2026

Declaration of Generative AI

Generative AI tools were used to improve grammar and vocabulary in the text. They were not used to generate original content. They were also used for suggestion of code improvements and documentation. All changes were reviewed and full responsibility for the final work is taken by the authors.

Vilmer Malm and Hugo Raimer, Gothenburg, June 2026

Contents

List of Figures

List of Tables

Glossary

Cartesian Coordinates A coordinate representation using orthogonal x-, y-, and z-axes to describe spatial positions in three-dimensional space. 23, 25, 26

Intra-class Correlation Coefficient A statistical measure used to determine how much of the total variation in a dataset is explained by differences between groups or individuals, indicating the consistency of measurements within the same subject. 36

Linear Mixed Models An extension of linear regression that include both fixed effects, such as demographics and approach angles in this study, and random effects, such as participant-level variation. 36

Proxemics The study of how humans perceive and use physical space to regulate social interaction to ensure comfort and safety. 1, 5

Spherical Coordinates A three-dimensional coordinate system in which positions are represented using a radial distance and angular components. 26

Wizard-of-Oz A research methodology where a participant interacts with a system they believe to be autonomous, while it is actually being operated or controlled by a human researcher. This approach is used to observe authentic user behavior and responses without the technical constraints or safety risks of full automation. xv, 9, 13, 14

List of Abbreviations

AR Augmented Reality. xiii, 2, 3, 11–14, 17, 18, 20, 21, 90, 92, 95

HAI Human-Agent Interaction. 8
HDI Human-Drone Interaction. 1–3, 9, 21, 83, 85
HRI Human-Robot Interaction. 1, 9, 83

ICC Intra-class Corellation Coefficient. 36, 67, 80, 96
IPD Interpersonal Distance. 2, 5, 6, 9, 33, 47, 80, 82, 83, 85

LMM Linear Mixed Models. xviii, 27, 35, 47, 67, 68, 80, 81, 83, 84, 92, 96
LOSOCV Leave-One-Subject-Out Cross-Validation. xii, xvii, xviii, 29, 32, 33, 35, 45, 62, 63, 65–67, 71, 72, 80, 84, 95

OLS Ordinary Least Squares. xvii, xviii, 32, 60, 62–66

RMSE Root Mean Squared Error. 34, 35, 66

UAV Unmanned Aerial Vehicle. 7

1

Introduction & Background

The emergence of social drones reflects a broader shift in robotics from industrial robots to agents that operate in close proximity to humans in everyday environments. While drones have traditionally been used in military, industrial, and logistical settings [36], recent research increasingly explores their role in domestic, service, and companion-oriented applications [37]. These social drones are autonomous aerial robots designed to operate in inhabited environments such as homes and cities [7], and perform supporting tasks such as delivery, navigation assistance, support well-being, and accessibility [54]. Unlike ground-based robots, drones can maneuver above obstacles and crowds, remain visible from a distance, and approach users from multiple vertical angles, which expands their interactive capabilities but also introduces new spatial considerations [71].

As drones begin to operate in populated environments, they must maintain socially acceptable spatial boundaries when approaching humans. The theory of proxemics, originally introduced by Hall [21], describes how humans regulate interpersonal distance as part of social interaction. Prior research in Human-Robot Interaction (HRI) shows that distance strongly affects a person's comfort levels around robots and affect how the robots are perceived [65]. Evidence suggests a preferred social distance of 50–100 cm is common for humans engaging with smaller (121 cm) humanoid robots [49]. Although drones can maintain an excessive distance to avoid unnecessary discomfort, overcompensating the distance maintained can lead to 'freezing,' where the system fails to reach its destination in dense environments [64]. Determining the appropriate approach distances is therefore essential to enable safe, timely, and socially acceptable drone behavior [9].

A multitude of factors impact human proxemic preferences toward robots and drones. These preferences are influenced by an agent's perceived emotions and social cues, such as gaze, which have been shown to affect interpersonal distancing in both robot and drone contexts [46], [8]. Kim and Mutlu [38] demonstrated that a robot's role significantly influences proxemic preferences, noting, for example, that these distances shift based on whether the robot is cooperative or competitive. Preferences also vary depending on whether the robot is perceived as autonomous or teleoperated [62]. Furthermore, personal characteristics such as age, height, gender, and culture have been shown to influence preferred interaction distances across HRI and Human-Drone Interaction (HDI) studies [46], [35], [71], [41]. These findings highlight the need to account for both drone-related and human-related variability when modeling human-drone proximity. While early HDI studies have identified preferred interaction distances between humans and drones [70], [71], the field still lacks a generalized model that explains how drone-related parameters such as approach angles, together with human characteristics, determine preferred interpersonal distance in three-dimensional space.

To enable drones to maintain socially acceptable distances in real-world operation, autonomous systems require predictive models capable of estimating preferred human-drone proximity under varying interaction conditions. Prior HRI research has proposed mathematical representations of personal space, including rubber-sheet transformation approaches that model proxemic boundaries as continuous and dynamically deformable based on interaction parameters [12]. While these approaches provide a conceptual and mathematical foundation, extending them to aerial robots introduces additional complexity due to three-dimensional movement and thus more approach trajectories, which must be incorporated into the distance prediction models. These rubber-sheet [12] models could be used to personalize proxemics by modeling proxemic preferences by utilizing a person's demographics or other variables.

The purpose of this study is to investigate how humans establish comfort boundaries when approached by an autonomous aerial drone. The study aims to quantify how approach angles together with selected human-related characteristics such as age and height, influence acceptable human-drone social distances and subsequently develop a predictive model that can automatically determine the optimal distance that a social drone should keep from a person.

This work aims to address the following research questions:

- **RQ 1**: What rubber-sheet model adapts well to human-drone proxemics in a 3D space?
- **RQ 2**: What human demographics most strongly influence proxemics during HDI?
  - **RQ 2.1**: What is the impact of individualist and collectivist cultures on preferred human-drone Interpersonal Distance (IPD)?

This study focuses specifically on spatial interaction by investigating human-drone proxemics in a controlled AR environment. The goal is to generate proxemics data and use it to develop a model capable of predicting acceptable HDI distances across varying drone approach angles for different users.

This study demonstrates how AR can be used as a controlled platform for collecting high-precision proxemics data in three-dimensional interaction scenarios. Compared to physical drone experiments, AR enables control of spatial parameters while ensuring participant safety and experimental repeatability. AR environments have also been shown to provide high spatial realism and strong external validity in proxemics research [33], [26].

From a practical perspective, the results could support the development of autonomous drones that can operate safely and acceptably around humans.

Based on the empirical findings and analytical frameworks evaluated in this study, the primary research contributions to the field of human-drone proxemics are summarized as follows:

- Demonstrates how rubber-sheet equations can extend into three dimensional rubber-bubble equations for use in human-drone proxemics.
- Estimates the predictive value of various parameters, such as age, gender, and height with statistical analysis, rubber-bubble models, linear mixed models, and Bayesian analysis.
- Proposes a simplified, zone-based mathematical formulation to reduce parameter complexity in areas with high angular correlation.
- Provides a cross-cultural comparison between individualist (Swedish) and collectivist (Vietnamese) cultures to examine how human-drone proxemics differ between them.
- Provides a Bayesian approach to HDI research.
- Provides an open-source AR environment to facilitate and effectivize future HDI research.

2

Related Work

The deployment of autonomous drones in human-occupied environments requires a thorough understanding of human spatial psychology. This chapter establishes the theoretical foundations of this thesis by reviewing the evolution of spatial boundary theory, tracing its trajectory from foundational human-human psychology to its modern applications in robotics.

2.1 From Human-Human to Human-Drone Proxemics

Proxemics is defined as the unconscious physical distances humans maintain during social interactions. Originally introduced by Hall, this framework suggests that humans regulate distance to manage social comfort. Hall's foundational work identified four discrete, concentric horizontal zones: intimate (< 0.45 m), personal ($0.45 - 1.2$ m), social (1.2˘3.6 m), and public (> 3.6 m), which serve as a baseline for understanding human spatial expectations [21], [22].

This concept has since been expanded to the personal space people keep to robots [59], [62], [46], [43]. People tend to treat robots as social agents, applying human-like spatial boundaries to them [62], including an automatic and instinctive response to robots violating said boundaries [9].

However, a robot's status as a non-human agent introduces unique dynamics, as they are often viewed along a spectrum between social entities and functional tools [62]. Research has noted that humans generally maintain smaller proxemic zones toward robots than they do toward other humans. While a human tends to keep an IPD of around 1.0 meter when talking to a stranger [25], it has been found that one prefers to keep a 121.0 cm tall, ground-based robot, at a distance of 0.5 - 1.0 m [49].

Human-robot proxemic zones are dynamically affected by the motion of the robot, e.g. approach angle and velocity [55, 50], as well as physical characteristics of the robot such as height and size [55]. They are further mediated by the individual characteristics of the human participant. Physical factors such as gender, age and height shape the required size of a person's proxemic zone [55], [62], [12], and even factors such as cultural background and previous experience with robots or pets affect zone size [62], [58], [48].

The three-dimensional operational space of drones introduces variables not present in HRI modeling research, such as the vertical approach angle [70, 9]. Previous proxemic studies have hypothesized that the 3D proxemic spatial zone around a person reflects a bubble in shape, meaning that the IPD should vary at different heights [21, 70]. Furthermore, the ability to fly is facilitated through the drone's propellers, which themselves impact the proxemic zones in other ways than ground robots do through their high-pitched noise, airflow and intimidating appearance [8], [71].

In their early investigation of drone proxemics, Duncan and Murphy examined interpersonal distances by having a small unmanned aerial vehicle approach participants at heights of 1.5 m and 2.1 m [15]. The participants, who averaged 1.7 m in height, indicated when the drone reached a minimum comfortable distance. The researchers set the closest possible distance to 0.6 m. They found that the comfortable distance ranged from 0.6 m - 0.9 m. Notably, the study found no significant difference in comfort levels based on whether the drone was positioned above or below eye level, a result that contradicts the bubble proxemic zone. Wögerbauer et al. [70] suggest that this lack of significance may be attributed to a high perceived safety level, as the drone was mounted on a moving platform rather than flying freely, or perhaps the variation in height was insufficient.

In contrast, the study conducted by Wögerbauer et al. demonstrated a significant height effect, revealing that the comfort zone boundary is C-shaped [70], see Figure 2.1. Within this C-shaped configuration, participants preferred the drone further away the higher and the lower it was positioned, also contradicting the bubble theory by inverting it. Users generally allowed the drone to come the closest at approximately 50% of their eye level, resembling a relaxed grasping position. They found the mean preferred distance to the drone at all angles to be 1.62 m.

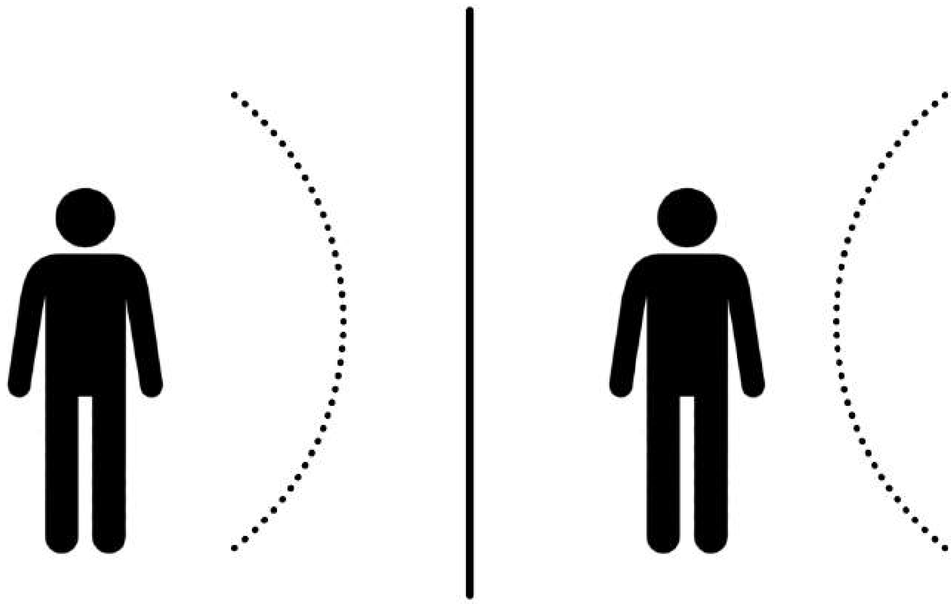


Figure 2.1: Schematic illustration of bubble (left) and C-shaped (right) proxemic zones.

Bretin et al. [9] showed that the purpose or "framing" of a drone is another factor that affects proxemic zones. They tested this by presenting the drone to people in two different ways. One group was given a social frame, where the drone was called "Happy" and described with pet metaphors. The other group was given a technical frame, where it was described as a Unmanned Aerial Vehicle (UAV) using reinforcement learning. The researchers expected the social frame to make people more comfortable and lead to closer distances, but the results showed no statistically significant effect. In fact, there was a trend where the social group actually stayed further away from the drone. This likely happened because of cognitive dissonance: the participants were told the drone was a friendly companion, but they were looking at a standard, noisy mechanical machine. The study mentions that simply providing a social description may be insufficient to make a drone feel social to a user. If the physical design and behavior of the drone do not match the framed expectations, the framing can act as a double-edged sword, increasing user distancing.

2.2 Modeling Proxemic Behaviors

The field has shifted away from Hall's discrete circular zones toward more sophisticated representations. For example, recent research has demonstrated that non-circular egg-shaped proxemic zones can be derived directly from human kinematics and the time required to avoid potential collisions. By modeling these boundaries as a function of an agent's heading and speed, researchers have established a mathematical link between physical trust and spatial behavior [11].

There are various other ways of measuring proxemics in Human-Agent Interaction (HAI) through mathematical modeling as well as deep learning methods. Researchers have employed reinforcement learning to adapt distances based on user body signals to estimate interaction success probabilities. Supervised deep learning has been applied to address human factors variability. For example, long-short-term-memory (LSTM)-based models have proven to be effective, as their ability to capture temporal dependencies allows them to accurately predict continuous probability distributions for comfortable stopping distances [20]. Although deep learning approaches show promise, they often require massive datasets and complex pre-processing steps to avoid sharp transitions in the predicted comfort zones when facing unseen approach angles [20].

To address the need for a robust, continuous boundary that functions effectively without the need for synthetic data, this study utilizes a mathematical modeling approach. One of the earliest mathematical approaches was done by Kirby et al. [39], who introduced the COMPANION framework for socially aware robot navigation. In this model, personal space was represented using two asymmetric 2D Gaussian halves, creating a larger safety region in front of a person. In another area, Camara and Fox [10] developed a method to infer proxemic utility functions rather than traditional bubbles zones. By applying Bayesian inference to pedestrian-vehicle interactions, they evaluated polynomial, Gaussian, and hyperbolic functions to quantify the psychological discomfort that occurs when approached by a machine, finding that a hyperbolic function best captures the dynamic "trust zones" required for safe interaction.

Further research by Neggers et al. [49] measured perceived comfort in relation to robot distance and passing-side through a fitted inverted Gaussian model. More recently, Camara et al. [12] has utilized rubber-sheet models to represent flexible, personalized boundaries. These models allow for the deformation of proxemic zones to account for varied comfort distances across different interaction parameters, making them a fitting choice for this study. To account for human variability they collected various characteristics of the subjects including: age, gender, height, experience with robots, pet ownership, educational background, profession, and dominant hand. They then performed a correlation analysis of the traits with the preferred proxemic distance with a goal to include only the most important variables in the final model. They found for example that shorter female participants have a smaller proxemic zone in front of them than taller female participants did and this was incorporated into the model. This modeling approach is described further in Chapter 5. When analyzing statistical methods for HCI, Robertson and Kaptein [57] mean that there is no best method to analyze data, suggesting that multiple statistical perspectives should be actively considered and combined. Exploring these boundaries through a multi-faceted analytical lens supports comprehensive and fair statistical communication within the field.

2.3 Cultural Dimensions in Proxemics

Having participants from multiple cultures requires an examination of the cultural impacts on human-robot proxemics. While the specific cultures in this study (Swedish and Vietnamese) have not been compared before in the context of HRI or HDI proxemics, existing research has established the significance of culture in HRI [5], [52]. For instance, Eresha et al. [18] demonstrated that cultural differences significantly dictate preferred proxemic distances between humans and robots. Under established proxemic theory, both Northern Europeans and Asians are classified as low-contact cultures, generally preferring larger interpersonal distances than Mediterranean or Arab cultures [23], [21]. However, Swedish and Vietnamese cultures diverge along the dimension of individualism and collectivism coined by Hofstede [28].

In Hofstede's work on cultural differences he found that people from collectivist cultures, such as Vietnam, rely on close intra-group relationships, whereas people from individualist cultures, such as Sweden, are highly independent and maintain strong feelings of autonomy within a group [29]. In a spatial context, these dimensions often translate to differences in IPD. In individualistic cultures individuals are described as preferring to stand apart to preserve their personal autonomy [31]. Those with collectivist backgrounds tend to be physically close to one another, as the cohesiveness of the in-group is prioritized over individual space. Although, collectivist cultures often keep further distance to people not in the in-group [31] which may translate to keeping further distances from unknown agents such as drones. Eresha et al. [17] noted differences in human-robot proxemics, finding that participants from Arab (collectivist) cultures preferred closer IPD than German (individualistic) participants did. Jane et al. [41] conducted a cross cultural study on human drone interaction comparing participants from China, a collectivist culture, and the United States, an individualistic culture. While some differences in preferred interaction distances were observed, participants from both cultures generally accepted the drone into their personal space and anthropomorphized it. Chinese participants were somewhat more likely to allow the drone within intimate distance, but the study overall highlighted similarities in proxemic behavior across cultures. They suggest that although cultural background influences HDI in other ways, proxemic preferences may be similar across cultures.

Conducting cross-cultural research within the participants' native environments is essential for maintaining ecological validity, as it avoids the confounding variables often associated with international student populations who may be undergoing cultural adaptation. Joosse et al. [34] emphasize that observing participants in their home countries ensures a more authentic representation of cultural norms compared to studies involving expats. Furthermore, the use of a Wizard-of-Oz approach provides a robust framework for investigating proxemics [34].

2.4 Research Gap

To the best of our knowledge, research on human-drone proxemics is limited and previous research does not explore a way to mathematically model autonomous drone interaction distances based on approach angles combined with human-characteristics. There is also limited research in evaluating the human-drone proxemic predictive power of the objective parameters used in this study such as culture, age, gender, and height.

3

User Study

This chapter describes the user study procedure and data collection methodology.

3.1 Ethics

The study was conducted in accordance with established ethical guidelines for research involving human participants. Informed consent was obtained from all participants before the experiment began.

Participation in the study was voluntary. Participants were informed that they could withdraw from the experiment at any time without providing a reason and without any negative consequences. They were also informed that their data could be removed upon request.

The experiment was conducted in a controlled environment using an AR setup in which participants interacted with a simulated drone. This approach enabled the study of human drone proxemics while avoiding the physical safety risks associated with operating a real drone in close proximity to participants. The procedure therefore posed minimal risk beyond ordinary everyday activities.

The data collected during the study consisted of participants' preferred interaction distances to the drone under different approach conditions, together with basic demographic and other information as described in Sec. 3.5. No personally identifying information was stored with the experimental data. Each participant was assigned an anonymous identifier to ensure confidentiality.

All collected data was stored securely and used exclusively for academic research purposes. Results are reported only in aggregated and anonymized form to ensure that individual participants cannot be identified.

3.2 Participant Recruitment

In Sweden, 43 people were recruited for the study, using convenience sampling. This resulted in mostly friends, family, professors, and fellow students participating. They received no compensation for volunteering. In the end, 38 of them identified as culturally Swedish and were used for the Swedish sample.

In Vietnam, 35 people were recruited for the study. In order to more quickly gather participants during the one week study, participants were offered compensation of 50,000 Vietnamese dong (approximately 1.9 USD at the time of writing) but only 14 of them accepted it.

Before the study began, all participants signed a research consent form (see Appendix F) agreeing to participate in the study and allowing their data to be used for the research.

3.3 Experimental Setup

This experiment tests the user's distance preferences to an approaching drone at various heights and approach angles. The user study was conducted in an AR environment where a virtual DJI Tello drone (see Figure 3.3 for reference) approached users from 21 different angles with the same speed of 0.445 m/s. The vertical and horizontal angles changed with 15° intervals. The vertical angles included -15°, 0°, and 15°, where 0° represents eye-level relative to the participant. The horizontal angles ranged from 45-135°, where 0° is directly to the left and 90° a direct approach from the front. This results in 7 horizontal angles and thus 21 approach angles in total (see Figure 3.1). The culmination of these angles are referred to as the "frontal zone" covering the 90° horizontal space in front of the user within a vertical of 30° (see Figure 3.2 for a visualization). The speed and size of the drone were constant as to not introduce too many variables, putting a larger focus on the relevant human characteristics.

The stopping distances in this study were recorded from the anchor point of the virtual drone model which was placed in the middle of the model. The drone was 22 cm long meaning that the outmost point of the drone (i.e. the edge of the rotor guards) were 11 cm from the anchor point. To get the distance of how close the drone is to a participant, one must subtract 11 cm from the recorded stopping distances. The distances in this study are always referred to from the anchor point if not stated otherwise.

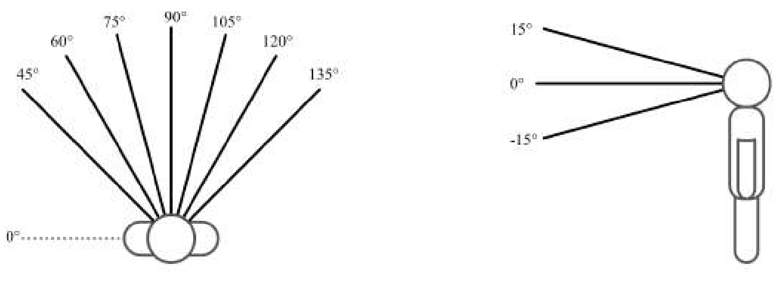


Figure 3.1: An illustration of the angular range of the experiment (horizontal angles to the left; vertical angles to the right).

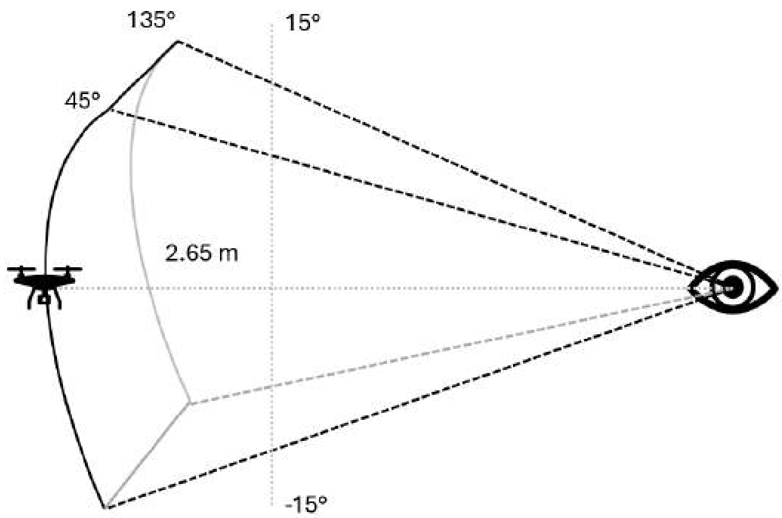


Figure 3.2: A schematic illustration of the frontal zone, showcasing the range of approach angles included in the trials. It stretches from -15° to 15° vertically, where 0° is eye level, and from 45° to 135° horizontally, where 0° is directly to the left of the participant.

Figure 3.3: The 3D drone model used during the experiment.

Participants were informed that they would interact with a social drone in an AR environment. The drone was framed as a helpful companion, and participants were instructed to perceive it as a supportive agent on its way to interact with them in some way. In Vietnam, participants were offered written instructions in Vietnamese (see Appendix G for reference), along with the same spoken instructions given to Swedish participants. For the participant, the drone is perceived as fully autonomous as a Wizard-of-Oz approach was utilized (see Figure 3.4 for reference). This resulted in users feeling like they were speaking to a voice controlled autonomous drone instead of to the researchers, aided by the fact that the researchers left the room during the study in Sweden. In Vietnam the conditions made it so that the researchers had to stay in the room but as the room was very spacious one could stand in the corner of to make the participant feel more alone with the drone.

Figure 3.4: Recreation of the Wizard-of-Oz approach in Sweden.

In Sweden, the room where the study was conducted measured 4.4 × 3.6 meters and was largely empty, except for two tables between which the participant was positioned (see Figure 3.5a). In Vietnam, the room measured 7.8 x 6.4 meters (see Figure 3.6). Two tape markings were placed on the floor (see Figure 3.5b). In Sweden, the first tape marking was located 2.2 m from the short (side) walls and 0.8 m from the long wall by the table. In Vietnam, it was placed 3.2 meters from the long (side) walls and 2.8 meters from the short wall cabinets so that the participants had the roughly same distance to the wall that they were facing in both countries (see Figure 3.6 for reference). A second tape marking was placed 0.4 m in front of the first. During the study, participants were told to stand on the first marking (see Figures 3.5a and 3.5b for reference). In Sweden, windows were closed off with a curtain during testing so the participant could not be influenced by seeing another person (see Figure 3.7 for reference). Participants were then equipped with a Meta Quest 3 headset, with the AR application pre-loaded and running.

Due to scheduling constraints during the research visit to Vietnam, the Vietnamese data collection was conducted in two separate experiment rooms, as the initial room became unavailable during the visit period. The experimental setup was reproduced as consistently as possible between locations to minimize potential environmental variation. In this second room, four participants participated, one of whom had to be removed for data integrity. For reference, see Figure 3.8 for the second room.

(a) Recreation of a participant in the Swedish experiment room.

(b) Tape markings on the floor in the Swedish experiment room, as seen from a participant.

Figure 3.5: Experimental setup and floor markings in Sweden.

Figure 3.6: Recreation of a participant in the initial Vietnam experiment room.

Figure 3.7: Curtains closed in the Swedish experiment room.

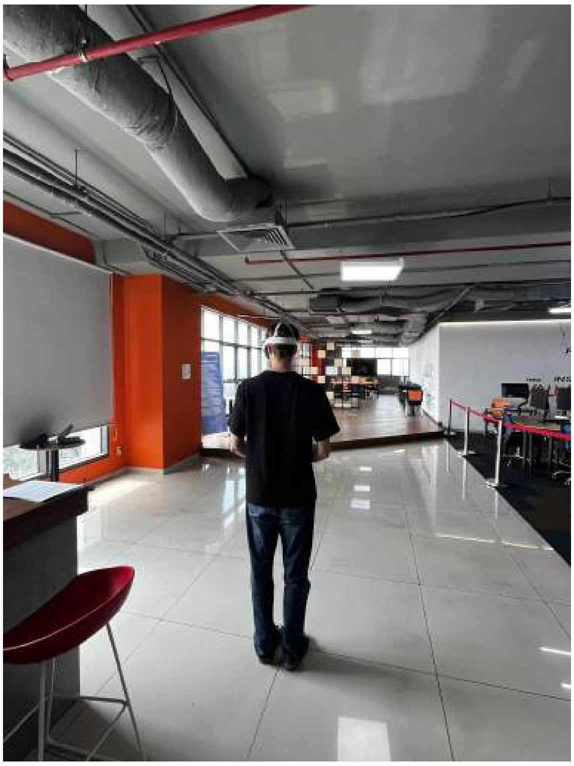


Figure 3.8: Recreation of a participant in the secondary Vietnam experiment room.

3.4 Running the Study

As the researchers exited the room (or relocated to the corner) a short warm-up phase was conducted in which the drone took-off 1.0 meter in front of the participant, moved 1.8 meters to the right and to the left of the participant. Then the drone landed in its initial position. This phase allowed participants to become familiar with the AR environment and the drone's movement and sound before the experiment began, reducing potential novelty effects. After this warm-up phase, participants were asked if they were ready to begin the experiment.

The experiment was initiated by the drone taking off and moving to one of the predefined starting points of an approach angle. The order of the angle of approach was randomized to reduce potential ordering effects, such as participants becoming more accustomed to the AR environment over time.

Once the drone reached its starting position and was oriented toward the participant, participants were instructed to say "start." The drone then began approaching the participant along a linear trajectory directed toward their eyes. Participants were asked to say "stop" at the point where any further approach would result in discomfort. The drone would then be stopped immediately by the researchers.

Two seconds after the drone stopped, an auditory signal, a "beep," was played. Upon hearing this signal, participants were instructed to look down at the marked position on the floor in front of them to avoid observing the drone repositioning between trials (see Figure 3.9 for reference). This was done to reduce potential anchoring effects from previous stopping distances. The delay was added so that the participant would have time to assess the stopping distance they chose. Once the drone had reached its next starting position, a second auditory signal prompted participants to look up and say "start" again.

Figure 3.9: Recreation of a participant looking down during drone relocation between trials.

This procedure was repeated for all 21 approach angles for each participant.

Throughout the experiment the researchers monitored the participant's view by casting the headset display onto a mobile device using the Meta Horizon iOS app (see Figure 3.4 for reference). This allowed them to ensure that instructions were followed correctly and to provide guidance if necessary.

After completing all trials, participants filled out a questionnaire consisting of three parts: a demographic section capturing relevant individual characteristics, a post-study section aimed at understanding the reasoning behind their stopping distances, and an AR comfort assessment. The latter was included to evaluate whether factors such as discomfort or nausea could have influenced the recorded distances.

The experimental procedure was conducted in a controlled environment and followed a structured sequence to ensure data consistency. The total experiment time for one participant was approximately 20-25 minutes.

(a) **Greetings and Consent (5 minutes):** Upon arrival, participants were welcomed and provided with a consent form.

(b) **Instructions (5 minutes):** After the signing of the consent form the participants were given an explanation and instructions for the test in verbal form (as well as written in Vietnam). The study's objectives regarding social drone proxemics were explained through a standardized script to ensure all participants received identical information regarding the experimental goals (see Appendix G for reference).

(c) **AR Familiarization and Baseline (1 minute):** To mitigate the novelty effect of the AR environment, participants were given a one-minute warm-up to observe the environment and the drone.

(d) **Instructions repeated after familiarization (1 minute):** The participant was asked if they were ready to start the test. If they were, main instructions were repeated again to make sure participants understood correctly and then they were given the opportunity to ask any final questions.

(e) **Experimental Trials (7 minutes):** The data collection phase consisted of repeated drone approaches at varying approach angles, as described in Sec. 3.3.

(f) **Post-Study Data Collection (5 minutes):** Following the trials, participants completed a digital form. This single step integrated the collection of demographic data (age, gender, height) with a questionnaire regarding subjective comfort levels and previous drone / VR experience, as well as their comfort level in the AR environment.

3.5 Questionnaires

All scale-based questions were answered using a 7-point Likert scale, where 1 corresponds to strongly disagree and 7 corresponds to strongly agree.

3.5.1 Demographic Questionnaire

The demographic questionnaire collected the following participant characteristics:

- Age
- Gender
- Height
- Dominant hand
- Profession
- Cultural identity (Optional)
- Country of residence
- Experience with free-roaming pets (e.g., dogs or cats):
  - Have pets
  - Have had pets
  - Never had pets

The question regarding cultural identity was asked in order to better capture the participants "subjective belief in their common descent." As Weber [68, p. 389] argues, this shared belief is the defining feature of an ethnic group.

3.5.2 Post-Study Questionnaire

The post-study questionnaire captured participants' prior experience and perceptions of the drone interaction:

- Previous experience with VR (scale)
- Previous experience with drones (scale)
- I was comfortable with the drone speed (scale)
- I was comfortable with the drone size (scale)
- The drone's sound impacted my decision (scale)

- The drone's appearance made me feel at ease (scale)
- The drone's movements were smooth and natural (scale)
- The drone felt dangerous (scale)
- I trusted the drone to behave safely at all times (scale)
- I could easily understand where the drone was going (scale)
- Overall, I felt relaxed during the interactions with the drone (scale)
- What aspects of the drone's behavior made you feel most comfortable? (Open-ended)
- What aspects of the drone's behavior made you feel most uncomfortable? (Open-ended)

3.5.3 Augmented Reality Comfort

To assess the validity of the collected data, participants' comfort in the AR environment was evaluated:

- I was comfortable in the AR environment (scale)
- Did you experience any discomfort in the AR environment?
  - None
  - Dizziness
  - Nausea
  - Eye strain
  - Other (open-ended)

4

Open-Source AR Environment Setup

The following section outlines how to navigate the AR application and how it is set up. It is meant to guide researchers in utilizing the setup for future HDI research.

The experiment application was developed in Unity 6 and targets the Meta Quest 3 headset, using the Meta XR SDK (OVR) for headset tracking and controller input. The application renders a virtual drone overlaid on the participant's real-world environment using the headset's passthrough camera feed.

4.1 Using and Extending the AR Environment

The `Unity_project` folder in the study's **GitHub repository** contains the files required to open and modify the AR application (`Assets`, `manifest.json` and `ProjectSettings`). The project can be used to reproduce the AR environment described in this study or as a basis for further development.

To set up the Unity project, follow the steps below:

1. Download `Unity_project` from the **GitHub repository**.
2. Open Unity Hub.
3. Select `Add` and then `Add project from disk`. Choose the `Unity_project` folder.
4. If Unity version `6000.3.5f2` is not installed, Unity Hub will warn about a version mismatch and prompt the user to install it. The project should preferably be opened using version `6000.3.5f2`, as this was the version used during development. Using a different version may result in broken scripts, missing APIs, or package incompatibilities.
5. Click `Open`. The initial loading process may take several minutes.
6. Follow **Meta's Unity setup guide for Meta Quest development** to en-

sure that the required packages and build settings are installed and used.

7. Open the `TESTVR.unity` scene from the `Unity_project` folder.

4.2 Running the Application on the Headset

To build and run the application on a Meta Quest headset, the Meta Quest build platform must first be selected in Unity's Build Profiles. This process is described in **Meta's guide for selecting the Meta Quest build platform**.

After the correct build platform has been selected, the headset should be connected to the development computer using a USB-C data cable. Developer mode must also be enabled on the headset, and USB debugging must be allowed from within the headset when prompted. Meta's guide describes this process in detail in **the section on enabling developer mode**.

Once the headset is connected and detected by Unity, the application can be deployed by selecting `Build and Run`.

4.3 Extracting Experimental Data from the Headset

Android Debug Bridge (ADB) is used to extract the logged data from the headset after the experiment has been completed. ADB must first be downloaded and installed on the development computer. After installation, the data can be retrieved as follows:

1. Connect the headset to the computer using a USB-C data cable and allow USB debugging from within the headset.
2. Open File Explorer and navigate to the folder where ADB was extracted, for example:

   `C:\platform-tools`

3. Click the address bar in File Explorer, type `cmd`, and press Enter. This opens a terminal window in the ADB directory.
4. Copy, paste, and run the following command in the terminal, replacing `<destination>` with the local folder where the extracted data should be saved:

```
adb pull /sdcard/Android/data/com.UnityTechnologies.com.unity.
    template.urpblank/files/Data/<destination>
```

This command copies the `Data` folder from the application's storage directory on the headset to the selected destination on the computer.

4.4 Experiment Orchestration

The central script for the entire application is `SC_RouteController`, which governs the full lifecycle of a session. The headset camera is used as the anchor which serves as the reference for all subsequent drone positioning, so `PlaceDroneAtFloor()` is used to place the drone at the floor, using the participants height (`userHeightCm`) and offsetting it to account the difference in height and eye-level of a person. `userHeightCm` defaults to 180 cm but is a public Inspector field in Unity, so it can be adjusted for each participant before the session, ensuring that the drone does not clip the floor or float above it. The camera's forward vector is projected onto the horizontal plane (Y zeroed out) so the drone spawns directly in front of the participant. The calculated floor position is saved and reused by the landing coroutine (`WarmupLand`) to know exactly where to descend to.

Routes are generated by taking the Cartesian product of the seven horizontal angles and three vertical angles, producing 21 unique approach vectors. The order of these are randomized once per session using a Fisher-Yates shuffle, ensuring counterbalanced ordering across participants.

4.5 Controls and Experiment Guardrails

The controller uses four OVR right-hand controller bindings:

- Grip (`PrimaryHandTrigger`): First press initiates a warmup sequence; second press begins the experiment after warm-up is completed.
- A button (`Button.One`): Executes the current trial, i.e. makes the drone fly toward the participant.
- B button (`Button.Two`): Stops the drone, logs the stopping distance, and advances the trial state.
- Index trigger (`PrimaryIndexTrigger`): Repositions the drone to the next route's start position.

Input is gated by a set of Boolean flags (`warmupDone`, `takeoffDone`, `isRepositioning`, `trialComplete`) to prevent conflicting commands during transitions. For example, it is not possible to stop a route during repositioning or before the route has been initiated.

4.6 Route Position Calculation

Each trial's start position is computed from its angle pair using a spherical-to-Cartesian conversion. The horizontal angle is offset by -90° before conversion, such that hAngle = 90° corresponds to directly forward in the XZ-plane, and the vertical angle elevates or depresses the position relative to the participant's eye height.

A default approach distance of 2.65 m defines the radius of the sphere. Using `CalculateStartPosition(float hAngle, float vAngle)`, the offset is calculated as:

$$\text{offset} = d \begin{pmatrix} \sin(\theta - 90^\circ)\cos\phi \\ \sin\phi \\ \cos(\theta - 90^\circ)\cos\phi \end{pmatrix} \tag{4.1}$$

where d is the default approach distance, θ is the horizontal angle, and ϕ is the vertical angle. Using the headset as the calculation origin point, this offset is then used to calculate the starting position of the route as:

$$p_{\text{start}} = p_{\text{user}} + \text{offset} = p_{\text{user}} + d \begin{pmatrix} \sin(\theta - 90^\circ)\cos\phi \\ \sin\phi \\ \cos(\theta - 90^\circ)\cos\phi \end{pmatrix} \tag{4.2}$$

For example, a route starting at `hAngle = 90°` (directly forward) and `vAngle = 0°` (eye level) would give the following calculation:

- `hAngle = 90`, `vAngle = 0`
- offset $= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \times 2.65\text{m} = \begin{pmatrix} 0 \\ 0 \\ 2.65 \end{pmatrix} \text{m} \iff$ 2.65 m directly in front of the user

4.7 Drone Repositioning

Between routes, `FlyToNextStartPosition()` executes a three-phase repositioning sequence:

1. The drone rotates on its Y-axis to face the target position.
2. It moves directly to the new position using smooth 3D interpolation.
3. It rotates on its Y-axis again to face the participant.

4.8 Drone Approach

When the A button is pressed, `FlyToHeadset()` flies the drone from the current route's start position (`flightStart`) directly toward the participant's headset position (`flightEnd`), which is sampled at the moment the command is issued. Movement is computed via linear interpolation over time with a fixed speed of approximately 0.445 m/s. A threshold of 0.3 m automatically triggers the stop-and-log behaviour if `B` is not pressed before the drone reaches that distance.

4.9 Animation and Audio

Physical movement and animation are handled by `AS_DroneMovements`, which operates through Unity's Animator component. The drone model uses a two-tier hierarchy: a `droneRoot` object handles position and yaw rotation, while a child object carries the Animator, whose states (lean forward, lean back, idle) represent flight pose.

The `AS_Levitate` component applies a continuous sinusoidal offset to the drone while hovering, primarily along the Y-axis, to replicate the natural swaying of a hovering drone. Levitation is explicitly disabled during takeoff, landing, and repositioning, and re-enabled once the drone reaches its next trial position.

A looping propeller hum (`FLY.wav`) plays while the drone is hovering or flying, using a custom audio 3D rolloff curve (linear from 1.0 at 0.5 m to 0.15 at 15 m) to add spatial realism. Discrete events (takeoff, rotation, landing) play one-shot clips. All audio is fully spatial.

4.10 Data Logging

On application launch, `SC_DataLogger` creates a timestamped CSV file with the format `DroneSession_YYYY-MM-DD_HH-mm-ss.csv`. A header row is written immediately and each route appends a single row containing the timestamp, route index, horizontal and vertical approach angles in degrees and the measured stopping distance in meters. The stopping distance is computed and logged the moment the `B` button is pressed as the distance between the drone's world position and the participant's headset position.

4.11 Data Extraction and Processing

Following each session, data was retrieved from the headset to a PC using `adb pull` from the Android Debug Bridge (ADB) tool, which enables file transfer between the Meta Quest 3 and PC over USB-C.

Each participant is assigned a sequential index in the order they complete the experiment; this index is used as the filename for their stop distance data and is also recorded alongside their demographic information, enabling the stop distance measurements to be linked to individual participant characteristics during subsequent analysis.

4.12 Data Validation

Each session file was passed through a validation script (`validate_runs.py`) to check data integrity. The script defines the complete set of expected angle combinations as the Cartesian product of the valid vertical and horizontal angle sets. It then loads each CSV file and extracts the first 21 rows, which constitute one full experimental run. The validator checks these 21 rows against three error conditions: duplicate entries (the same angle combination appearing more than once), missing entries (an expected combination absent from the file), and unexpected entries (angle pairs outside the valid set). Any file with fewer than 21 rows is immediately flagged as incomplete, but files with more than 21 rows are simply stripped to the first 21 rows. The script reports if each passed or failed the validation and why.

4.13 Data Visualization

A visualization script (`visualize_stop_distance.py`) was used to produce an interactive 3D representation of the stopping distances across all sessions. The script loads all CSV files and applies the same 21-row extraction used during validation. Because the raw data is recorded in spherical coordinates, the script transforms each measurement into Cartesian space to allow correct 3D rendering. The conversion uses the standard spherical-to-Cartesian formulas:

$$p = r \begin{pmatrix} \cos(\phi)\sin(\theta - 90^\circ) \\ \cos(\phi)\cos(\theta - 90^\circ) \\ \sin(\phi) \end{pmatrix} \tag{4.3}$$

where r is the stopping distance, θ is the horizontal angle, and ϕ is the vertical angle. The participant position is placed at the origin.

Since the approach angles are defined relative to each participant, the data from all participants can be expressed in a collective egocentric coordinate system, with all participants positioned at the origin, no matter their height.

5

Data Analysis and Modeling Methods

This chapter presents the analytical frameworks used to analyze the study results and answer the research questions. It explains statistical analysis, linear regression for rubber-bubble modeling, LMM, and Bayesian modeling.

5.1 Statistical Analysis

To select appropriate statistical tests, a Shapiro-Wilk test [67] was first run on all 14 subgroups to check whether the data followed a normal distribution. Thirteen of the fourteen groups passed this check. The only exception was the short height group (p = .025), however, the Gen Z age group (N = 60) nearly violated the assumption (p = .071). Given this violation, and since stopping distance data has a physical lower limit, non-parametric tests were chosen for the primary statistical analyses [13]. See Table 5.1 for full reporting.

Table 5.1: Normality checks using Shapiro-Wilk tests.

Variable	Group	N	p
Gender	Male	43	.1169
Gender	Female	26	.1007
Pet ownership	No pets	25	.1026
Pet ownership	Has pets	44	.2762
Student status	Non-student	24	.2891
Student status	Student	45	.1508
Age group	Gen Z (≤ 29)	60	.0715
Age group	Millennial (30–45)	5	.3793
Age group	Gen X (46+)	4	.0952
Height group	Short (< 168 cm)	23	.0254
Height group	Medium (168–175 cm)	24	.7435
Height group	Tall (≥ 175 cm)	22	.5626
Culture	Swedish	37	.3145
Culture	Vietnamese	32	.1709

Height and age were split into three bins each since this was necessary for the statistical testing. Height was divided into three groups using tertile splitting, i.e. the sample's own 33rd and 67th percentiles. Age was split by conventional generations. The argument can definitely be made not to categorize continuous variables, such as height and age, but this is addressed by including them as standardized continuous variables in all modeling.

For comparisons between two groups, the Mann-Whitney U test was used [67], with effect size reported as rank-biserial r (range -1 to +1). For comparisons between three groups, the Kruskal-Wallis H test was used [67], with effect size reported as eta-squared (η^2), which should be interpreted as an approximation of explained variance based on ranked data rather than raw stopping distances, and may therefore underrepresent the magnitude of actual differences between groups [13, 19]. Effect sizes for r were interpreted as negligible ($< .10$), small (.10-.29), medium (.30-.49), or large ($\geq .50$), while η^2 values were categorized as negligible ($< .01$), small (.01-.05), medium (.06-.13), or large ($\geq .14$) [14], [42].

Given the violations of normality, non-parametric methods were used as the primary basis for frequentist statistical analysis. However, there are arguments for using parametric ANOVA tests despite these violations [53]. For this reason, ANOVA analyses were conducted [67], [61] and are reported in Appendix B.

To satisfy the independence assumption underlying the ANOVA, Mann-Whitney U and Kruskal-Wallis tests, stopping distance was averaged across trials for each participant before statistical testing. This reduced the repeated-measures data to one mean stopping-distance value per participant, ensuring that each value represented an independent participant rather than multiple correlated trials from the same individual.

Since making multiple comparisons increases the risk of false positives, the tests conducted should be evaluated with a Bonferroni-corrected significance threshold in mind:

$$\alpha_{\text{corrected}} = \frac{\alpha}{m} \tag{5.1}$$

where α is the chosen significant threshold (the conventional $\alpha = 0.05$ in this study) and m is the number of tests conducted. Although Bonferroni can be overly conservative when tested variables are related, such as student status and age overlap [32], the goal is to do exploratory data analysis, not to prove a hypothesis. The correction is made to be honest and transparent in the awareness of the increased risk of false positives.

Some of the post-study questionnaire items were measured on a 7-point Likert scale. Because Likert responses can reasonably be interpreted in different ways and there is debate regarding what is best test for Likert responses [16], three complementary tests were reported. Spearman's correlation treats the items as ordinal and tests for monotonic associations [60], Kruskal-Wallis provides a non-parametric comparison across response levels [16], and one-way ANOVA treats response levels as groups for comparing mean stopping distances [53]. Since each approach is defensible but rests on different assumptions and preferences, all three are reported for transparency and to allow readers to judge the results according to their preferred interpretation. All three are done using [67].

For Spearman's rank correlation, the correlation coefficient ρ was used as the effect size and interpreted by its absolute magnitude as negligible ($|\rho| < .10$), small ($.10 \leq |\rho| \leq .29$), medium ($.30 \leq |\rho| \leq .49$), or large ($|\rho| \geq .50$).

Both the Kruskal-Wallis and ANOVA tests use η^2 for the effect size. The difference between the two is that the Kruskal-Wallis η^2 should be interpreted as an approximation of explained variance based on ranked data, while the ANOVA η^2 represents the proportion of variance in the raw mean stopping-distance values explained by Likert response group. However, the same qualitative thresholds were used to classify their magnitude as described earlier in this Section: negligible ($< .01$), small (.01-.05), medium (.06-.13), and large ($\geq .14$).

To satisfy the independence assumption of these tests, stopping distance was first averaged across trials for each participant. Each Likert response was then paired with that participant's mean stopping distance, ensuring that each observation represented an independent participant rather than multiple correlated trials from the same individual.

Since multiple tests were conducted, a Bonferroni-correction (cf. Eq. 5.1) should be considered when interpreting the findings, as it provides a more conservative view on potential identified effects [32].

5.2 Testing Generalizability

The process initially utilized a standard train-test split approach to validate the rubber-bubble framework. K-fold cross validation was implemented to provide a more generalized version of a personalized model. However, the evaluation also tried a subject-aware framework, specifically Leave-One-Subject-Out Cross-Validation (LOSOCV) to measure the correlation of the investigated parameters with preferred stopping distances across new subjects. LOSOCV is a subject-aware variation of leave-one-out cross validation where the model trains on all subjects except one for each fold, repeated once for all subjects in the data set.

This transition helps ensure the validation set aligns with the assumption that data points should be independent and identically distributed (i.i.d.) to provide a realistic performance estimate across a general population [2]. As noted by [56], repeated measurements within the same individual tend to be more similar, meaning that models can inadvertently fit to the grouping structure itself rather than the underlying behavior, such as a high correlation between approach angles for a single user. Partitioning the data by subjects addresses this by preventing the model from adapting too closely to the low within-subject variation of specific participants. By ensuring that no model is tested on a subject included in its training set, this approach provides a more conservative and reliable estimate of how the model might perform for unseen individuals [24].

5.3 Rubber-Sheet to Rubber-Bubble Models

A primary objective of this thesis was to test the validity of the rubber-sheet proxemics model [12] as a function for autonomous drones to predict stopping distances. This mathematical framework utilizes polynomials to represent elastic shapes, making it suited for modeling the fluid boundaries of human-agent proxemics. Rubber-sheet transformations can account for multiple predictive features, such as the robot's approach angle and user-specific attributes like height and age, to generate a personalized approximation of comfortable stopping distances that can vary at different angles to the user [12].

The general 2^{nd} order polynomial for a model utilizing two features and an interaction term is defined as follows:

$$R_i(\lambda_i, \phi_i) = a + b_i\lambda_i + c_i\phi_i + d_i\lambda_i^2 + e_i(\lambda_i \times \phi_i) + f_i\phi_i^2 \tag{5.2}$$

In this equation:

- i: Index ranging from 1 to n.
  - n: number of data points.
- R_i: Predicted stopping distance.
- λ_i, ϕ_i: Predictive features, such as angle, height, or age.
- a: Intercept.
- b, c, d, f: Model coefficients.
- e: Interaction coefficient between λ_i and ϕ_i.

The rubber-sheet model for ground-based robots [12] is visualized in Figure 5.1. The figure shows how training a model on many subjects (blue sheets) then concludes in a final (orange) sheet that is stretched and contracted based on the current subject that the robot is interacting with. A drone would need a three dimensional zone as vertical angle must be included, which means that this rubber-sheet is stretched upwards and formed into a "rubber-bubble" instead (see Figure 5.2). The idea is the same, that the drone would identify what configuration of the rubber-bubble proxemic zone that a user should receive. Working best for users who have similar demographics to ones highly represented in the training data.

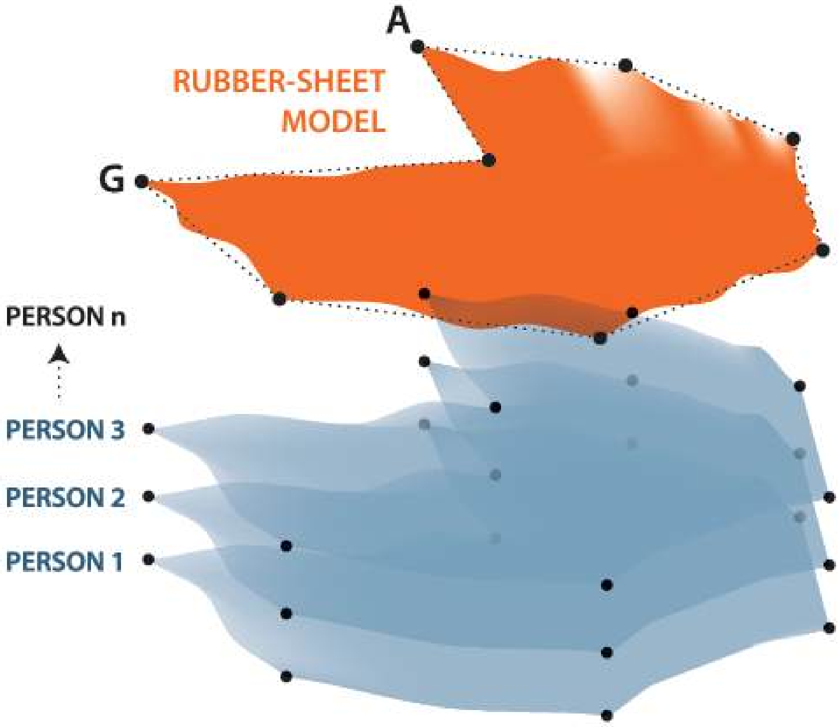


Figure 5.1: Schematic of the rubber-sheet model for different users. From Camara et al. [12], reproduced with permission.

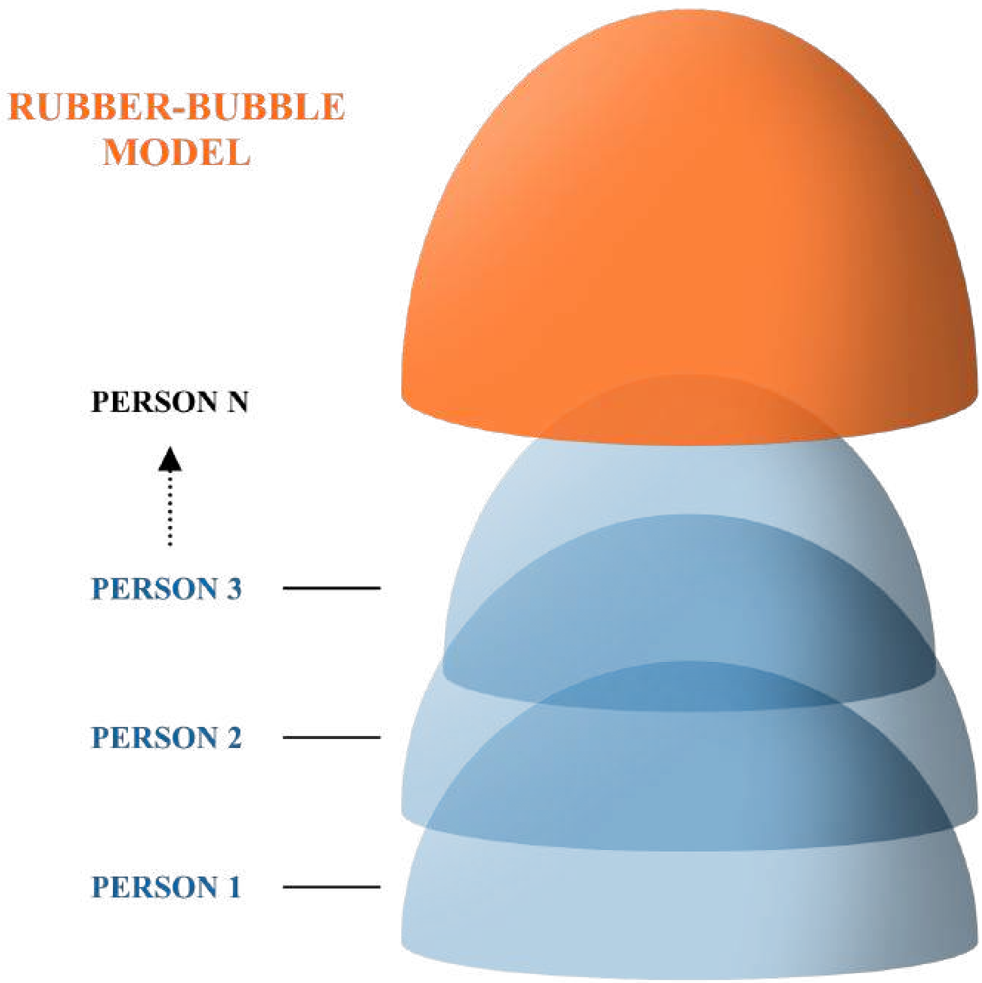


Figure 5.2: Schematic of the rubber-bubble model for different users. Keep in mind that the bubble is more deformed in reality since a perfectly smooth surface means the same preferred stopping distances at all approach angles.

Model coefficients and variables were determined through an evaluation of all possible feature combinations of our subsets in standard OLS models. Ridge, and lasso regression models were also used in an attempt at improving the models. Following the findings of [12], which suggest that models incorporating a larger number of variables may benefit from higher degrees, polynomial models of degrees one through ten were evaluated for subset one and polynomials of degrees one through five for the LOSOCV subset two evaluation. However, it is important to note that for real-world deployment, the model should not be too complex as the drone needs to calculate stopping distances rapidly after identifying a user in order to avoid approaching too closely before determining the appropriate distance.

The process for finding the best model involved an exhaustive search across all combinations of objective parameters and chosen polynomial degrees to identify the most accurate configuration. For the purposes of this study, "objective parameters" refers to data points objectively associated with each participant. These were categorized into two distinct subsets based on their practical application in autonomous navigation.

The first subset consists of demographical features that a socially intelligent drone could reasonably estimate or utilize in a real-time approach scenario to determine an optimal IPD. The second subset incorporates broader contextual factors to evaluate whether their inclusion significantly enhances the model's predictive power and thus looks for a best-case scenario model for new subject prediction but that is of less use in a realistic scenario.

- **Subset 1 (Real-time Applicable Features):**
  - Horizontal approach angle
  - Vertical approach angle
  - Age
  - Gender
  - Height
- **Subset 2 (Comprehensive Demographic Features):**
  - All features included in Subset 1
  - Student status (0 = is not student, 1 = is student)
  - History of owning free-roaming pets (0 = have never had pets, 1 = have had or has pets)

Subset two was only investigated in LOSOCV since they are not as applicable for a drone in a real life scenario. This was done simply to test their predictive power. Rubber-bubble models were stratified by culture as an autonomous drone could be equipped with a culturally fitting model to where it is deployed.

The modeling was implemented using the scikit-learn library in Python. To generate the higher order terms for the rubber-bubble framework, the PolynomialFeatures class was used to transform input features into the specified polynomial degrees before fitting. Since the models benefit from standardized input for interpretability, and Ridge and Lasso additionally require it to ensure the regularization penalty is applied equally across features, all inputs were standardized using StandardScaler prior to fitting. For the regularized models, hyperparameter optimization was performed by testing a range of alpha values between 10^{-6} and 10^{3} using a logarithmic grid search. While their alpha tuning used a standard k-fold approach, this does not compromise subject-based validation integrity, as the inner loop operates exclusively within the outer training fold and held-out subjects are never exposed during hyperparameter selection. Since Ridge and Lasso are designed to shrink or remove unnecessary parameters, they were primarily tested with the full parameter set from

Subset two to allow the models to perform their own feature selection [27], [63]. The final models were selected by comparing every combination of objective parameters and selecting the best ones. Performance was measured using R^2 and Root Mean Squared Error (RMSE) scores. While multicollinearity exists between certain demographic features, such as height and gender, it does not impact overall predictive capabilities [40]. Because the primary objective of the rubber-bubble model is prediction rather than isolating causal effects or interpreting individual coefficients, this correlation does not compromise the validity of the predictions. See Appendix A for a demonstration of how to calculate the output of a standardized model.

Equation (5.3) shows a calculation of the number of parameters in a polynomial regression model as a function of the number of degrees d and the number of input features k, including all interaction and higher-order terms generated by the polynomial expansion.

$$P(d,k) = \frac{(d+k)!}{d! \times k!} \tag{5.3}$$

In this equation:

- $P(d,k)$: Total number of parameters (including the intercept term) in the polynomial regression model.
- d: Maximum polynomial degree of the model.
- k: Number of input features included in the regression model.

The following are the evaluation methods used for the rubber-bubble models. RMSE is used to measure the average deviation between the predicted stopping distances and the observed values. It penalizes larger errors more heavily than smaller ones, which is useful for identifying models that fail significantly on specific outliers. It is expressed in the same units as the target variable (meters).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5.4}$$

In this equation:

- n: Total number of observations.
- y_i: The actual observed stopping distance for trial i.
- $\hat{y}_i$: The distance predicted by the model for trial i.

R^2 represents the proportion of the variance in the stopping distance that is predictable from the independent variables. In the context of this study, it indicates how much of the variation in human proxemic behavior is captured by the model compared to a simple mean baseline.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5.5}$$

In this equation:

- SS_{res}: The sum of squares of residuals (the uncaptured variance).
- SS_{tot}: The total sum of squares (the total variance in the data).
- $\bar{y}$: The mean of all observed stopping distances.

5.4 Linear Mixed Model

To account the multiple trials nested within participants, a linear mixed effects model (LMM) was fitted using the `statsmodels` library in Python [61]. Unlike the polynomial regression models, which treat all observations as independent, the LMM explicitly models the dependency between repeated measurements from the same participant [4]. The model was specified as:

$$StopDistance \sim HAngle + VAngle + Demographics + (1|ParticipantID) \tag{5.6}$$

Fixed effects included horizontal approach angle, vertical approach angle, and demographic variables such as age, height, gender, pet ownership, and student status. A random intercept per participant was included to capture stable individual differences in baseline stopping distance. This random intercept allows each participant to have their own baseline comfort distance while sharing the same fixed effect slopes across the population.

Three model variants were compared: angles only, angles combined with objective parameters, and objective parameters only. All models were fitted using restricted maximum likelihood (REML), which provides unbiased estimates of the variance components. For model comparison via AIC and BIC, models were refitted using full maximum likelihood (ML), as REML based likelihoods are not comparable across models with different fixed effect structures [66].

To assess generalization to unseen participants, LOSOCV was performed. In each fold, predictions for the held out participant were generated using only the fixed effect component. This is done by setting the random intercept to zero, which is equivalent to treating the participant as a new, previously unseen individual. LOSOCV RMSE was averaged across all folds and compared against the standard deviation of the target variable to contextualize predictive accuracy.

For the Linear Mixed Models, R^2 is divided into marginal and conditional components following the method by Nakagawa and Schielzeth [47]:

Marginal R^2 represents the proportion of variance explained by the fixed effects alone, such as approach angles and demographics. It describes how much of the total variation can be predicted by the general population trends without knowing the specific identity of the participant.

$$R_m^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\epsilon^2} \tag{5.7}$$

Conditional R^2 represents the variance explained by the entire model, including both fixed effects and random effects (individual subject offsets). This metric describes how well the model fits the existing data when the identity of each participant is known.

$$R_c^2 = \frac{\sigma_f^2 + \sigma_\alpha^2}{\sigma_f^2 + \sigma_\alpha^2 + \sigma_\epsilon^2} \tag{5.8}$$

In these equations:

- σ_f^2: Variance explained by the fixed effects.
- σ_α^2: Variance explained by the random effects (between-subject variance).
- σ_ϵ^2: Residual variance (within-subject noise).

The intra-class correlation coefficient (Intra-class Corellation Coefficient (ICC)) is used to quantify the degree of clustering in the data. In this study, it measures how much of the total variation in stopping distance is due to permanent differences between participants versus random noise between trials. A high ICC justifies the use of a Mixed Model, as it confirms that observations within the same participant are correlated, making linear mixed models more fitting than Linear Regression.

$$ICC = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2} \tag{5.9}$$

In this equation:

- σ_α^2: The between-subject variance (variance explained by the random intercept).
- σ_ϵ^2: The residual variance (within-subject noise).

5.5 Bayesian Model

Unless otherwise stated, all theoretical and methodological presentation in this section follows McElreath's *Statistical Rethinking* [44].

The Bayesian model was included as an additional modeling perspective, because the fitted model's posterior can be used for further data analysis, particularly in answering RQ2.

The Bayesian approach also provides a natural way to estimate uncertainty. Parameters are themselves modeled as outcomes of their own distribution, producing multiple levels of uncertainty that are carried forward all the way to final inference. Thus, instead of producing only a single output for each parameter, the model estimates a posterior distribution using the distributions of the parameters. In this thesis, this means that the effects of gender, culture, age, and height can be interpreted together with the uncertainty around those effects. The same applies to the participant-level variation. This is useful because proxemic behavior is expected to vary between individuals, and the data is not only used to estimate average patterns but also to understand how much variation remains between participants after accounting for the included predictors.

Specifically, this study incorporates a Bayesian multilevel Generalized Linear Model (GLM) with partial pooling. The model has more than one level because individual data points come from a participant, who in turn comes from a population. Thus, the data has the following structure:

$$\text{Trial Observation} \rightarrow \text{Subjects} \rightarrow \text{Population}$$

Each stopping distance measurement depends on a participant who has their own baseline stopping distance and is assumed to come from a population that shares a common distribution. In this way, model parameters exist at more than one level, such as the subject level and the population level. Partial pooling provides a compromise between the dichotomy of estimating each subject completely separately and assuming that all participants are uniform within the shared population. Each subject is allowed to have their own baseline estimate, but this estimate is informed not only by the subject's own data, but also by the overall population pattern. Each subject-specific intercept is shrunk toward the population intercept according to the amount of information available for that subject and the estimated between-subject variation. In this way, the model preserves individual variation while simultaneously allowing general trends to emerge in the population. Another advantage of partial pooling is that it prevents over-sampled clusters from unfairly dominating inference; a participant, or group of participants, with disproportionately many observations can still be estimated more precisely, but they do not dominate the analysis solely by weight of repetition.

In Bayesian inference, the posterior distribution is proportional to the prior distribution multiplied by the likelihood, as per:

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood} \tag{5.10}$$

The prior expresses assumptions about plausible parameter values before observing the data, while the likelihood describes how compatible different parameter values are with the observed data. The posterior therefore represents the updated plausibility of the parameters after combining the model assumptions with the evidence provided by the data.

5.5.1 Causal Assumptions and Directed Acyclic Graph (DAG)

The Directed Acyclic Graph (DAG) presents the assumed causal structure between the demographic variables included in the Bayesian model and the observed stopping distance. The purpose of the DAG is not to claim that the structure can be proven, but to provide a transparent representation of how demographic variables are assumed to causally relate to each other and the preferred stopping distance. See Figure 5.3 for a visualization, or Listing 5.1 for the code used to render it using DAGitty.

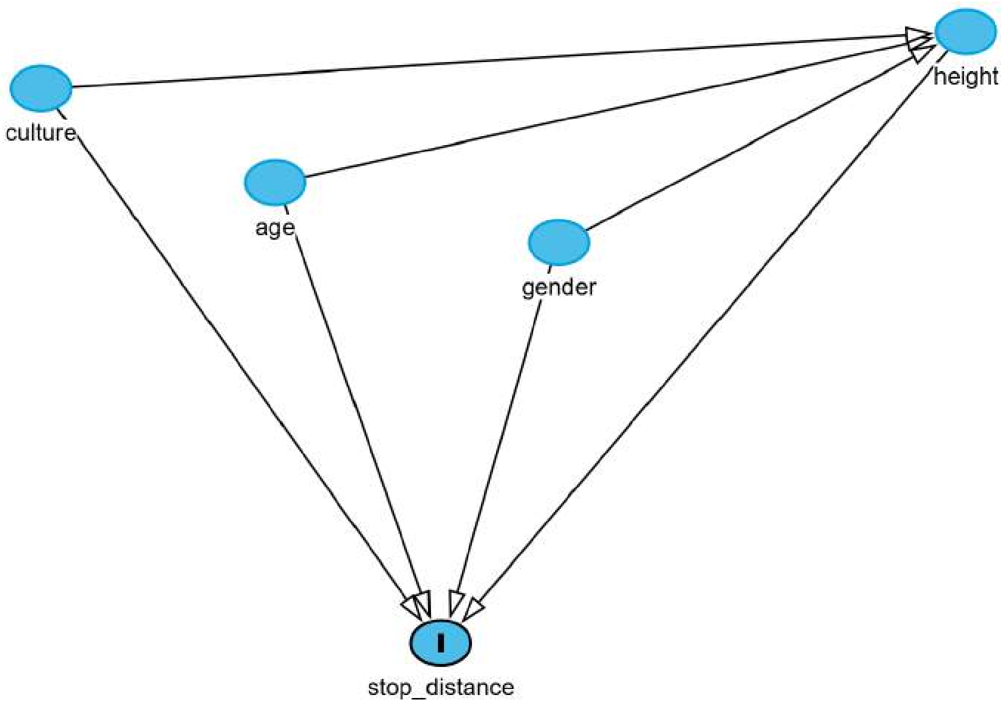


Figure 5.3: Visualization of the DAG.

Listing 5.1: DAGitty code for the DAG.

```
dag {
bb="0,0,1,1"
age [pos="0.417,0.245"]
```

```
culture [pos="0.364,0.196"]
gender [pos="0.482,0.277"]
height [pos="0.574,0.166"]
stop_distance [outcome,pos="0.455,0.489"]
age -> height
age -> stop_distance
culture -> height
culture -> stop_distance
gender -> height
gender -> stop_distance
height -> stop_distance
}
```

The selected variables correspond to the "realistic subset" of features: age, gender, height, and culture. These variables were chosen because they are either directly observable or realistically available in a deployment context, unlike more subjective variables that would require questioning or prior knowledge of the user.

Age, gender, culture, and height are all modeled as possible direct predictors of stopping distance. This is supported by prior work in human-robot and human-drone proxemics, where personal characteristics such as age, height, gender and culture have been reported to influence preferred interaction distances [46], [35], [71], [41]. In the DAG, these direct paths therefore represent the assumption that each variable may contribute to how close a person allows a drone.

Height is modeled as being influenced by age, gender, and culture. The path from gender to height reflects that males tend to be taller than females. The path from culture to height reflects population-level differences in height distributions that arise from genetic and nutritional factors. Even though most of the participants in this study were adults, there were younger participants as well, making this relationship necessary to include.

The DAG should be understood as a modeling assumption rather than a definitive causal claim. It is presented to make the assumed structure visible and open to criticism. In this study, the DAG provides a basis for including age, gender, height, and culture in the Bayesian model, while acknowledging that the estimated associations should not necessarily be interpreted as isolated direct causal effects. Instead, they may also capture indirect relationships through other unmeasured factors or differences in how participants interpreted the interaction.

5.5.2 Bayesian Model Specification

The Bayesian model was specified as a multilevel Generalized Linear Model (GLM) with partial pooling. The outcome variable was the participant's stopping distance to the drone in centimeters.

Gender, culture, age and height were included as predictors because the DAG assumes that each may be associated with stopping distance. Gender was coded as $G = 0$ for male and $G = 1$ for female, and culture was coded as $C = 0$ for Swedish participants and $C = 1$ for Vietnamese participants.

Age and height were standardized before modeling, so their coefficients represent changes associated with a one standard deviation increase in the predictor. A continuous variable x with standard deviation σ_x can be standardized using:

$$x_i = \frac{x_i - \bar{x}}{\sigma_x} \tag{5.11}$$

meaning the age is standardized using:

$$A_i = \frac{\text{age}_i - \bar{\text{age}}}{\sigma_{\text{age}}} \tag{5.12}$$

and the height is standardized using:

$$H_i = \frac{\text{height}_i - \overline{\text{height}}}{\sigma_{\text{height}}} \tag{5.13}$$

The model used a Gamma likelihood for stopping distance:

$$D_i \sim \text{Gamma}(\mu_i, \text{scale}) \tag{5.14}$$

where D_i is the observed stopping distance for observation i, μ_i is the expected stopping distance, and scale controls the dispersion of the Gamma distribution.

Ontologically, the model treats stopping distance as a positive continuous outcome generated by repeatedly stopping an approaching drone. The Gamma likelihood reflects this because distance cannot be negative and because the distribution of stopping distances may be right-skewed, with most observations near a typical value and some larger values.

Epistemologically, the Gamma likelihood represents uncertainty about each observed distance. The model does not assume that a participant has one perfectly fixed stopping distance that would be reproduced exactly in every trial. Instead, it assumes that there is an expected stopping distance for a given participant and condition, but that each actual observed stop is a noisy realization around that expectation. The variation between a realized stopping distance and the expected one captures factors that are not fully observed in the model, such as perception, timing, comfort, attention, and whatever other differences participants have in judging the approaching drone. The Gamma distribution is therefore not used because stopping distances are assumed to be literally Gamma distributed by nature, but because it provides a reasonable probabilistic description of positive, continuous, and variable stopping-distance observations.

The predictors in the model were gender, culture, standardized age, and standardized height, in accordance with the established DAG:

$$\log(\mu_i) = \alpha_{\text{subject}[i]} + \beta_G G_i + \beta_C C_i + \beta_A A_i + \beta_H H_i \tag{5.15}$$

where $\alpha_{\text{subject}[i]}$ is the intercept for the participant associated with observation i. The coefficients β_G, β_C, β_A, and β_H estimate the conditional association between each predictor and stopping distance, after accounting for the other predictors and the participant-specific intercept.

As can be seen in Eq. 5.15, rather than modeling stopping distance directly with an unconstrained linear model, the expected distance was modeled through a log link. This means that the linear predictor operates on the logarithmic scale, while the expected stopping distance on the original centimeter scale is obtained by exponentiation. Using a log link, a positive parameter can be mapped onto a linear model. This is suitable because the expected stopping distance cannot be negative, while the linear predictor itself can vary freely across the real number line. Since a log link is used, these coefficients are interpreted on the log scale. Therefore, a coefficient is not interpretable as a multiplier until after exponentiating.

For gender, β_G represents the expected log-scale difference between female and male participants, given the coding used in the model. For culture, β_C represents the expected log-scale difference between Vietnamese and Swedish participants. For age and height, the coefficients represent expected differences associated with a one standard deviation increase in the standardized predictor. These interpretations are conditional on the model structure and should not be read as direct causal effects unless the causal assumptions behind the model are also accepted fully.

5.5.3 Subject-level Variation and Partial Pooling

The data contains repeated stopping distance observations from each participant. Treating all observations as fully independent would ignore the fact that responses from the same person are likely to be more similar to each other than to responses from other participants. The model therefore includes a subject-specific intercept:

$$\alpha_j \sim \text{Normal}(\alpha, \sigma_{\text{subject}}) \tag{5.16}$$

where α_j is the intercept for participant j, α is the average population-level intercept, and σ_{subject} describes how much participant baselines vary around that population average.

The parameter σ_{subject} is meaningful in the context of this study. A small value would indicate that participants have similar stopping distances after accounting for the predictors. A larger value would indicate that participants differ in their preferred stopping distances for reasons not captured by the included predictors.

5.5.4 Prior Specification

The priors were chosen to keep the model within plausible ranges while still allowing the data to update the parameter estimates. In particular, weakly informative and regularizing priors can reduce overfitting by making the model skeptical of extreme values. Such priors reduce fit to the sample but often improve predictive accuracy, since they discourage the model from adapting too strongly to noise in the observed data.

The population intercept was assigned the prior:

$$\alpha \sim \text{Normal}(\log(133), 0.5) \tag{5.17}$$

Because the model uses a log link, the intercept is defined on the log-distance scale. Centering the prior at log(133) places the prior expectation for the baseline stopping distance around 133 cm before accounting for predictors or subject-level variation. This value was chosen because it is the halfway point of the distance between the participant and the drone at its starting position for any route rounded up ($\frac{265}{2} = 132.5$). This does not force the posterior estimate to remain near 133 cm, but it gives the model a reasonable starting assumption for the expected distance. The standard deviation of 0.5 on the log scale allows variation around this baseline while discouraging unrealistically extreme baseline values before the data are considered.

The predictor coefficients were assigned the following priors:

$$\beta_G \sim \text{Normal}(0, 0.25) \tag{5.18}$$

$$\beta_C \sim \text{Normal}(0, 0.25) \tag{5.19}$$

$$\beta_A \sim \text{Normal}(0, 0.25) \tag{5.20}$$

$$\beta_H \sim \text{Normal}(0, 0.25) \tag{5.21}$$

These priors are centered on zero, meaning that before seeing the experimental data the model does not assume any effect by gender, culture, age, or height on the preferred stopping distance. The standard deviation of 0.25 makes the priors regularizing rather than flat. Gaussian priors centered on zero shrink estimates toward zero, with narrower priors expressing stronger skepticism toward large effects. The prior standard deviation of 0.25 allows meaningful predictor effects, while simultaneously making very large effects less likely before seeing the data. This is appropriate because the measured demographic predictors are expected to potentially be relevant, but not so dominant that they overwhelm participant-level variation entirely.

The subject-level standard deviation was assigned the prior:

$$\sigma_{\text{subject}} \sim \text{Exponential}(6) \tag{5.22}$$

This parameter must be positive because it represents variation between participant intercepts. The exponential prior constrains it to positive values and regularizes very large between-subject variation. This prior expresses that some individual variation is expected, but that large variation should require support from the data.

The Gamma scale parameter was also assigned an exponential prior:

$$\text{scale} \sim \text{Exponential}(2) \tag{5.23}$$

This parameter must also be positive, since it controls the spread of the Gamma likelihood around the expected stopping distance.

Prior predictive checks were done to assess whether the priors imply reasonable stopping distances before the model is fitted to the observed data. This is done by simulating from the prior distribution alone, using the sampled parameter values to generate possible outcomes and then evaluating whether those outcomes are plausible for this study. This is useful because prior predictive checks help diagnose poor prior choices by showing what the model expects before seeing the data, i.e. asking the question "What stopping distance would our assumptions generate if we had no data?" In this study, prior predictive checks therefore provide a way to judge whether the chosen priors generate realistic stopping distances for this specific experiment before the posterior model is estimated. Figure 5.4 shows the simulated stopping distances using only the priors established above visually. As shown in Table 5.2, the chosen priors are reasonable for the experimental setting. Across conditions, the central 80% of simulated stopping distances basically covers the full approach distance of 2.65 m. Although the priors occasionally generate extreme values of > 20m, these are rare, and with 4000 random samples from the prior it is expected that the maximum may include unusually large values from the tail of the distribution. This mainly indicates that the priors remain permissive before seeing the data; once conditioned on the observed stopping distances, the posterior is expected to place most probability on values supported by the experiment.

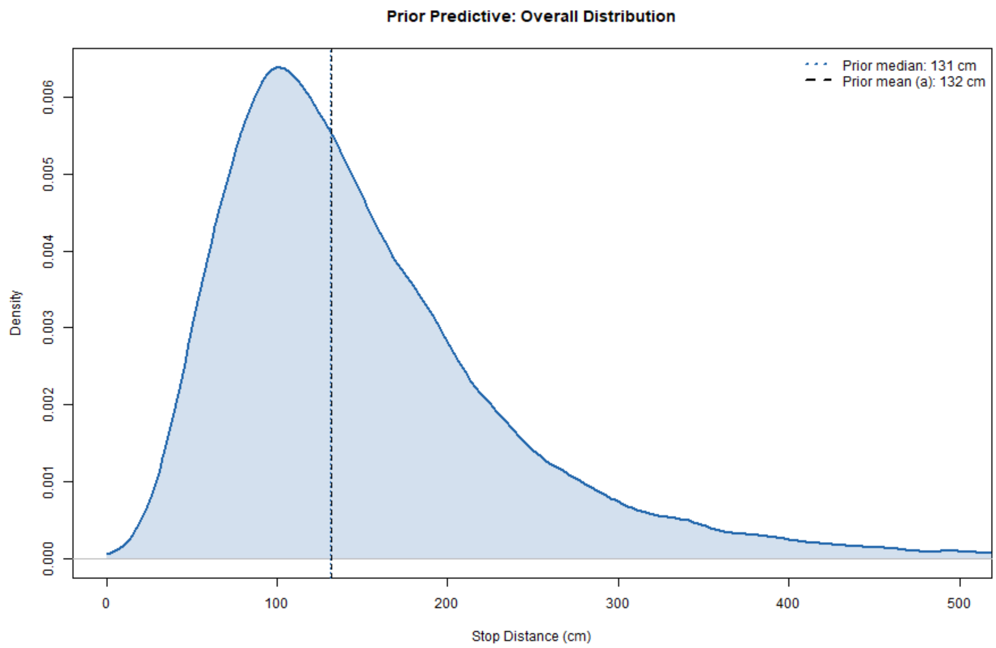


Figure 5.4: Simulated stopping distance distribution using only the priors.

Table 5.2: Prior predictive stopping distance summaries by different conditions.

Condition	Median (cm)	10% (cm)	90% (cm)	Max (cm)
Baseline (M, Swe, avg)	131.4	64.9	266.7	1263.7
Female	131.7	60.9	283.1	961.5
Vietnamese	133.3	60.7	286.1	2608.7
Young (−1 SD age)	130.4	61.1	285.2	2334.5
Old (+1 SD age)	132.2	60.8	291.6	2396.1
Short (−1 SD height)	131.8	59.8	284.8	2162.2
Tall (+1 SD height)	130.9	60.7	286.8	4314.3

5.5.5 Model Diagnostics

The Bayesian model was evaluated using standard Markov Chain Monte Carlo (MCMC) diagnostics to assess whether the posterior samples were sufficiently reliable for inference. This is achieved by assessing the convergence and sampling efficiency using the $\hat{R}$ statistic and the effective sample size, n_{eff}.

The $\hat{R}$ statistic compares variation within and between independent MCMC chains. If the chains have converged to the same posterior distribution, the between-chain and within-chain variation should be similar, and $\hat{R}$ should be close to 1. Values noticeably above 1 indicate that the chains may not have mixed properly or may still be exploring different regions of the parameter space. In this study, $\hat{R} < 1.01$ was used as the convergence criterion.

The effective sample size, n_{eff}, estimates how many independent posterior samples the MCMC output is approximately equivalent to. This is necessary because consecutive MCMC samples are typically autocorrelated and therefore contain less independent information than the raw number of samples suggests. Low effective sample sizes indicate inefficient sampling and imply that posterior summaries, such as means and compatibility intervals, may be less stable. In this study, parameters were considered sufficiently sampled when $n_{\text{eff}} > 400$, even though as few as 200 can be enough for a good estimate of the posterior.

Trace plots were also inspected as a visual diagnostic to ensure that chains mixed well and showed no systematic trends over the sampling process.

5.5.6 Null and Full Models Comparison

To evaluate whether the demographic and cultural predictors improve out-of-sample prediction, two different Bayesian models were compared: a null model and a full model. Both models used the same likelihood, link function, subject-level structure, and hyperpriors. The difference between them was that the null model included only subject-level variation, while the full model also included gender, culture, age, and height as population-level predictors. In simple terms, the null model assumes that stop distances vary between individuals, but that gender, culture, age, and height do not explain these differences. The full model makes the same allowance for individual variation, but additionally estimates whether gender, culture, age, and height affect the expected stop distance. The comparison therefore isolates the added predictive value of these population-level predictors.

The null model was specified as:

$$\begin{aligned} D_i &\sim \text{Gamma}(\mu_i, \text{scale}) \\ \log(\mu_i) &= \alpha_{\text{subject}[i]} \\ \alpha_j &\sim \text{Normal}(\alpha, \sigma_{\text{subject}}) \end{aligned} \tag{5.24}$$

The full model was specified as:

$$\begin{aligned} D_i &\sim \text{Gamma}(\mu_i, \text{scale}) \\ \log(\mu_i) &= \alpha_{\text{subject}[i]} + \beta_G G_i + \beta_C C_i + \beta_A A_i + \beta_H H_i \\ \alpha_j &\sim \text{Normal}(\alpha, \sigma_{\text{subject}}) \end{aligned} \tag{5.25}$$

If the full model predicts held-out subjects better than the null model, this suggests that the included predictors contain information about stopping distance that generalizes to new participants. If the two models perform similarly, this suggests that most of the prediction is already captured by the population-level subject variation and that the added predictors contribute little to out-of-sample prediction.

The models were compared using LOSOCV.

Predictive performance was measured using the log predictive density for each held-out subject. For each subject, the comparison computes:

$$\Delta\mathrm{LPD} = \mathrm{LPD}_{\mathrm{full}} - \mathrm{LPD}_{\mathrm{null}} \tag{5.26}$$

where a positive value indicates better predictive performance for the full model. Using LPD, deviance can be expressed as:

$$\Delta\mathrm{deviance} = -2 \times \Delta\mathrm{LPD} \tag{5.27}$$

With this transformation, negative values favor the full model and positive values favor the null model.

This comparison should be interpreted as a comparison of generalized predictive performance, not as a direct causal test. In this study, the comparison is therefore used to evaluate whether gender, culture, age, and height improve prediction of stopping distances for unseen participants.

6

Results

This chapter presents the results of the data analysis and predictive modeling. A data analysis is presented on the combined dataset ($N = 69$) to examine the distributions of stopping distances and evaluate whether any objective predictors demonstrate statistical significance. The polynomial rubber-bubble models are evaluated. The results of the LMM and Bayesian modeling are then presented. The predictive capabilities of the post-study questions are also evaluated.

6.1 Data Cleaning

In this chapter, uppercase N stands for number of subjects and lowercase n stands for individual data points (i.e. angles).

A total of $N = 78$ subjects participated in the study, of which 43 were in Sweden and 35 in Vietnam.

To obtain as good a cultural comparison as possible, participants who did not culturally identify as Swedish or Vietnamese were removed, resulting in the removal of five participants from the Sweden subset.

A total of eight data points (i.e. eight individual routes) were removed for data integrity, one in Sweden and seven in Vietnam. One participant (no. 74) in Vietnam was removed because they admitted they had a hard time seeing the drone since they had removed their glasses, thus letting it come closer than they may have otherwise.

Some participants let the drone approach to its closest allowed point of 30 cm (meaning the rotor blades were 19 cm away) and would have let it come closer had it not stopped automatically. For the purpose of a prediction model this would be unrealistic and causes severe safety concerns for users. Letting it come this close may be attributed to the perceived safety of the environment, specifically the participants' awareness that the virtual drone posed no physical risk of collision (lack of suspension of disbelief). Some users conceptualized the interaction as a delivery scenario requiring close proximity, though this does not provide a complete explanation, as one would not want a delivery drone close enough to be a safety risk. A cutoff point was chosen at 45 cm in accordance with Hall's intimate zone border of 45 cm [21] and Neggers et. al observed minimum human-robot IPD of 50 cm [49] to ensure outliers like these were excluded from the dataset. Some leeway is granted through this cutoff since it is at the anchor point and not at the front most part of the drone. This was done to not exclude close preferences due to the fact that users' may want robots closer than they would want humans [49]. This threshold is more conservative than the 60 cm minimum distance observed in [15] or the 130 cm observed by Wögerbauer et al. [70]. This resulted in the removing of 130 data points or 8.6% of the remaining 1504 points, including the complete removal of participants no. 51, as well as 39 and 77 since this cut-off left them with only one data point. All other participants have at least 10 data points remaining. This resulted in a total of $n = 1372$ data points remaining with 52 participants having all their 21 angles left.

This resulted in a total of $N = 69$ participants with 37 who identify as Swedish and 32 who identify as Vietnamese.

6.2 Sample Overview

Figure 6.1 summarizes the demographic composition of the Swedish participant group. The sample had a relatively balanced gender distribution and included both students and non-students. Participant ages were concentrated primarily within two clusters: a dominant 20-30 age range and a smaller 55-60 age range. The boxplot whiskers are contained within the 20-30 interval, emphasizing the dominance of younger participants in the sample. Height distributions displayed moderate variability across the sample.

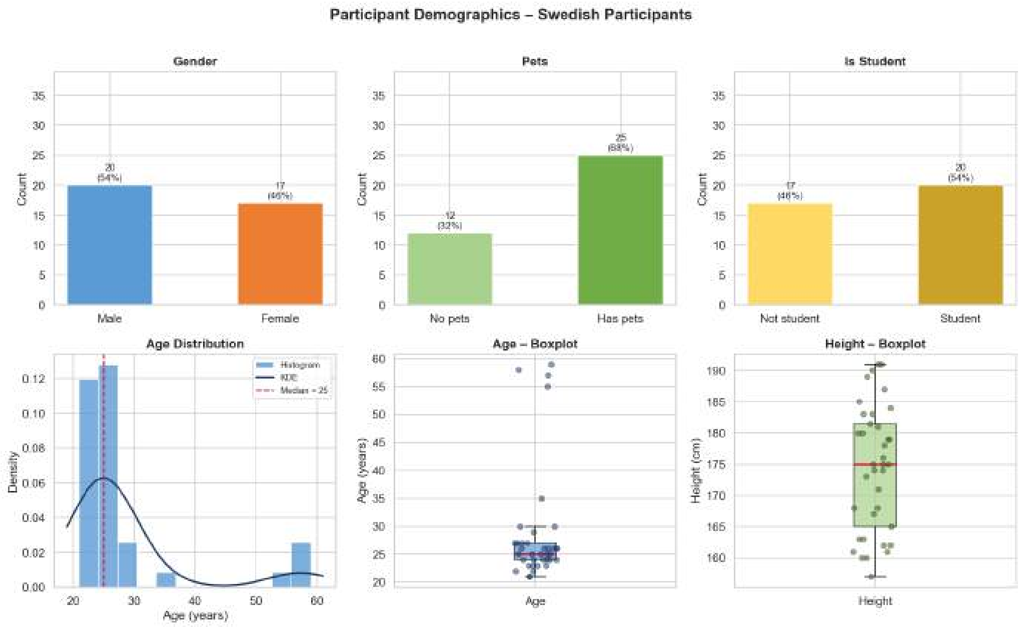


Figure 6.1: Demographics for the Swedish sample.

Figure 6.2 summarizes the demographic composition of the Vietnamese participant group. This sample contains a majority of male participants and only two non-students. Participant ages are almost all in the lower twenties with only two participants above this, both in their thirties. The Vietnamese sample has a median height of 170 cm compared to the Swedish 175 cm, even though the Vietnamese sample is dominated by males (72% compared to Sweden's 54%).

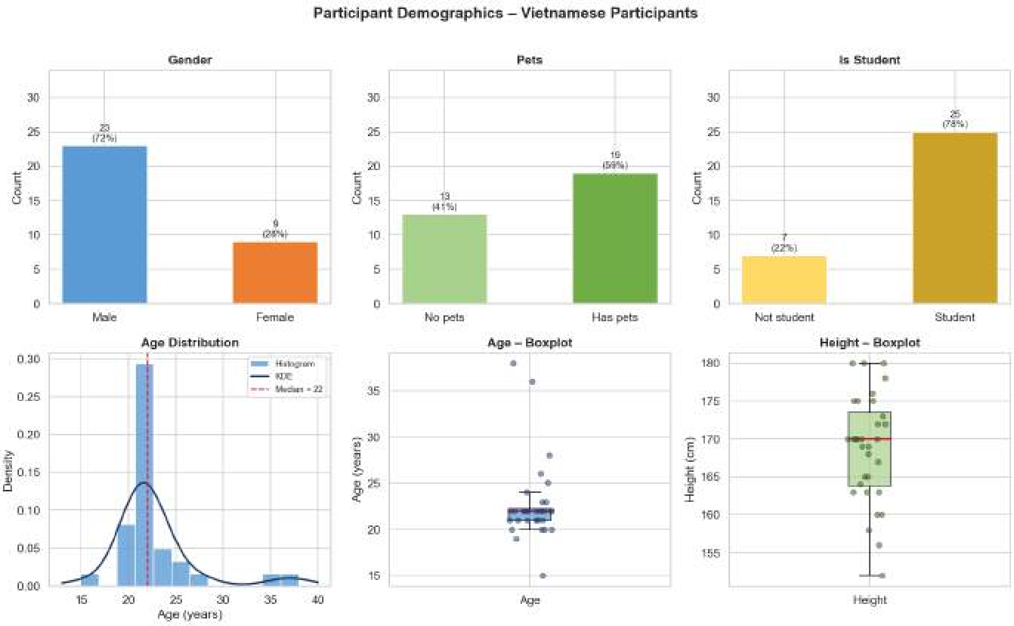


Figure 6.2: Demographics for the Vietnamese sample.

Figure 6.3 shows the combined sample set used for the general models.

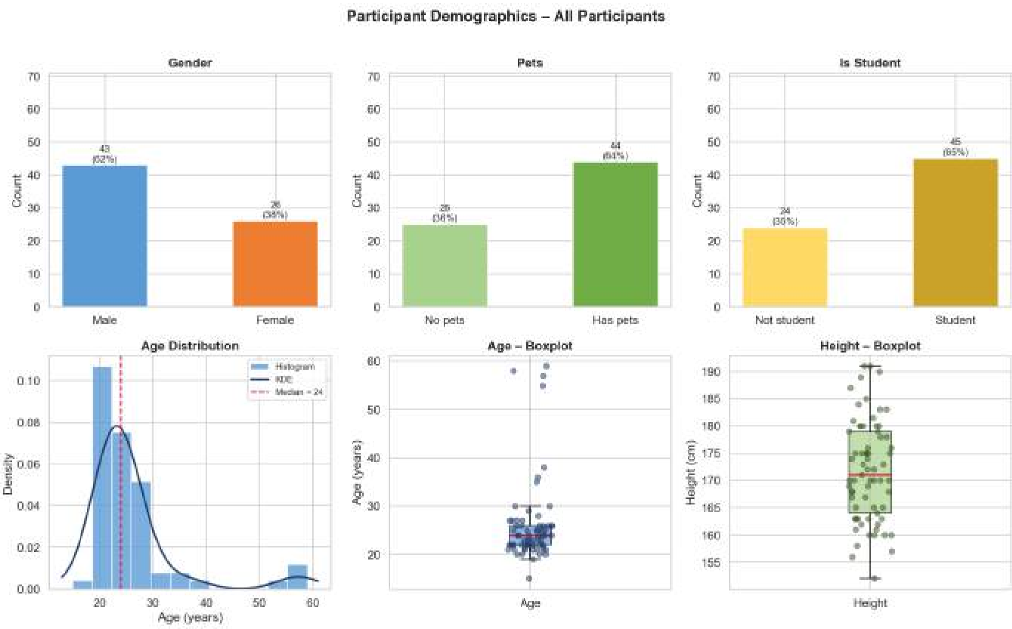


Figure 6.3: Demographics for the full sample.

Figures 6.4, 6.5, and 6.6 show the distribution of stopping distances. The Swedish dataset has fewer stop distances around the intimate-zone cutoff point (0.45 m) and further away than the Vietnamese. However, both subsets have similar means lying between 0.9 to 1.0 m. Note that *Mdn* refers to the median, *M* refers to the mean, and *SD* refers to the standard deviation.

Swedish (n = 746): Mdn = 0.903 m, M = 0.945 m, SD = 0.267 m, Min = 0.452 m, Max = 1.809 m

Vietnamese (n = 626): Mdn = 0.933 m, M = 0.996 m, SD = 0.365 m, Min = 0.451 m, Max = 2.449 m

General (n = 1372): Mdn = 0.918 m, M = 0.968 m, SD = 0.316 m, Min = 0.451 m, Max = 2.449 m

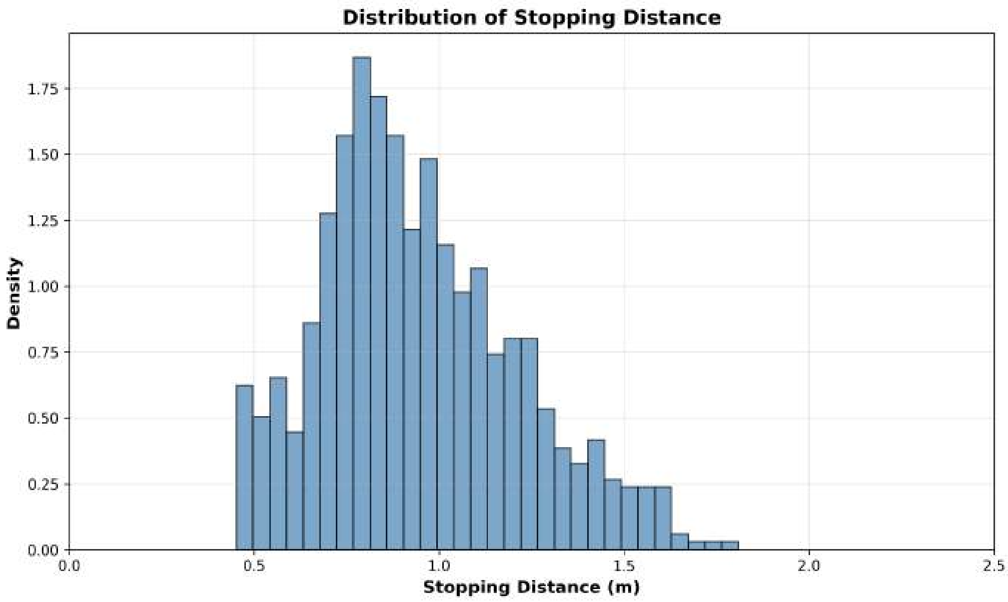


Figure 6.4: Stopping distance distribution for the Swedish sample.

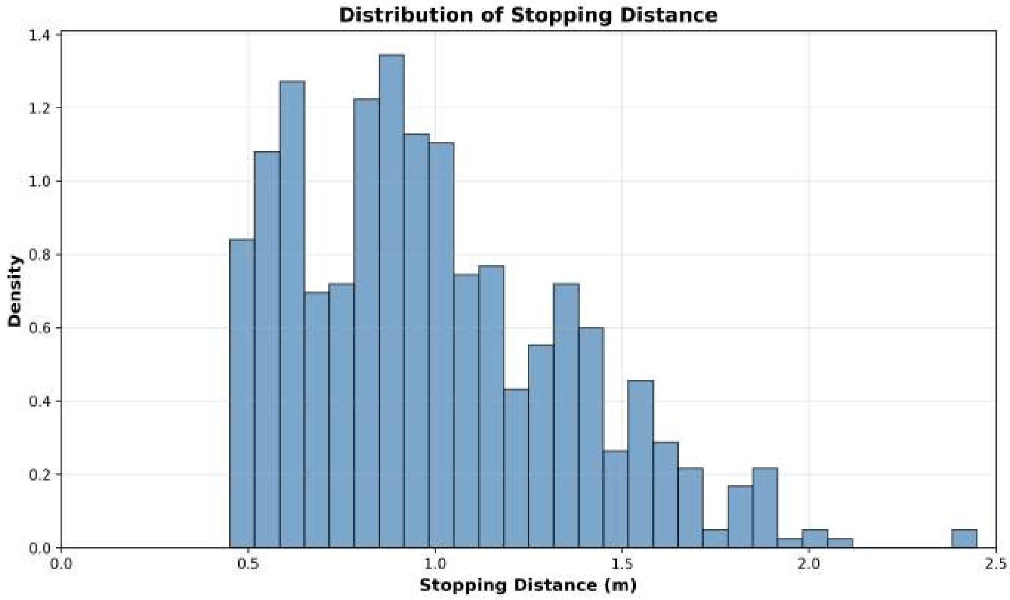


Figure 6.5: Stopping distance distribution for the Vietnamese sample.

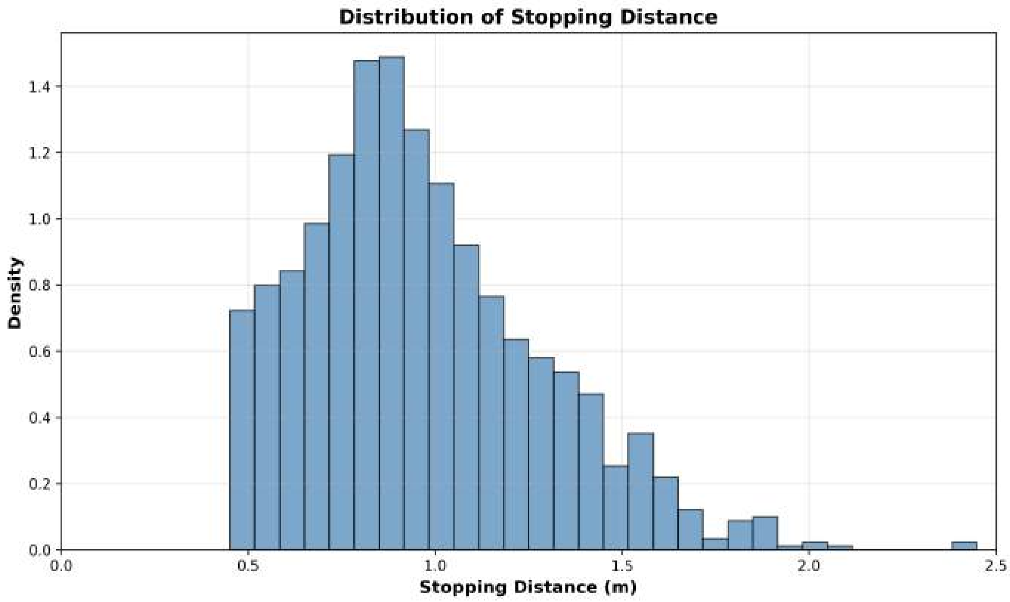


Figure 6.6: Stopping distance distribution for the full sample.

Figures 6.7 and 6.8 show the Pearson correlations between angles in both subsamples, demonstrating that Vietnamese have much larger deviations between some angles, reaching a lowest correlation of 0.19 compared to the Swedish lowest of 0.64. Figure 6.9 highlights the high correlation between angles in the general dataset. Figure 6.10 displays the distribution of stopping distances, capturing both variation across trials for individual subjects and broader differences between subjects. The standard deviation between subjects is more than twice as large as that observed within individual subjects, emphasizing that personal preferences vary much more than a single user's response to different approach angles.

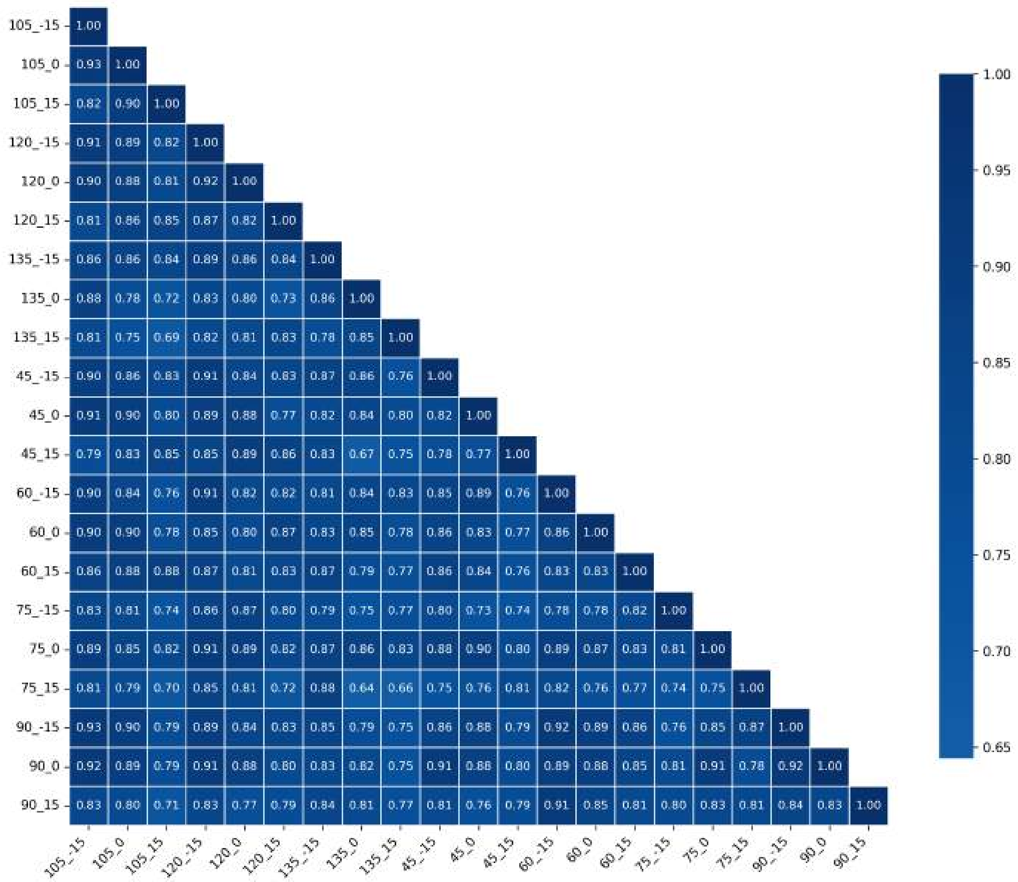


Figure 6.7: Correlations between approach angles based on stopping distances in the Swedish dataset.

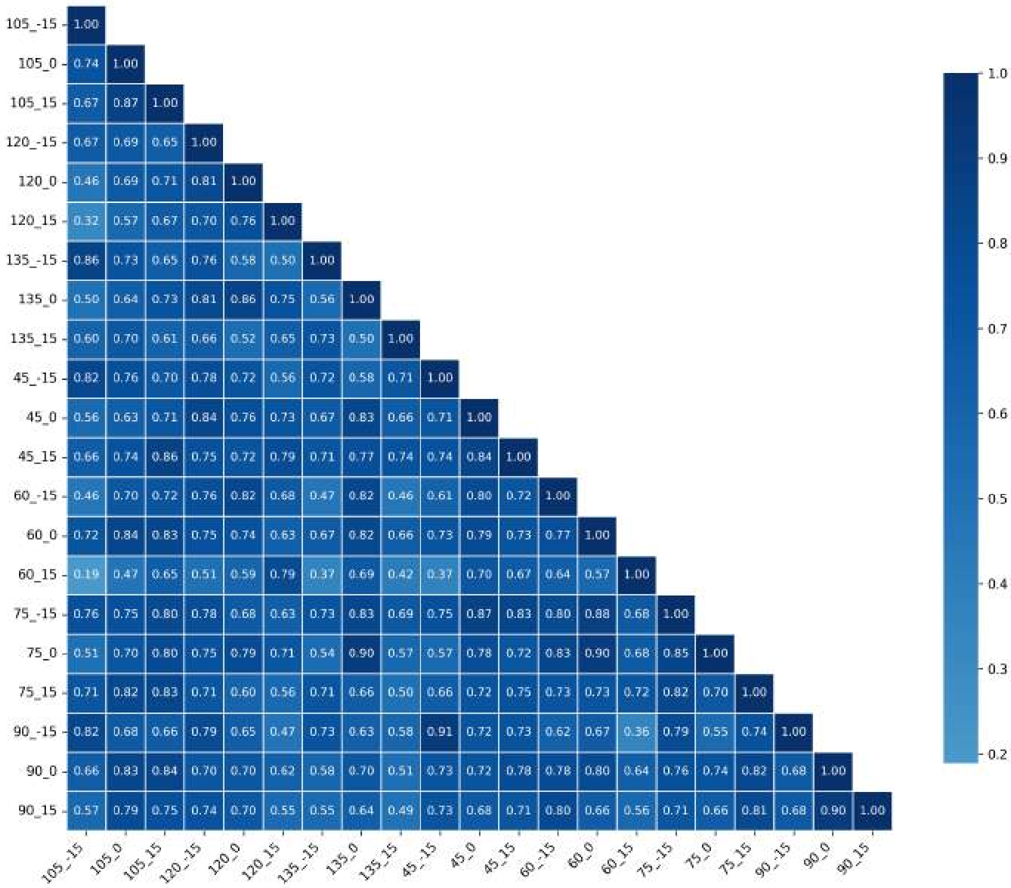


Figure 6.8: Correlations between approach angles based on stopping distances in the Vietnamese dataset.

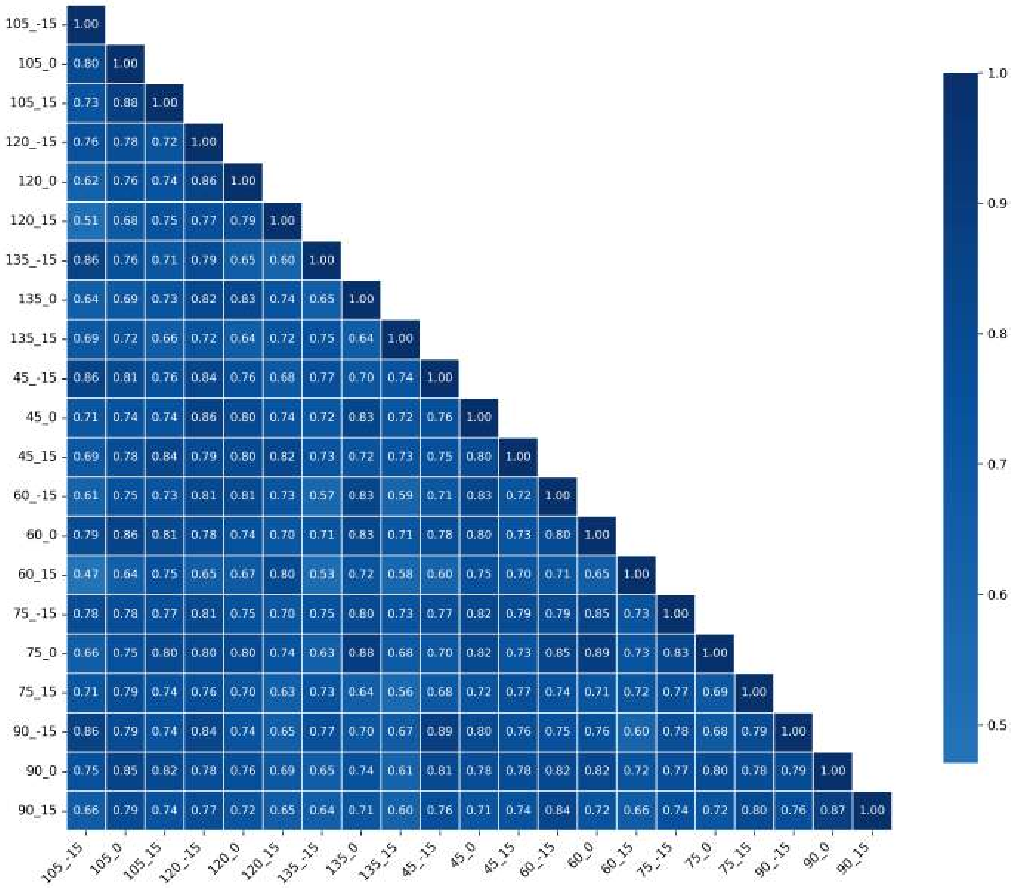


Figure 6.9: Correlations between approach angles based on stopping distances in the full dataset.

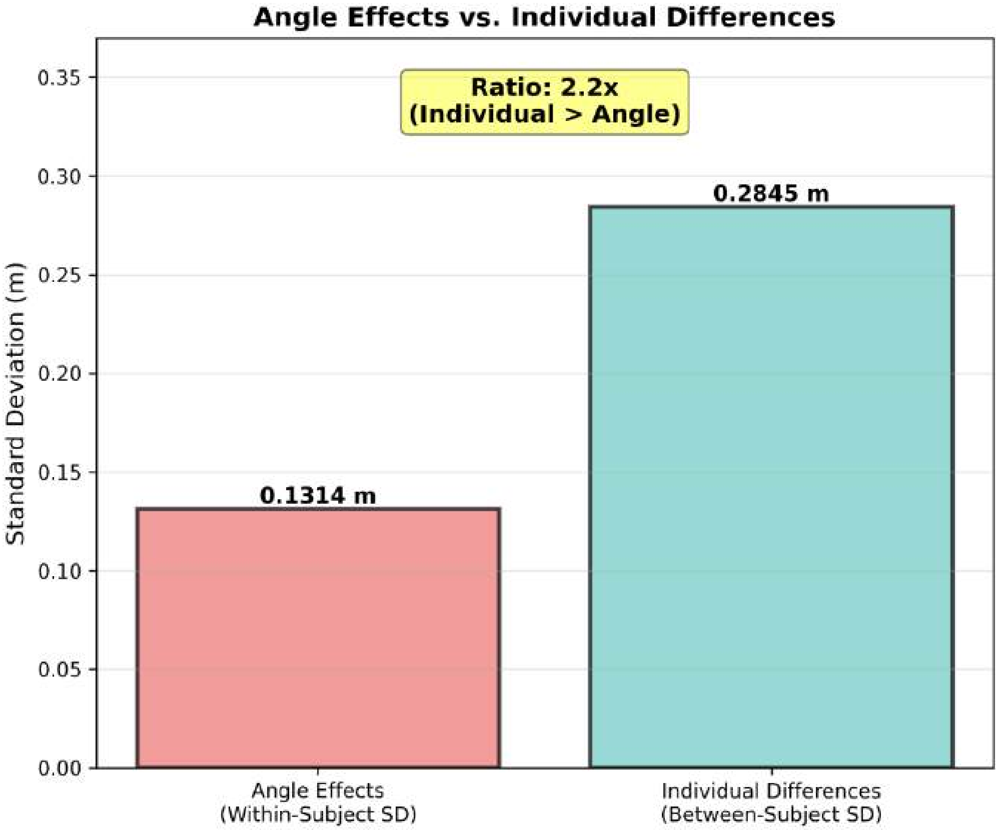


Figure 6.10: The standard deviation between stopping distances at different approach angles for the same subject (left) compared to the standard deviation of stopping distances between subjects (right).

6.3 Statistical Analysis

6.3.1 Main Effects (Full Sample)

Analyses of the full sample ($N = 69$) revealed no statistically significant main effects across any of the six primary variables when evaluated against the adjusted significance threshold ($\alpha = .0083$). The preferred stopping distances remained statistically uniform regardless of culture, gender, pet ownership, student status, age or height.

- **Culture:** The difference in stopping distances between Swedish and Vietnamese participants was not statistically significant ($U = 547$, $p = .592$, $r = -0.076$, negligible effect).

  Swedish ($N = 37$): $Mdn = 0.885$ m, $M = 0.929$ m, $SD = 0.254$

  Vietnamese ($N = 32$): $Mdn = 0.948$ m, $M = 0.969$ m, $SD = 0.319$

- **Gender:** No significant differences were observed between male and female

Table 6.1: Group sizes.

Variable	Group	N
Culture	Swedish	37
Culture	Vietnamese	32
Gender	Male	43
Gender	Female	26
Pet ownership	No pets	25
Pet ownership	Has pets	44
Student status	Non-student	24
Student status	Student	45
Age group	Gen Z ($\leq$ 29)	60
Age group	Millennial (30–45)	5
Age group	Gen X (46+)	4
Height group	Short (< 168 cm)	23
Height group	Medium (168–175 cm)	24
Height group	Tall ($\geq$ 175 cm)	22

participants ($U = 652$, $p = .252$, $r = +0.166$, small effect).

Male ($N = 43$): $Mdn = 0.912$ m, $M = 0.984$ m, $SD = 0.288$

Female ($N = 26$): $Mdn = 0.908$ m, $M = 0.888$ m, $SD = 0.274$

- **Pet Ownership:** Ownership of domestic pets did not significantly alter preferred boundaries ($U = 481$, $p = .393$, $r = -0.125$, small effect).

  No Pets ($N = 25$): $Mdn = 0.885$ m, $M = 0.903$ m, $SD = 0.245$

  Has Pets ($N = 44$): $Mdn = 0.919$ m, $M = 0.974$ m, $SD = 0.304$

- **Student Status:** No significant main effect was found for student status ($U = 484$, $p = .484$, $r = -0.104$, small effect).

  Non-student ($N = 24$): $Mdn = 0.902$ m, $M = 0.913$ m, $SD = 0.294$

  Student ($N = 45$): $Mdn = 0.928$ m, $M = 0.967$ m, $SD = 0.281$

- **Age Group:** The effect of age group was not statistically significant ($H = 3.288$, $p = .193$, $\eta^2 = .020$, small effect). The Millennial and Gen X subgroups fell below cell size $N = 5$, and their values should be interpreted with caution.

  Gen Z ($N = 60$): $Mdn = 0.926$ m, $M = 0.969$ m, $SD = 0.281$

  Millennial ($N = 5$): $Mdn = 0.634$ m, $M = 0.733$ m, $SD = 0.254$

  Gen X ($N = 4$, *caution*): $Mdn = 0.783$ m, $M = 0.905$ m, $SD = 0.325$

- **Height Group:** Categorisation by height tertiles yielded no significant differences ($H = 1.202$, $p = .548$, $\eta^2 = -.012$, negligible effect).

  Short (<168 cm, $N = 23$): $Mdn = 0.924$ m, $M = 0.918$ m, $SD = 0.301$

  Medium (168–175 cm, $N = 24$): $Mdn = 0.936$ m, $M = 1.007$ m, $SD = 0.302$

  Tall ($\geq$175 cm, $N = 22$): $Mdn = 0.880$ m, $M = 0.915$ m, $SD = 0.248$

6.3.2 Pairwise Demographic Interactions

Exploratory evaluations of all ten pairwise demographic combinations revealed no statistically significant interaction effects across the strata. Most pairs yielded negligible, uniform effects. For full reporting, see Appendix C. Two descriptive interaction patterns were identified:

- **Pet Ownership and Gender Interaction:** A directional reversal was observed when examining pet ownership within distinct gender categories. Among male participants, pet owners preferred larger distances than non-owners ($M = 1.030$ m vs. $Mdn = 0.864$ m, $r = -.33$, medium effect). Conversely, among female participants, non-owners preferred larger distances than pet owners ($M = 0.938$ m vs. $M = 0.838$ m, $r = +.29$, small effect). Neither of these localised variations reached statistical significance ($p = .096$ for males; $p = .218$ for females), but they highlight an opposing descriptive trend between genders.
- **Age and Student Status Interaction:** Among non-student participants, a weak descriptive trend indicated that Gen Z individuals stopped the drone farther away than their Millennial and Gen X counterparts ($H = 2.796$, $p = .247$, $\eta^2 = .04$, small effect); however, this was not statistically significant. Among student participants, an interaction analysis could not be conducted because all 45 students fell within the Gen Z age group, eliminating the necessary variance for comparison.

6.3.3 Cross-cultural Analysis

When analyses were repeated separately within each cultural group, no statistically significant interaction between culture and any demographic variable was found. However, descriptive differences in the patterns between Swedish and Vietnamese participants were observed.

- **Gender Stratified by Culture:** Gender had no meaningful effect within the Swedish subsample, but a significant divergence emerged among Vietnamese participants, with males maintaining larger distances than females ($U = 151$, $p = .049$, $r = +0.459$, medium effect). However, it should be noted that the tested subgroups are imbalanced and that this effect is not considered significant under the Bonferroni correction described in Eq. 5.1.

Swedish ($N = 37$): $U = 164$, $p = .867$, $r = -0.035$ (negligible)

Male ($N = 20$): $Mdn = 0.880$ m, $M = 0.920$ m, $SD = 0.228$

Female ($N = 17$): $Mdn = 0.892$ m, $M = 0.941$ m, $SD = 0.289$

Vietnamese ($N = 32$): $U = 151$, $p = .049$ (significant), $r = +0.459$ (medium)

Male ($N = 23$): $Mdn = 0.984$ m, $M = 1.040$ m, $SD = 0.325$

Female ($N = 9$): $Mdn = 0.924$ m, $M = 0.789$ m, $SD = 0.227$

- **Pet Ownership Stratified by Culture:** Pet ownership showed a negligible effect in the Swedish subsample. In Vietnam, pet owners descriptively maintained a larger buffer zone than non-owners, though this did not reach significance (medium effect).

Swedish ($N = 37$): $U = 158$, $p = .808$, $r = +0.053$ (negligible)

No Pets ($N = 12$): $Mdn = 0.920$ m, $M = 0.938$ m, $SD = 0.244$

Has Pets ($N = 25$): $Mdn = 0.881$ m, $M = 0.925$ m, $SD = 0.264$

Vietnamese ($N = 32$): $U = 86$, $p = .156$, $r = -0.304$ (medium)

No Pets ($N = 13$): $Mdn = 0.873$ m, $M = 0.870$ m, $SD = 0.252$

Has Pets ($N = 19$): $Mdn = 1.069$ m, $M = 1.037$ m, $SD = 0.347$

- **Student Status Stratified by Culture:** Student status was non-significant in both cultural subsamples. Descriptively, Vietnamese students maintained larger distances than non-students, while the pattern in Sweden was negligible and reversed.

Swedish ($N = 37$): $U = 173$, $p = .939$, $r = +0.018$ (negligible)

Non-student ($N = 17$): $Mdn = 0.912$ m, $M = 0.946$ m, $SD = 0.325$

Student ($N = 20$): $Mdn = 0.878$ m, $M = 0.915$ m, $SD = 0.182$

Vietnamese ($N = 32$): $U = 63$, $p = .281$, $r = -0.280$ (small)

Non-student ($N = 7$): $Mdn = 0.892$ m, $M = 0.832$ m, $SD = 0.195$

Student ($N = 25$): $Mdn = 0.971$ m, $M = 1.008$ m, $SD = 0.338$

- **Age Stratified by Culture:** Age effects were non-significant in both subsamples. All non-Gen Z groups fell below cell size $N = 5$ and must be interpreted with caution. In the Vietnamese subsample, Gen X was absent ($N = 0$), reducing the comparison to Gen Z versus Millennial ($N = 2$); the borderline result ($p = .062$) cannot support inferential conclusions at that cell size.

Swedish ($N = 37$): $H = 0.566$, $p = .754$, $\eta^2 = -.042$ (negligible)

Gen Z ($N = 30$): $Mdn = 0.889$ m, $M = 0.942$ m, $SD = 0.249$

Millennial ($N = 3$, *caution*): $Mdn = 0.912$ m, $M = 0.830$ m, $SD = 0.303$

Gen X ($N = 4$, *caution*): $Mdn = 0.783$ m, $M = 0.905$ m, $SD = 0.325$

Vietnamese ($N = 32$): $H = 3.491$, $p = .062$, $\eta^2 = +.08$ (medium, borderline) — Gen X excluded ($N = 0$)

Gen Z ($N = 30$): $Mdn = 0.960$ m, $M = 0.995$ m, $SD = 0.312$

Millennial ($N = 2$, *caution*): $Mdn = 0.587$ m, $M = 0.587$ m, $SD = 0.066$

Gen X ($N = 0$, *cannot be tested*)

- **Height Stratified by Culture:** Height tertiles did not significantly influence stopping distances in either cultural group. Medium-height Vietnamese participants maintained the largest distances, though this pattern was not statistically significant.

Swedish ($N = 37$): $H = 0.598$, $p = .742$, $\eta^2 = -.041$ (negligible)

Short ($N = 11$): $Mdn = 0.956$ m, $M = 0.983$ m, $SD = 0.314$

Medium ($N = 9$): $Mdn = 0.881$ m, $M = 0.927$ m, $SD = 0.258$

Tall ($N = 17$): $Mdn = 0.874$ m, $M = 0.896$ m, $SD = 0.218$

Vietnamese ($N = 32$): $H = 2.388$, $p = .303$, $\eta^2 = +.01$ (small)

Short ($N = 12$): $Mdn = 0.894$ m, $M = 0.858$ m, $SD = 0.288$

Medium ($N = 15$): $Mdn = 1.069$ m, $M = 1.054$ m, $SD = 0.324$

Tall ($N = 5$): $Mdn = 0.984$ m, $M = 0.981$ m, $SD = 0.354$

6.4 Rubber-Bubble Model

The rubber-bubble models are presented here, showing the difference in the models by culture and gender. The results cover the Swedish models first, followed by the Vietnamese.

6.4.1 Swedish

6.4.1.1 No subject-aware grouping

Table 6.2 shows the results of personalized hold-out modeling. It shows high $R^2 > 0.8$, plateuing around five degrees, and excluding angles in all of the best models.

Table 6.2: Swedish: Top 10 Models by R^2. Both genders. Hold-out modeling. Compared all feature combinations from subset one, degrees one through ten with OLS.

Degree	Model	R^2	RMSE	Features
10	LinearRegression	0.8349	0.1131	height, age
9	LinearRegression	0.8349	0.1131	height, age
7	LinearRegression	0.8349	0.1131	height, age, gender
8	LinearRegression	0.8349	0.1131	height, age, gender
10	LinearRegression	0.8349	0.1131	height, age, gender
6	LinearRegression	0.8349	0.1131	height, age, gender
9	LinearRegression	0.8349	0.1131	height, age, gender
8	LinearRegression	0.8349	0.1131	height, age
5	LinearRegression	0.8204	0.1179	height, age, gender
7	LinearRegression	0.8203	0.1179	height, age

Table 6.3 shows that models of order lower than five fail to match the high scores achieved by a fifth-degree model, and that fifth-degree models that do not incorporate gender also drop in R^2 score.

Table 6.3: Swedish: Top 7 Models by R^2. Both genders. Hold-out modeling. Compared all feature combinations from subset one, degrees one through five with OLS.

Degree	Model	R^2	RMSE	Features
5	LinearRegression	0.8204	0.1179	height, age, gender
5	LinearRegression	0.8020	0.1238	vertical angle, height, age, gender
5	LinearRegression	0.7974	0.1253	horizontal angle, height, age, gender
5	LinearRegression	0.5141	0.1940	height, age
5	LinearRegression	0.4773	0.2012	vertical angle, height, age
5	LinearRegression	0.4738	0.2018	horizontal angle, height, age
4	LinearRegression	0.4307	0.2099	height, age, gender

Table 6.4 provides an illustrative snapshot of the fitted coefficients for various second-order models. This demonstrates that the second-order rubber-bubble models generally yield better performance when trained and tested separately on gender-stratified data rather than the combined dataset.

Introducing demographic features alters model performance. For instance, adding height to the angle models improves metrics for both, giving males a slightly positive R^2 (RMSE = 0.2223, R^2 = 0.0824) but fails to create a positive fit for females (RMSE = 0.2546, $R^2 = -0.0195$). Conversely, adding age to the angles benefits the female group (RMSE = 0.2335, R^2 = 0.1421) but degrades the male model (RMSE = 0.2493, $R^2 = -0.1539$).

Ultimately, the best performing model entirely excludes the approach angles, combining only height and age. This configuration provides the lowest error and highest coefficient of determination for both female data (RMSE = 0.2200, R^2 = 0.2385) and male data (RMSE = 0.2075, R^2 = 0.2002).

Table 6.4: Swedish: Model coefficients for 2^{nd} order rubber-bubble models stratified by gender with standardized features. V = Vertical Angle and H = Horizontal Angle. ↓ indicates lower is better, ↑ indicates higher is better.

Model Features	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**j**	**k**	**RMSE (↓)**	**R^2 (↑)**
$\{V, H\}_{\text{Both}}$	0.9323	0.0052	0.0401	0.0088	-0.0034	-0.0348					0.2816	-0.0241
$\{V, H, \text{Height}\}_{\text{Both}}$	0.9323	-0.0113	-0.1463	0.3025	0.0073	-0.0054	0.0195	-0.0381	0.1962	-0.3873	0.2773	0.0069
$\{V, H, \text{Age}\}_{\text{Both}}$	0.9323	-0.0170	0.0335	0.0420	0.0094	-0.0046	0.0253	-0.0404	0.0197	-0.0702	0.2858	-0.0553
$\{\text{Height}, \text{Age}\}_{\text{Both}}$	0.9323	0.5548	0.0840	-0.6124	0.0323	-0.1315					0.2787	-0.0032
$\{V, H\}_{\text{Female}}$	0.9733	0.0272	0.0774	-0.0050	-0.0151	-0.0932					0.2575	-0.0431
$\{V, H, \text{Height}\}_{\text{Female}}$	0.9733	0.2666	-0.5838	-5.5190	-0.0055	-0.0299	-0.2229	-0.0754	0.6498	5.3970	0.2546	-0.0195
$\{V, H, \text{Age}\}_{\text{Female}}$	0.9733	0.0072	0.1321	0.9023	0.0040	0.0062	0.0048	-0.1259	-0.0355	-0.9518	0.2335	0.1421
$\{\text{Height}, \text{Age}\}_{\text{Female}}$	0.9733	-5.6307	1.6118	5.6906	-0.8438	-0.8572					0.2200	0.2385
$\{V, H\}_{\text{Male}}$	0.9232	0.0055	-0.1096	0.0007	-0.0051	0.1226					0.2350	-0.0257
$\{V, H, \text{Height}\}_{\text{Male}}$	0.9232	0.3073	-0.2083	1.6205	-0.0014	-0.0084	-0.3039	0.1121	0.1087	-1.7066	0.2223	0.0824
$\{V, H, \text{Age}\}_{\text{Male}}$	0.9232	0.0175	-0.0502	-0.3591	0.0027	-0.0111	-0.0101	0.1057	-0.0650	0.4868	0.2493	-0.1539
$\{\text{Height}, \text{Age}\}_{\text{Male}}$	0.9232	4.1379	8.2398	-3.5975	-7.4805	-0.5713					0.2075	0.2002

Table 6.5 shows 5-fold cross validation on the Swedish subset showing similar results to the hold-out method used previously due to the high correlation between subject angles. Up to 20-fold cross validation was tried with similar results.

Table 6.5: Swedish: Top 3 Models by R^2 for degrees 2, 5, and 10 using 5-Fold cross-validation. Both genders. Compared all feature combinations from subset one using OLS.

Degree	Model	R^2	RMSE	Features
10	LinearRegression	0.8136	0.1129	height, age, gender
5	LinearRegression	0.7880	0.1205	height, age, gender
2	LinearRegression	0.0769	0.2536	height, age, gender

6.4.1.2 Subject-aware grouping

Tables 6.6, 6.7, and 6.8 show the results of modeling on Swedish data using LOSOCV on subset two to evaluate the models' overall generalizability with all objective parameters available. Any models with $R^2 < 0$ perform worse than simply picking the mean stopping distance for the dataset each time. All of the models, both the ones stratifying by gender and the ones using both genders, give negative scores, and thus do not generalize to new subjects. The results were the same when removing all subjects without all angles left in their datasets.

When evaluating the Swedish subset using both genders across subset two, the predictive models demonstrated a severe lack of generalizability, marked by highly negative R^2 values (see Table 6.6). The highest-performing feature combination within this unstratified group linked horizontal angle and height, yielding a peak R^2 of -15.4786 and an RMSE of 0.2570 on a third degree model. Across all these models, height consistently was incorporated as a feature, though these R^2 scores are so low that it is difficult to draw any reasonable conclusions from this.

Table 6.6: Swedish: Top 5 Models by R^2. Both genders. Compared all feature combinations from subset two, degrees one through five with OLS, Ridge, and Lasso models using LOSOCV.

Deg	**Model**	R^2	**RMSE**	**Features**
3	LinearRegression	-15.4786	0.2570	horizontal angle, height
3	LinearRegression	-15.5081	0.2567	height
3	LinearRegression	-15.5215	0.2569	vertical angle, height
3	LinearRegression	-15.5384	0.2577	horizontal angle, vertical angle, height
4	LinearRegression	-15.7740	0.2614	vertical angle, height

Isolating the male Swedish subset forced the model selection toward simpler, first-degree models. While this reduction in polynomial complexity mitigated some variance compared to the other strata, test metrics remained negative. The best-performing configuration combined horizontal angle with height, both with and without the inclusion of gender, establishing a performance ceiling of $R^2 = -8.3183$ and an RMSE of 0.2339.

Table 6.7: Swedish: Top 5 Models by R^2. Males only. Compared all feature combinations from subset two, degrees one through five with OLS, Ridge, and Lasso models using LOSOCV.

Deg	Model	R^2	RMSE	Features
1	LinearRegression	-8.3183	0.2339	horizontal angle, height
1	LinearRegression	-8.3183	0.2339	horizontal angle, height, gender
1	LinearRegression	-8.3217	0.2339	height
1	LinearRegression	-8.3217	0.2339	height, gender
1	LinearRegression	-8.3305	0.2341	horizontal angle, vertical angle, height

The female Swedish subset suffered from the most pronounced generalization deficits, resulting in the poorest overall predictive scores within this feature set. Second-degree polynomial equations proved optimal but also failed to yield viable predictions. The best configuration utilized height paired with gender (as well as height isolated), securing a top R^2 of only -22.9021 and an RMSE of 0.3071.

Table 6.8: Swedish: Top 5 Models by R^2. Females only. Compared all feature combinations from subset two, degrees one through five with OLS, Ridge, and Lasso models using LOSOCV.

Deg	Model	R^2	RMSE	Features
2	LinearRegression	-22.9021	0.3071	height, gender
2	LinearRegression	-22.9021	0.3071	height
2	LinearRegression	-22.9063	0.3072	vertical angle, height
2	LinearRegression	-22.9063	0.3072	vertical angle, height, gender
2	LinearRegression	-22.9733	0.3078	horizontal angle, height

6.4.2 Vietnamese

6.4.2.1 No subject-aware grouping

Utilizing subset one using OLS across high-degree polynomials (degrees one through ten) revealed positive R^2 values. The highest performance was achieved at higher complexities, specifically degrees eight through ten, where a combination of vertical angle, age, and height reached a peak R^2 of 0.4516 and an RMSE of 0.2620. Notably, the feature combination of vertical angle, age, and height consistently dominated the top four rankings, demonstrating that vertical angles may have a larger impact in this subset.

Table 6.9: Vietnamese: Top 10 Models by R^2. Both genders. Compared all feature combinations from subset one, degrees one through ten with OLS.

Degree	Model	R^2	RMSE	Features
10	LinearRegression	0.4516	0.2620	Vertical angle, age, height
9	LinearRegression	0.4516	0.2620	Vertical angle, age, height
8	LinearRegression	0.4516	0.2620	Vertical angle, age, height
7	LinearRegression	0.4409	0.2646	Vertical angle, age, height
7	LinearRegression	0.4401	0.2648	Vertical angle, age, gender, height
10	LinearRegression	0.4401	0.2648	Vertical angle, age, gender, height
8	LinearRegression	0.4401	0.2648	Vertical angle, age, gender, height
9	LinearRegression	0.4401	0.2648	Vertical angle, age, gender, height
6	LinearRegression	0.4301	0.2671	Vertical angle, age, gender, height
9	LinearRegression	0.3998	0.2741	Age, gender, height

As in the Swedish subset, below are examples of second order personalized rubber-bubble models. Models entirely dependent on angular variables ($\{V, H\}$) suffered from poor generalization, resulting in negative R^2 values across all cohorts. However, the inclusion of demographics increased model accuracy. The combination of height and age ($\{\text{Height}, \text{Age}\}$) proved to be the most powerful predictor across the board. This was most pronounced within the female subset ($\{\text{Height}, \text{Age}\}_{\text{Female}}$), which captured the highest score with an R^2 of 0.2718 and a minimized RMSE of 0.2217.

Table 6.10: Vietnamese: Model coefficients for 2^{nd} order rubber-bubble models stratified by gender with standardized features. V = Vertical Angle and H = Horizontal Angle. ↓ indicates lower is better, ↑ indicates higher is better.

Model Features	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**j**	**k**	**RMSE (↓)**	**R^2 (↑)**
$\{V, H\}_{\text{Both}}$	0.9994	-0.0284	-0.0018	0.0233	0.0024	-0.0020					0.3570	-0.0176
$\{V, H, \text{Height}\}_{\text{Both}}$	0.9994	0.1999	-0.6935	2.6138	0.0259	-0.0047	-0.2214	0.0131	0.6815	-2.6312	0.3533	0.0031
$\{V, H, \text{Age}\}_{\text{Both}}$	0.9994	-0.0871	0.0198	0.1787	0.0163	-0.0017	0.0631	0.0218	-0.0452	-0.2558	0.3415	0.0688
$\{\text{Height}, \text{Age}\}_{\text{Both}}$	0.9994	4.0403	1.6394	-3.6714	-1.3119	-0.4459					0.3311	0.1242
$\{V, H\}_{\text{Female}}$	0.8234	0.0007	0.0205	0.0401	-0.0246	-0.0243					0.2626	-0.0220
$\{V, H, \text{Height}\}_{\text{Female}}$	0.8234	-0.3556	0.3757	4.2906	0.0385	-0.0167	0.3571	-0.1165	-0.2753	-4.2984	0.2555	0.0330
$\{V, H, \text{Age}\}_{\text{Female}}$	0.8234	0.0174	-0.0139	-0.1007	0.0211	-0.0060	-0.0319	0.0231	-0.0084	0.0148	0.2465	0.0996
$\{\text{Height}, \text{Age}\}_{\text{Female}}$	0.8234	-0.4780	-6.5035	-0.5781	6.1050	0.1286					0.2217	0.2718
$\{V, H\}_{\text{Male}}$	1.0695	-0.0166	0.0678	-0.0019	-0.0072	-0.0779					0.3580	-0.0362
$\{V, H, \text{Height}\}_{\text{Male}}$	1.0695	-0.3330	-0.6932	3.0674	0.0031	-0.0161	0.3259	-0.0957	0.7816	-3.1334	0.3538	-0.0117
$\{V, H, \text{Age}\}_{\text{Male}}$	1.0695	-0.0922	0.0090	-0.4174	-0.0018	-0.0424	0.1095	-0.0623	0.0463	0.2686	0.3496	0.0118
$\{\text{Height}, \text{Age}\}_{\text{Male}}$	1.0695	4.0611	0.5260	-3.9335	-0.8298	0.1767					0.3361	0.0865

Table 6.11 shows the best models for a 5-fold cross validation on the Vietnamese subset over the degrees two, five, and ten. It shows slightly better results than the fold that the standard hold-out modeling provided.

Table 6.11: Vietnamese: Top 3 Models by R^2 for degrees 2, 5, and 10 using 5-Fold cross-validation. Both genders. Compared all feature combinations from subset one using OLS.

Degree	Model	R^2	RMSE	Features
10	LinearRegression	0.5817	0.2339	height, age, gender
5	LinearRegression	0.5817	0.2339	height, age, gender
2	LinearRegression	0.2128	0.3221	height, age, gender

6.4.2.2 Subject-aware grouping (LOSOCV)

The predictive performance for the entire Vietnamese dataset across subset two yielded generally lower, negative R^2 values under LOSOCV, indicating cross-subject variability and low predictive capabilities for the chosen variables (see Table 6.12). Second-degree models consistently outperformed first-degree variants. The top-performing configuration utilized a combination of vertical angle, age, and gender, achieving the highest R^2 of -8.3361 and an RMSE of 0.3218. The demographic features age and gender appeared as in all five top-ranking models.

Table 6.12: Vietnamese: Top 5 Models by R^2. Both genders. Compared all feature combinations from subset two, degrees one through five with OLS, Ridge, and Lasso models using LOSOCV.

Degree	Model	R^2	RMSE	Features
2	LinearRegression	-8.3361	0.3218	vertical angle, age, gender
2	LinearRegression	-8.3503	0.3213	age, gender
2	LinearRegression	-8.4196	0.3228	horizontal, vertical angle, age, gender
2	LinearRegression	-8.4271	0.3221	horizontal angle, age, gender
1	LinearRegression	-8.4437	0.3200	horizontal, vertical angle, age, gender

When isolating the male Vietnamese subjects, rubber-bubble models demonstrated a slight improvement in error rates, though R^2 scores remained negative (see Table 6.13). Quadratic implementations dominated the top five models. The best configuration used horizontal angle, height, and pet ownership ($R^2 = -7.1835$, RMSE = 0.3432).

Table 6.13: Vietnamese: Top 5 Models by R^2. Male only. Compared all feature combinations from subset two, degrees one through five with OLS, Ridge, and Lasso models using LOSOCV.

Degree	Model	R^2	RMSE	Features
2	LinearRegression	-7.1835	0.3432	horizontal angle, height, pets
2	LinearRegression	-7.1835	0.3432	horizontal angle, height, gender, pets
2	LinearRegression	-7.1910	0.3426	height, pets
2	LinearRegression	-7.1910	0.3426	height, gender, pets
2	LinearRegression	-7.2213	0.3446	horizontal angle, vertical angle, height, pets

The female-only Vietnamese subgroup yielded the strongest overall predictive accuracy within subset two, producing less negative R^2 values and lower RMSE compared to both the male and combined gender models. Second-degree models proved optimal for this subset as well. The highest predictive accuracy was achieved by combining the vertical angle, height, and pet ownership features, yielding a top R^2 of -5.1400 and an RMSE of 0.1973.

Table 6.14: Vietnamese: Top 5 Models by R^2. Female only. Compared all feature combinations from subset two, degrees one through five with OLS, Ridge, and Lasso models using LOSOCV.

Degree	Model	R^2	RMSE	Features
2	LinearRegression	-5.1400	0.1973	vertical angle, height, pets
2	LinearRegression	-5.1400	0.1973	vertical angle, height, gender, pets
2	LinearRegression	-5.2806	0.1965	height, gender, pets
2	LinearRegression	-5.2806	0.1965	height, pets
2	LinearRegression	-5.4267	0.2031	horizontal angle, vertical angle, height, pets

6.5 Linear Mixed Model

Table 6.15 presents the fixed effects for the full subset two model. None of the individual subset two variables reached statistical significance ($\alpha = .05$). Gender was closest to this threshold, with female participants maintaining a mean stopping distance approximately 0.171 m shorter than male participants ($\beta = -0.171, p = .068$). Approach angles, age, height, pet ownership, and student status all showed negligible associations.

Table 6.15: Fixed effects estimates for the full LMM (subset two, $N = 69$).

Predictor	$\hat{\beta}$	**SE**	z	p	**95% CI**
Intercept	2.376	0.888	2.674	.008**	[0.634, 4.117]
Horizontal angle	0.000	0.000	0.443	.658	[−0.000, 0.000]
Vertical angle	−0.000	0.000	−0.596	.551	[−0.001, 0.000]
Age	−0.003	0.005	−0.677	.498	[−0.013, 0.006]
Height (cm)	−0.008	0.005	−1.542	.123	[−0.017, 0.002]
Gender	−0.171	0.094	−1.822	.068	[−0.356, 0.013]
Pet ownership	0.066	0.074	0.886	.376	[−0.080, 0.211]
Student status	−0.013	0.087	−0.151	.880	[−0.184, 0.158]

$^{**}p < .01$. $N = 69$ subjects, 1372 observations.

Table 6.16 shows the results of the LMM fitted on subset two. The ICC shows that differences between participants accounted for 75.7% of the total variance in stopping distances, whereas within-person trial-to-trial noise accounted for the remaining 24.3%. The fixed effect predictors (subset two) explained only 5.7% of overall variance (marginal $R^2 = 0.057$). LOSOCV evaluation yielded an RMSE of 0.294 m whereas the standard deviation of the sample is 0.317, confirming that it fails to meaningfully generalize to unseen individuals.

Table 6.16: Variance components and fit statistics for the LMM model (subset two, $N = 69$).

Statistic	**Value**
Random intercept variance (between subjects)	0.079
Residual variance (within subject)	0.025
ICC	0.757
Marginal R^2	0.057
Conditional R^2	0.771
AIC (ML)	−856.84
RMSE (m)	0.294

Table 6.17 compares performance across three specifications. Model optimization via AIC favored the angles-only model, indicating that incorporating objective predictors increases complexity without improving model fit.

Table 6.17: Comparison of three LMM specifications. All models include a random intercept per participant. AIC computed using maximum likelihood (ML). Lower AIC and lower RMSE indicate better performance.

Model	**Marg.** R^2	**Cond.** R^2	**AIC**	**RMSE (m)**
Angles only	0.000	0.759	$-861.01^{\dagger}$	0.284
Subset two	0.057	0.771	-856.84	0.294
Objective parameters only	0.057	0.771	-860.28	0.293

†Best AIC. All models include a random intercept per subject.

6.6 Bayesian Model

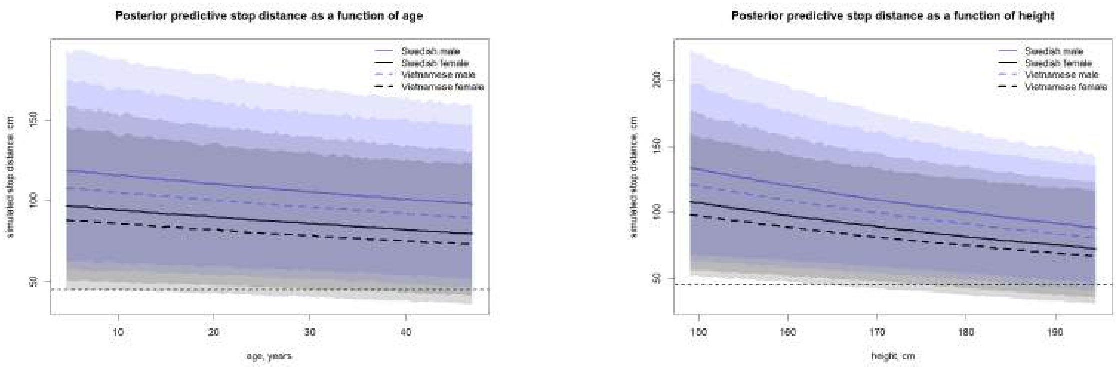
Table 6.18: Final Bayesian model population-level parameter diagnostics.

Parameter	Mean	SD	5.5%	94.5%	$\widehat{R}$	n_{eff}
a	4.6326	0.0730	4.5131	4.7495	1.0056	502.4
bG	-0.2057	0.0962	-0.3509	-0.0454	1.0069	477.0
bC	-0.0935	0.0880	-0.2310	0.0484	1.0042	480.5
bA	-0.0378	0.0356	-0.0957	0.0189	1.0051	1138.8
bH	-0.0811	0.0520	-0.1604	0.0054	1.0040	651.2
σ_{subject}	0.2910	0.0259	0.2525	0.3342	1.0016	3107.7
scale	2.3906	0.0927	2.2491	2.5445	1.0027	3384.8

The model passed all diagnostics thresholds, as reported in full in Appendix D. Table 6.18 offers a summary.

The population-level intercept was estimated as a = 4.6326, with an 89% CI from 4.5131 to 4.7495. This corresponds to an expected stop distance of approximately 103 cm for the reference Swedish male at average age and height.

The model estimated a negative association for gender, bG = -0.2057, with an 89% CI from -0.3509 to -0.0454. Because the model used a log link, this corresponds to female participants having an estimated stop distance approximately 19% shorter than male participants, conditional on the model.



(a) Curve for age. **(b)** Curve for height.

Figure 6.11: Posterior predictive curves for continuous variables. The y-axis is truncated at 45 cm, as this represents the intimate-zone cutoff point used during data cleaning (see Sec. 6.1).

The remaining predictors were also negative on average, but their compatibility intervals included zero. The culture coefficient was bC = -0.0935, with an 89% CI from -0.2310 to 0.0484, suggesting an estimated lower stop distance for Vietnamese participants compared with Swedish participants, although with substantial uncertainty. The age coefficient was bA = -0.0378, with an interval from -0.0957 to 0.0189, and the height coefficient was bH = -0.0811, with an interval from -0.1604 to 0.0054. Thus, the model estimated shorter stop distances for participants who were older and for participants who were taller (see Figures 6.11a and 6.11b). However, the CI for both age and height overlapped zero, meaning that these associations were highly uncertain.

The estimated standard deviation of the subject-level intercepts was $\sigma_{subject}$ = 0.2910, with an 89% CI from 0.2525 to 0.3342. This indicates meaningful between-subject variation in baseline stop distance when accounting for gender, culture, age and height.

The Gamma scale parameter was estimated as scale = 2.3906, with an 89% CI from 2.2491 to 2.5445.

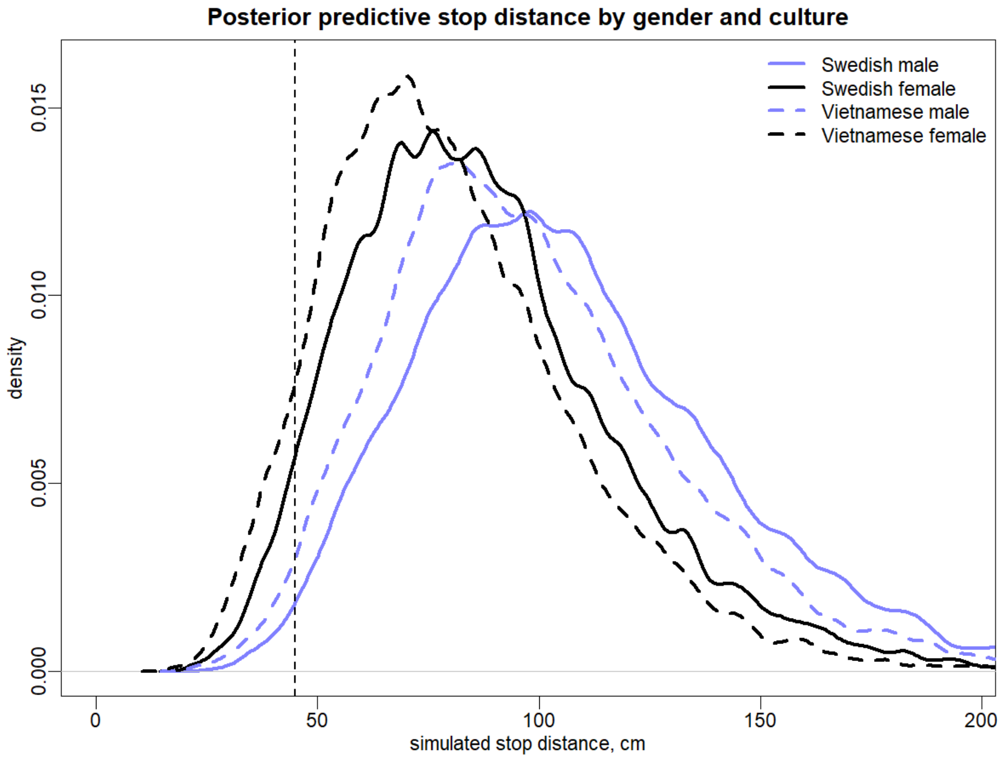


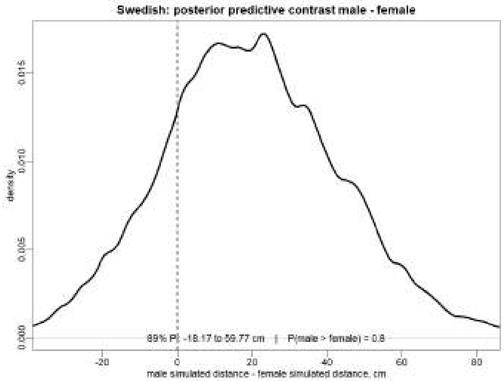
Figure 6.12: Posterior predictive stop distance by gender and culture. The x-axis is truncated at 45 cm, as this represents the intimate-zone cutoff used during data cleaning (see Sec. 6.1).

Table 6.19: Posterior predictive stop distances by gender and culture. Intervals represent 89% posterior intervals.

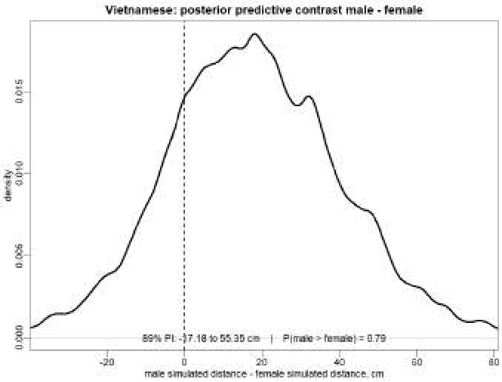
Group	Mean	5.5%	94.5%
Swedish male	107.50	57.48	171.59
Swedish female	87.72	46.26	143.64
Swedish male – female	19.78	-18.17	59.77
Vietnamese male	98.06	52.34	156.65
Vietnamese female	80.07	41.20	130.97
Vietnamese male – female	17.99	-17.18	55.35

Figure 6.12 visualizes the posterior predictive distributions, which showed a shift by gender. In both cultural groups, the distributions for males were shifted toward larger stop distances, while the distributions for females were shifted toward shorter stop distances. As can be seen in Table 6.19, Swedish participants had a mean posterior predictive stop distance of 107.50 cm for males and 87.72 cm for females. For Vietnamese participants, the corresponding values were 98.06 cm for males and 80.07 cm for females. Thus, the model predicted shorter stop distances for females than for males in both cultural groups.

The posterior predictive distributions also suggested lower predicted stop distances for Vietnamese participants compared to Swedish participants. Swedish males had the highest predicted stop distances, followed by Vietnamese males, Swedish females and Vietnamese females. However, the distributions overlapped substantially, indicating that the model does not separate these groups cleanly at the level of individual predicted outcomes.



(a) Swedish



(b) Vietnamese

Figure 6.13: Posterior predictive difference between males and females, with the dotted line representing no difference.

To examine the gender difference closer, posterior predictive contrasts were calculated as male simulated stop distance minus female simulated stop distance (see Figures 6.13a and 6.13b). Positive values indicate longer predicted stop distances for males than for females. For Swedish participants, the mean posterior predictive contrast was 19.78 cm, with an 89% prediction interval from -18.17 to 59.77 cm. The posterior probability that males had longer predicted stop distances than females was 0.80. For Vietnamese participants, the mean posterior predictive contrast was 17.99 cm, with an 89% prediction interval from -17.18 to 55.35 cm. The posterior probability that males had longer predicted stop distances than females was 0.79. The posterior predictive intervals for individual simulated outcomes included zero in both cultural groups. This indicates that although the model estimated a gender-based difference, there remained substantial uncertainty in predicted stop distances.

To evaluate whether the demographic predictors improved out-of-sample prediction, the full Bayesian model was compared with a null model using LOSOCV. The comparison was performed at the subject level to assess how well each model generalized to unseen participants. The full LOSO comparison is provided in Appendix E, while Table 6.20 summarizes the results.

Table 6.20: LOSO predictive-performance summary for the null and full Bayesian models.

Quantity	Value
Total LPD, null	-5804.50
Total LPD, full	-5808.28
Total deviance, null	11608.99
Total deviance, full	11616.57
Delta deviance, full–null	7.57
Mean per-subject delta deviance	0.1098
95% CI for mean delta deviance	[-0.0753, 0.2948]

The null model had a total log predictive density of -5804.50, compared with -5808.28 for the full model. On the deviance scale, the null model had a total deviance of 11608.99, while the full model had a total deviance of 11616.57. The $full - null$ difference in deviance was therefore +7.57, indicating slightly better predictive performance for the null model. However, the mean per-subject delta deviance was small, +0.1098, with a 95% interval of [-0.0753, 0.2948], which included zero. Thus, the LOSOCV comparison showed no clear predictive advantage for either model.

6.7 Post-Study Questions

6.7.1 Sweden

Figure 6.14 summarizes the Swedish participants responses to the post-study Likert-scale questionnaire items. Overall, participants reported high levels of comfort with the drone's behavior, movement, and predictability. Ratings related to perceived danger were low, while trust and perceived safety received consistently high scores. Furthermore, the distributions indicate relatively low disagreement among participants for several core measures, particularly regarding comfort with drone size and overall interaction safety.

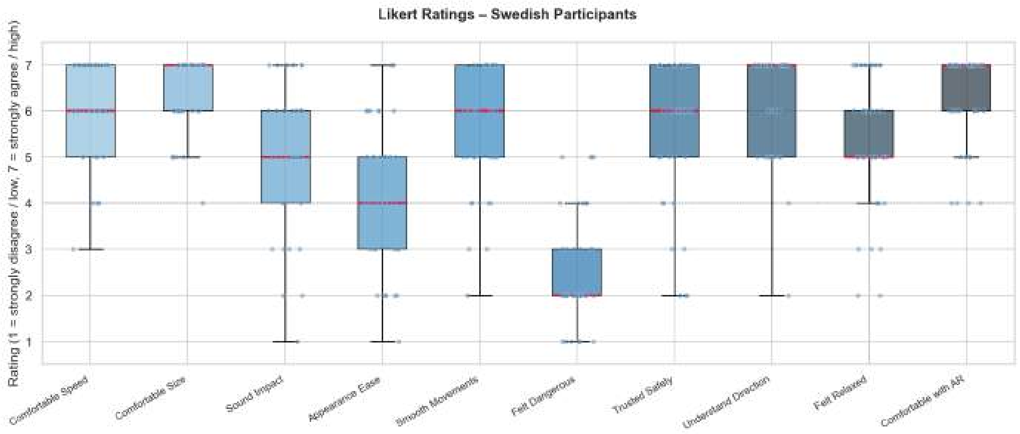


Figure 6.14: Boxplots of the post-study Likert questions in the Swedish sample.

Statistical testing on the Swedish post-study Likert metrics revealed that subjective user perceptions generally had a negligible impact on mean stopping distances, with one exception. The item "Appearance ease" (The drone's appearance made me feel at ease) demonstrated a statistically significant relationship across all three testing frameworks using the conventional $\alpha = .05$ threshold (ANOVA $p = 0.0100$, K-W $p = 0.0110$, Spearman $p = 0.0130$), accompanied by a strong negative correlation ($\rho = -0.404$) and a large effect size ($\eta^2 = 0.381$). This indicates that Swedish subjects who self-reported feeling more psychologically at ease with the drone's physical appearance maintained shorter mean stopping distances. However, this effect would not survive a Bonferroni-correction (cf. Eq. 5.1). No other subjective metrics reached statistical significance within this subset.

Table 6.21: Inferential Statistical Summary for Post-Study Likert Items: Swedish Sample ($N = 37$).

Likert Item	**Test Type**	**p-value**	**Effect Size (η^2 or ρ)**
Comfortable speed	ANOVA	0.6796	$\eta^2 = 0.045$
	K-W	0.7887	$\eta^2 = -0.072$
	Spearman	0.4577	$\rho = -0.126$
Comfortable size	ANOVA	0.9073	$\eta^2 = 0.006$
	K-W	0.4050	$\eta^2 = -0.003$
	Spearman	0.3409	$\rho = -0.161$
Sound impact	ANOVA	0.1455	$\eta^2 = 0.230$
	K-W	0.1148	$\eta^2 = 0.141$
	Spearman	0.8521	$\rho = -0.032$
Appearance ease	ANOVA	0.0100	$\eta^2 = 0.381$
	K-W	0.0110	$\eta^2 = 0.352$
	Spearman	0.0130	$\rho = -0.404$
Smooth movements	ANOVA	0.6186	$\eta^2 = 0.079$
	K-W	0.4494	$\eta^2 = -0.009$
	Spearman	0.6988	$\rho = +0.066$
Felt dangerous	ANOVA	0.4200	$\eta^2 = 0.111$
	K-W	0.2656	$\eta^2 = 0.038$
	Spearman	0.1091	$\rho = +0.268$
Trusted safely	ANOVA	0.1030	$\eta^2 = 0.246$
	K-W	0.2038	$\eta^2 = 0.072$
	Spearman	0.5361	$\rho = -0.105$
Understand direction	ANOVA	0.6060	$\eta^2 = 0.031$
	K-W	0.7279	$\eta^2 = -0.061$
	Spearman	0.8506	$\rho = +0.032$
Felt relaxed	ANOVA	0.4838	$\eta^2 = 0.129$
	K-W	0.2991	$\eta^2 = 0.035$
	Spearman	0.1938	$\rho = -0.219$
Comfortable with AR	ANOVA	0.7764	$\eta^2 = 0.032$
	K-W	0.8164	$\eta^2 = -0.063$
	Spearman	0.9542	$\rho = +0.010$

6.7.2 Vietnam

Figure 6.15 summarizes the Vietnamese participants responses to the post-study Likert-scale questionnaire items. While participants consistently rated high comfort scores for the speed and size the Vietnamese subsample stood out rating felt dangerous and trusted safely questions worse than Swedish participants. This could be due to the high level of convenience sampling used in Sweden resulting in participants trusting the researchers intentions to a higher degree.

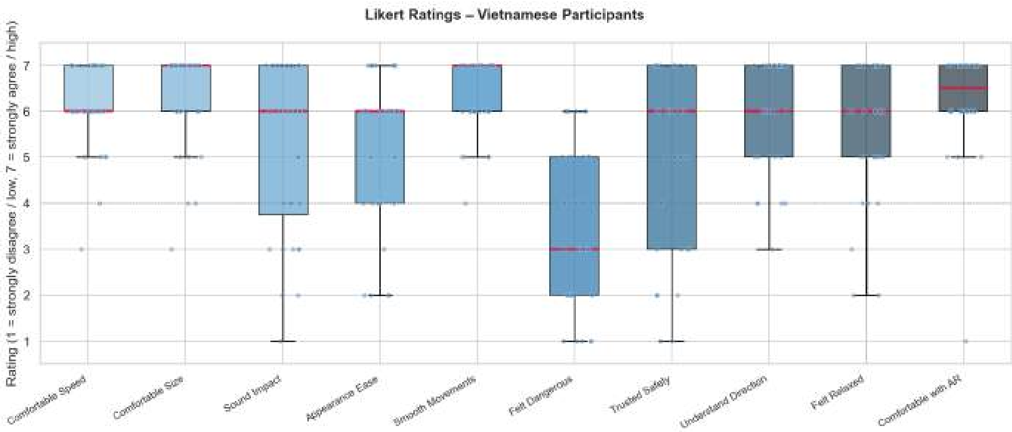


Figure 6.15: Boxplots of the post-study Likert questions in the Vietnamese sample.

Analysis of the Vietnamese post-study Likert items indicated no conventionally statistically significant relationships between subjective user responses and mean stopping distances. However, two items showed noteworthy trends. The strongest trend emerged from the "Felt dangerous" metric, which approached significance under the Spearman rank correlation ($p = 0.0729$, $\rho = +0.321$), while also showing relatively low p-values and moderate effect sizes under ANOVA ($p = 0.1133$, $\eta^2 = 0.277$) and Kruskal-Wallis ($p = 0.1409$, $\eta^2 = 0.127$). The positive correlation suggests that participants who reported stronger feelings of danger tended, although not significantly, to maintain longer stopping distances.

A second weaker trend appeared for "Felt relaxed". This item also produced relatively low ANOVA and Spearman p-values, with ANOVA indicating a moderate effect size ($p = 0.1021$, $\eta^2 = 0.249$) and Spearman showing a negative association ($p = 0.1622$, $\rho = -0.253$). The Kruskal-Wallis test was weaker ($p = 0.1791$, $\eta^2 = 0.100$). The test suggests that participants who felt more relaxed tended to allow the drone to come closer, although this relationship was not statistically reliable. Overall, these patterns point in the expected psychological direction: perceived danger was loosely associated with larger stopping distances, while relaxation was loosely associated with shorter stopping distances. Still, because none of the tests reached the conventional $\alpha = 0.05$ threshold, let alone the Bonferroni-corrected threshold (cf. Eq. 5.1), these findings should be interpreted as purely exploratory trends.

Table 6.22: Inferential Statistical Summary for Post-Study Likert Items: Vietnamese Sample ($N = 32$).

Likert Item	**Test Type**	p**-value**	**Effect Size** (η^2 **or** ρ)
Comfortable speed	ANOVA	0.5017	$\eta^2 = 0.050$
	K-W	0.4389	$\eta^2 = -0.009$
	Spearman	0.4541	$\rho = -0.137$
Comfortable size	ANOVA	0.7177	$\eta^2 = 0.048$
	K-W	0.6651	$\eta^2 = -0.060$
	Spearman	0.4465	$\rho = +0.139$
Sound impact	ANOVA	0.5728	$\eta^2 = 0.135$
	K-W	0.5732	$\eta^2 = -0.049$
	Spearman	0.3605	$\rho = +0.167$
Appearance ease	ANOVA	0.7957	$\eta^2 = 0.060$
	K-W	0.8936	$\eta^2 = -0.128$
	Spearman	0.9328	$\rho = -0.016$
Smooth movements	ANOVA	0.7930	$\eta^2 = 0.016$
	K-W	0.4546	$\eta^2 = -0.014$
	Spearman	0.7162	$\rho = -0.067$
Felt dangerous	ANOVA	0.1133	$\eta^2 = 0.277$
	K-W	0.1409	$\eta^2 = 0.127$
	Spearman	0.0729	$\rho = +0.321$
Trusted safely	ANOVA	0.5680	$\eta^2 = 0.136$
	K-W	0.6423	$\eta^2 = -0.070$
	Spearman	0.8514	$\rho = +0.034$
Understand direction	ANOVA	0.5837	$\eta^2 = 0.068$
	K-W	0.7089	$\eta^2 = -0.069$
	Spearman	0.8954	$\rho = +0.024$
Felt relaxed	ANOVA	0.1021	$\eta^2 = 0.249$
	K-W	0.1791	$\eta^2 = 0.100$
	Spearman	0.1622	$\rho = -0.253$
Comfortable with AR	ANOVA	0.4134	$\eta^2 = 0.061$
	K-W	0.3658	$\eta^2 = 0.006$
	Spearman	0.2862	$\rho = +0.194$

6.7.3 General

With the general sample ($N = 69$), testing post-study questions revealed an effect for the "Felt dangerous" item under the Spearman test ($\rho = +0.291$, $p = 0.0152$), while the Kruskal-Wallis test showed a near-significant trend ($\eta^2 = 0.089$, $p = 0.0593$) and the ANOVA result was also near-significant ($\eta^2 = 0.142$, $p = 0.0796$). Together, these results reveal a slight pattern that cross-culturally, participants who experienced higher perceived danger during the trials had longer mean stopping distances.

Table 6.23: Inferential Statistical Summary for Post-Study Likert Items: Full Sample ($N = 69$).

Likert Item	**Test Type**	p**-value**	**Effect Size** (η^2 **or** ρ)
Comfortable speed	ANOVA	0.1988	$\eta^2 = 0.088$
	K-W	0.3945	$\eta^2 = 0.001$
	Spearman	0.2602	$\rho = -0.137$
Comfortable size	ANOVA	0.3241	$\eta^2 = 0.052$
	K-W	0.5851	$\eta^2 = -0.018$
	Spearman	0.8989	$\rho = -0.016$
Sound impact	ANOVA	0.1064	$\eta^2 = 0.151$
	K-W	0.0958	$\eta^2 = 0.077$
	Spearman	0.6115	$\rho = +0.062$
Appearance ease	ANOVA	0.2203	$\eta^2 = 0.104$
	K-W	0.2501	$\eta^2 = 0.030$
	Spearman	0.1433	$\rho = -0.178$
Smooth movements	ANOVA	0.9371	$\eta^2 = 0.013$
	K-W	0.6920	$\eta^2 = -0.031$
	Spearman	0.9555	$\rho = +0.007$
Felt dangerous	ANOVA	0.0796	$\eta^2 = 0.142$
	K-W	0.0593	$\eta^2 = 0.089$
	Spearman	0.0152	$\rho = +0.291$
Trusted safely	ANOVA	0.0961	$\eta^2 = 0.155$
	K-W	0.1945	$\eta^2 = 0.043$
	Spearman	0.7677	$\rho = -0.036$
Understand direction	ANOVA	0.9141	$\eta^2 = 0.008$
	K-W	0.9728	$\eta^2 = -0.066$
	Spearman	0.8135	$\rho = +0.029$
Felt relaxed	ANOVA	0.2769	$\eta^2 = 0.093$
	K-W	0.3836	$\eta^2 = 0.004$
	Spearman	0.0972	$\rho = -0.201$
Comfortable with AR	ANOVA	0.7701	$\eta^2 = 0.017$
	K-W	0.5542	$\eta^2 = -0.015$
	Spearman	0.3296	$\rho = +0.119$

7

Discussion

This chapter interprets the results gathered by the analytical approaches. It starts by answering RQ 1 in section 7.1 and then moves on to evaluate the predictive values of all parameters in the study, answering RQ 2 in 7.3, and RQ 2.1 in 7.4.

7.1 Proxemic Zones and Degrees of Personalization

7.1.1 Personalized

In real-world deployment scenarios, such as an in-home personal assistant drone repeatedly interacting with the same individual, a personalized rubber-bubble model could be well-suited to the task. The drone can undergo an initial calibration phase with a user and employ continual learning [3] to refine its model for that user over time. A continuous rubber-bubble equation is advantageous here because it can interpolate across untested approach angles and one can add new variables such as drone speed or drone size if more personalization is required. Transforming the rubber-sheet model into a rubber-bubble model answers **RQ 1**: "What rubber-sheet model adapts well to human-drone proxemics in a 3D space?"

The results of personalized rubber-bubble modeling, following the rubber-sheet framework proposed by Camara et al. [12] for the Swedish dataset, are illustrated in Tables 6.2, 6.3, and 6.4. With same-subject data in training and testing (subject-unaware), these models achieved a maximum $R^2 \approx 0.83$, personalizing well to the subjects and suggesting that the rubber-bubble models can geometrically adapt to individual proxemic zones (see Figure 5.2). This upper bound reflects the theoretical maximum achievable accuracy given the repeated-measures structure of the data, where the remaining variance constitutes the noise between a user's angles [45]. Lower degree models result in less personalization in favor of more similar proxemic bubbles across people with demographics that are within the range of the data that it was trained on. An illustrative example of second-degree rubber-bubble formulations, is detailed in Table 6.4 illustrating how, for this dataset prioritizing demographics makes for the best results. Including the two angles and demographic parameters in a model quickly makes it complex. For example, a third degree model, that includes vertical angle, horizontal angle, age and height would have 35 parameters that would need to be calculated every time a subject is presented to the drone (see Eq. 5.3). Combining the angle parameters into a single zone parameter would bring the number of parameters down to 20 instead. This simplified model is explained below in section 7.2. See Table 7.1 for an overview of the number of parameters for different model configurations.

Table 7.1: Polynomial Feature Complexity by Parameter Count and Degree.

Parameters (n)	**Degree (d)**	**Calculation**	**Total Features**
3 (H, V, Age)	2	$\frac{5!}{3!2!}$	10
3 (H, V, Age)	3	$\frac{6!}{3!3!}$	20
4 (+ Height)	2	$\frac{6!}{4!2!}$	15
4	3	$\frac{7!}{4!3!}$	35
5 (+ Gender)	3	$\frac{8!}{5!3!}$	56

Higher-degree polynomial models with additional features can be introduced to increase personalization (as seen in Tables 6.2 and 6.9), though a fail-safe mechanism, such as defaulting to the population mean plus a standard deviation, is advisable for new subjects due to the large overfit. As the dataset grows, the model may gradually begin to wrongly generalize over individuals whose characteristics are well-represented in the training data and thus require more specific subject identifiers such as face identification or simply more features to reach maximum personalization, although that is not the goal here. The fully personalized models are not recommended for new subjects as even lower degree, simpler models, may not generalize well due to the objective parameters showing such low predictive power in this dataset.

7.1.2 Generalized

The strong personalized performance did not extend to unseen subjects. Subject-aware evaluation via LOSOCV (Table 6.6) showed that demographic parameters were insufficient to build generalizable models, with negative R^2 scores across all models, including the ones stratified by genders. This means that guessing the population mean currently yields more accurate predictions for new subjects than any demographic-based rubber-bubble model. Even though two people share similar demographics they may not share similar human-drone IPD preferences. However, this would need to be confirmed with larger datasets than the ones used in this study. These fully generalized models would be useful in contexts where the drone has a lower number of recurring interactions with a specific user, i.e. out in crowded environments where subjects quickly pass through, such as a busy street.

The LMM explains the lack of generalization. An ICC of 0.76 indicates that 76% of the variance in stopping distance comes from stable between-subject differences rather than from the measured predictors. Likewise, the large gap between the conditional R^2 of 0.76 and the marginal R^2 of 0.057 shows that participants were highly consistent with themselves, while the demographic and angle-based features explained little of the variation between individuals.

Developing effective generalized models would require predictors with stronger and more consistent associations with stopping distance. Some variables have shown predictive relevance in related work, such as cultural background in [17] and vertical approach angles in [9], suggesting that a larger and more demographically diverse dataset may be necessary to reach meaningful generalizability. The present findings establish that within the frontal zone and the objective parameters evaluated in this study, there is not enough correlation between the features and a person's stopping distance for a strong generalized model.

7.1.3 The Middle-Ground

For circumstances where personalization is desired, along with some generalization, a subject-unaware K-fold cross validation approach could be utilized. This could reduce the overfit to specific subjects and thus give them slightly less favorable distancing for the benefit of giving new subjects more generalized experiences. This helps with privacy as well by not having the drone identify a specific user but generalize more over the collected demographics in its training data. This model would be applicable in, for example, office spaces where a drone would have recurring interactions with the same users, but also encounter plenty of new people.

Tables 6.5 and 6.11 show that in this case, the K-fold models do not get lower R^2 scores than the hold-out ones, perhaps due to the high correlations between angles in the dataset. Furthermore, to get the required result and reach subjectively accepted generalizability for each new person, one would (as always with generalizability) preferably have a larger dataset and find features with more predictive power than the ones in this study.

7.2 Approach Angle Effects

A consistent finding across both the Swedish and Vietnamese datasets was that approach angles within the frontal zone carried little predictive value for stopping distance. The best rubber-bubble models, both with and without subject-aware evaluation, repeatedly excluded horizontal and vertical angles or added them with no tangible improvement. Tables 6.9 and 6.12 hint, with vertical angle included in the best models, that participants differed in their vertical preferences in the Vietnamese sample more than in the Swedish sample. While the angle correlations did span a greater range in Vietnam (see Figure 6.8), the Vietnamese 5-fold cross-validation resulted in all three best models in their respective degrees (see Table 6.11), not utilizing vertical angle. This could mean that for Table 6.9 the hold-out method simply got an advantageous fold for the vertical angle, but when averaging it across folds, thus getting a more honest answer, vertical angle does not prove as significant. This was further confirmed by the LMM, in which neither horizontal approach angle ($\beta \approx 0.000$, $p = 0.77$) nor vertical approach angle ($\beta = -0.0003$, $p = 0.43$) were significant predictors of stopping distance in the general dataset.

This result is largely explained by the fact that the frontal zone only covers an area in front of the subjects and they are always looking at the drone as it approaches them. When the drone approaches from anywhere within the frontal zone, participants tended to stop it at approximately the same distance regardless of the specific trajectory. For studies that examine the full space around a participant, such as Camara et al. [12], there is larger variations between frontal and lateral or rear approach directions, meaning that the rubber-bubble should go from perfectly smooth to more uneven as angles affect stop distances. The present study's focus on the frontal zone was motivated by the aim of obtaining a dense dataset within the primary interaction zone, as one would most often turn toward an approaching drone, as well as by the spatial constraints of the testing environment. The angular range was narrow enough that participants may have perceived all approach trajectories as broadly equivalent.

The finding for vertical angles is also notable in the context of prior work. [9] and [70] reported significant differences in IPD as a function of vertical approach angle, with drones approaching from above (but also below) eliciting larger stopping distances than those approaching at eye level. The present study did not replicate this effect, which may reflect the narrower vertical range tested here compared to the broader ones examined in [9].

Within the frontal zone, stopping distance behaves more like a zone-level response than a continuously angle-dependent one. Rather than treating the approach angle as a continuous predictor, a zone-based rubber-bubble formulation may be more appropriate and efficient for practical deployment. In such a model, the frontal zone is treated as a single spatial zone, with discrete zone indicators replacing the continuous angle parameters. Below is an example of a first degree zonal model:

$$R_i(Z_i, x_i) = a + \sum_{k=1}^{m} \beta_k \mathbb{I}(Z_i = k) + c_i x_i \tag{7.1}$$

where i is an index ranging from 1 to n, n is the number of data points, R_i is the predicted stopping distance, m is the total number of spatial zones, $\mathbb{I}(Z_i = k)$ is an indicator function equal to 1 if observation i falls in zone k and 0 otherwise, x_i represents predictive features such as height or age, a is the intercept for the reference zone, β_k are zone-specific distance shifts relative to the intercept, and c_i are coefficients for the non-angle features. This formulation retains the flexibility of the rubber-sheet framework while reducing model complexity, and would work with areas that, like the frontal zone, contain large correlations between angles.

7.3 Demographic Predictors of Human-Social Drone Distancing

This section examines whether any of the measured objective parameters, gender, age, height, pet ownership, and student status, can explain why individuals differ from one another, thus answering **RQ 2**: What human demographics most strongly influence proxemics during HDI? Evidence is drawn from the non-parametric statistical tests, the LMM, and the Bayesian full model on the predictive value of these variables.

7.3.1 Gender

Gender produced the most consistent predictive signal across methods. The LMM estimated a negative coefficient for gender ($\beta = -0.17$, $p = 0.069$), suggesting that female participants maintained somewhat shorter stopping distances on average, though this did not reach the conventional significance threshold. The Bayesian full model produced similar evidence: the 89% CI for the gender coefficient was the only one that did not overlap zero, with the posterior mean suggesting that female participants had expected stopping distances approximately 19% shorter than male participants, conditional on culture, age, height, and subject-level variation. The central 89% of the posterior distribution placed this difference between approximately 4% and 30%. However, the Bayesian coefficient-level results should be interpreted cautiously. The full model describes associations within the fitted model, not necessarily causal effects. Any estimated predictor association, such as gender, does not by itself show that the predictor causally affects stop distance. The same applies to culture, age and height. These predictors may be associated with stop distance directly, indirectly or through other unmeasured variables not included in the model.

The probability that a male participant had a longer predicted stopping distance than a female participant was estimated at 0.80 for Swedish participants and 0.79 for Vietnamese participants (see Figure 6.13). However, posterior predictive intervals for gender contrast included zero for both cultural groups (see Figure 6.13), indicating that this group-level tendency did not translate into reliable separation at the level of individual predictions. In other words, the model suggests a tendency for males to have longer predicted stop distances (by allowing some posterior predictive simulations in which females have longer stop distances than males), while still allowing considerable overlap between individuals.

The Bayesian model comparison further showed that including gender did not improve out-of-sample prediction, suggesting that while gender may carry some directional association within the fitted model, it does not generalize as a predictor for unseen individuals in this dataset. These findings are in partial agreement with prior HRI literature, which has occasionally observed female participants maintaining shorter IPD with robots [62].

7.3.2 Age, Height, Pet Ownership, and Student Status

The remaining demographic variables provided little to no evidence of a meaningful relationship with stopping distance in any of the analyzes. The Bayesian model estimated negative associations for age and height (see Figures 6.11a and 6.11b), suggesting that older and taller participants might maintain somewhat shorter distances, but the CI for both coefficients overlapped zero. The statistical tests similarly found no significant effects for age group ($p = .193$) or height group ($p = .548$), with the effect size being either negligible or small. Pet ownership and student status, included on the basis of potential prior exposure to uncontrolled autonomous agents or familiarity with technology, showed no meaningful association with stopping distance under any framework.

7.3.3 Individual Variation as the Dominant Factor

The gap between the LMM marginal R^2 of 0.057 and the conditional R^2 of 0.76 shows that participants are highly consistent with themselves across repeated trials, but the reason why one participant's baseline is 0.6 m while another's is 1.4 m is not revealed by their age, height, gender, pet ownership, or student status. The LMM LOSOCV RMSE of 0.284 m for the best model, corresponding to a normalized RMSE of more than 0.90 relative to the standard deviation means that for a new, unseen participant, the model is about as good as guessing the population mean.

These results suggest that the determinants of individual proxemic preferences for drone interaction could be psychological rather than demographic. Variables such as prior experience with drones, general anxiety, or personal attitudes toward autonomous systems may carry stronger and more consistent associations with stopping distance than the objective parameters examined here.

7.3.4 Overall Findings

Across all analyzes, none of the demographic variables emerged as consistent and reliable predictors of stopping distance. The Mann-Whitney U tests found no statistically significant main effects for any of the six variables examined before or after applying the adjusted Bonferroni threshold (see Eq. 5.1). The LMM confirmed this picture: adding age, height, gender, pet ownership, and student status as fixed predictors raised the marginal R^2 to only 0.057, meaning that the full set of demographic variables together explained just 5.7% of the total variance in stopping distance. The Bayesian model similarly found no increase in predictive performance from adding gender, age, culture, and height after accounting for between-subject variation.

LMM comparison using AIC favored the simplest specification, a null model with only the random intercept and no demographic predictors, over any model that included fixed effects. Similarly, the Bayesian model comparison using LOSOCV found that the full model, incorporating gender, culture, age, and height, did not improve predictive performance for held-out subjects relative to the null model, which only included subject-level intercepts. In other words, accounting for differences between individuals was more important than knowing a participant's gender, culture, age and height. The results should be understood as evidence that the added predictors had no clear predictive value for new subjects in this dataset. It should not be interpreted as evidence that these parameters do not causally influence stop distance.

7.4 Cultural Differences: Sweden and Vietnam

A primary aim of this study was to evaluate whether individualist or collectivist cultural backgrounds influenced preferred stopping distance during HDI. Across all analyzes, the evidence for a meaningful cultural difference was weak and inconsistent, thus answering **RQ 2.1**: What is the impact of individualist and collectivist cultures on preferred human-drone IPD? However, some findings are still worth discussing.

The Mann-Whitney U test found no statistically significant difference in mean stopping distance between Swedish and Vietnamese participants ($U = 547$, $p = .592$, $r = -0.076$, negligible effect). Vietnamese participants did maintain slightly larger mean distances ($M = 0.969$ m, $Mdn = 0.948$ m, $SD = 0.319$) compared to the Swedish sample ($M = 0.929$ m, $Mdn = 0.885$ m, $SD = 0.254$). The Bayesian full model estimated a negative association for culture, suggesting that Vietnamese participants had somewhat shorter expected stop distances conditional on the included predictors, but the CI for this coefficient included zero, meaning the evidence is not conclusive. The apparent reversal between the descriptive means and the Bayesian coefficient is likely due to the fact that the Vietnamese subset was mostly male, and male participants tended toward larger stopping distances, thus inflating the group mean. Once gender is held constant in the Bayesian model, the cultural estimate shifts in the opposite direction.

In the Swedish subsample, gender had no meaningful effect on stopping distance ($U = 164$, $p = .867$, $r = -0.035$, negligible). Within the Vietnamese subsample, male participants maintained larger distances than female participants ($U = 151$, $p = .049$, $r = +0.459$, medium effect), with Vietnamese males averaging $M = 1.040$ m compared to $M = 0.789$ m for Vietnamese females. This is the strongest statistical signal in the cultural analysis and suggests that gender moderates proxemic behavior differently across the two cultures. The interaction effect between gender and culture also approached significance in the ANOVA analysis, $p = .062$, $\eta_p^2 = .053$, further suggesting a possible culture-specific gender pattern. Although, neither test result is considered significant under the corrected Bonferroni threshold (see Eq. 5.1) and the Vietnamese male and female subgroups are imbalanced.

Within the Vietnamese subsample of the culture-stratified age analysis, the age-group comparison approached conventional significance, with Gen Z participants showing larger stopping distances than Millennial participants. This result should be interpreted with caution, as the comparison was based on small and imbalanced cells where the Vietnamese Millennial group contained only two participants and the Gen X group was empty.

The posterior predictive distributions from the Bayesian model show that the distributions of predicted stopping distances for Swedish and Vietnamese participants overlapped substantially, with the cultural shift appearing weaker and less consistent than the estimated gender effect. Although the trend suggests that Vietnamese participants may have shorter predicted stopping distances than equivalent Swedish participants, the available evidence is insufficient to make a definitive claim about such a cultural effect.

These results suggest that the cultural context in which drone interactions take place may be less important than individual variation or gender-based factors in determining proxemic preferences. While the pattern of Vietnamese males maintaining noticeably larger distances than Vietnamese females deserves further investigation, the overall distributions of stopping distances across the two subsets were broadly comparable. This aligns with prior human-drone interaction literature, such as [41], where proxemics between individualist and collectivist backgrounds were not that different. Further cross-cultural research would be needed to determine whether the patterns observed here generalize beyond this specific cohort.

7.5 Post-Study Questionnaire Analysis

The post-study questionnaire analysis suggest that most self-reported experience measures were not strong predictors of stopping distance. Across the Swedish, Vietnamese, and combined samples, the subjective ratings collected in this study did in general not reliably distinguish participants by their stopping distances. However, a few exploratory patterns indicate that subjective perception of the drone may still be relevant. Since none of the below mentioned effects survived correction for multiple comparisons (cf. Eq 5.1), these findings should be treated as exploratory indicators rather than predictors.

In the Swedish subsample, the clearest exploratory signal was whether the drone's appearance made participants feel at ease. Participants who rated the drone's appearance more positively tended to allow it to approach closer. This suggests that visual design may influence proxemic comfort: a drone that looks less threatening or more acceptable may reduce the distance people feel they need to maintain, which aligns with [71]. However, this pattern did not appear in the Vietnamese subsample.

In the Vietnamese subsample, the most relevant trends were related to perceived danger and relaxation. Participants who felt the drone was more dangerous tended to stop it farther away, while participants who felt more relaxed tended to let it come closer. Although these relationships were not statistically significant, their direction is meaningful: both trends point toward the same underlying interpretation that emotional responses to the drone may shape stopping behavior.

The almost equal and opposite direction of the "Felt dangerous" and "Felt relaxed" trends suggests that the two items may partly reflect two ends of the same comfort dimension. Higher perceived danger would be expected to coincide with lower relaxation, meaning that one item may contribute limited additional information once the other is known, i.e. the items may be partially redundant. However, these items should not be treated as interchangeable without directly testing their relationship.

In the combined sample, perceived danger was the strongest signal. Participants who rated the drone as more dangerous tended to maintain larger stopping distances, which aligns with the intuitive expectation that perceived threat increases avoidance. However, this effect was modest and was not significant within either cultural subgroup individually. This means the result should not be overinterpreted as a confirmed predictor, but it does suggest that perceived threat is a promising variable for future human-drone proxemics research.

The post-study questionnaire analysis suggest that subjective perception of the drone may help explain some of the individual variation observed in stopping distances. Future models could include psychological and emotional measures more systematically, rather than relying mainly on demographics to explain individual proxemic preferences, although such a model would likely be more difficult to develop and deploy.

7.6 The Four Analytical Frameworks

It should be noted that utilizing four distinct analytical frameworks introduces a degree of redundancy, as the statistical, regression, mixed-effects, and Bayesian models evaluate similar demographic relationships. However, because research into three dimensional human-drone proxemics remains largely exploratory, this approach tests the viability of different frameworks. Each method was retained because it approaches the data with a different utility: the statistical tests establish conventional statistical metrics, the rubber-bubble model acts as the practical geometric framework for drone deployment, the linear mixed-effects model isolates the individual variance components, and the probabilistic Bayesian framework incorporates all uncertainty and assumptions made all the way to the final distribution used for inference. The results of all frameworks overlap and reach similar conclusions; therefore, readers can apply the specific approach that they find most appropriate.

7.7 Limitations

7.7.1 Warm-up Anchoring Effect

In light of the mean stopping distance remaining consistently around 0.9 m to 1.0 m across different groups, it is important to consider whether the initial warm-up procedure may have introduced an anchoring effect on participants' preferred distances [69]. During the warm-up, the drone is positioned 1.0 m in front of the participant before beginning its movement sequence, potentially establishing this distance as a subconscious reference point.

Bretin et al. [9] found their means in a similar drone proxemics test to be between 0.92 m and 1.15 m depending on the vertical angle, meaning that there may be relevance to this distance as a general mean. The observed stopping distances exhibited substantial dispersion around the mean, as can be seen in Figure 7.1. The sample standard deviation was 0.316 m, yielding a coefficient of variation of approximately 30% [1], which shows a high degree of variability in stopping distances throughout the sample, which mostly stems from between-subject variability as shown in 6.10.

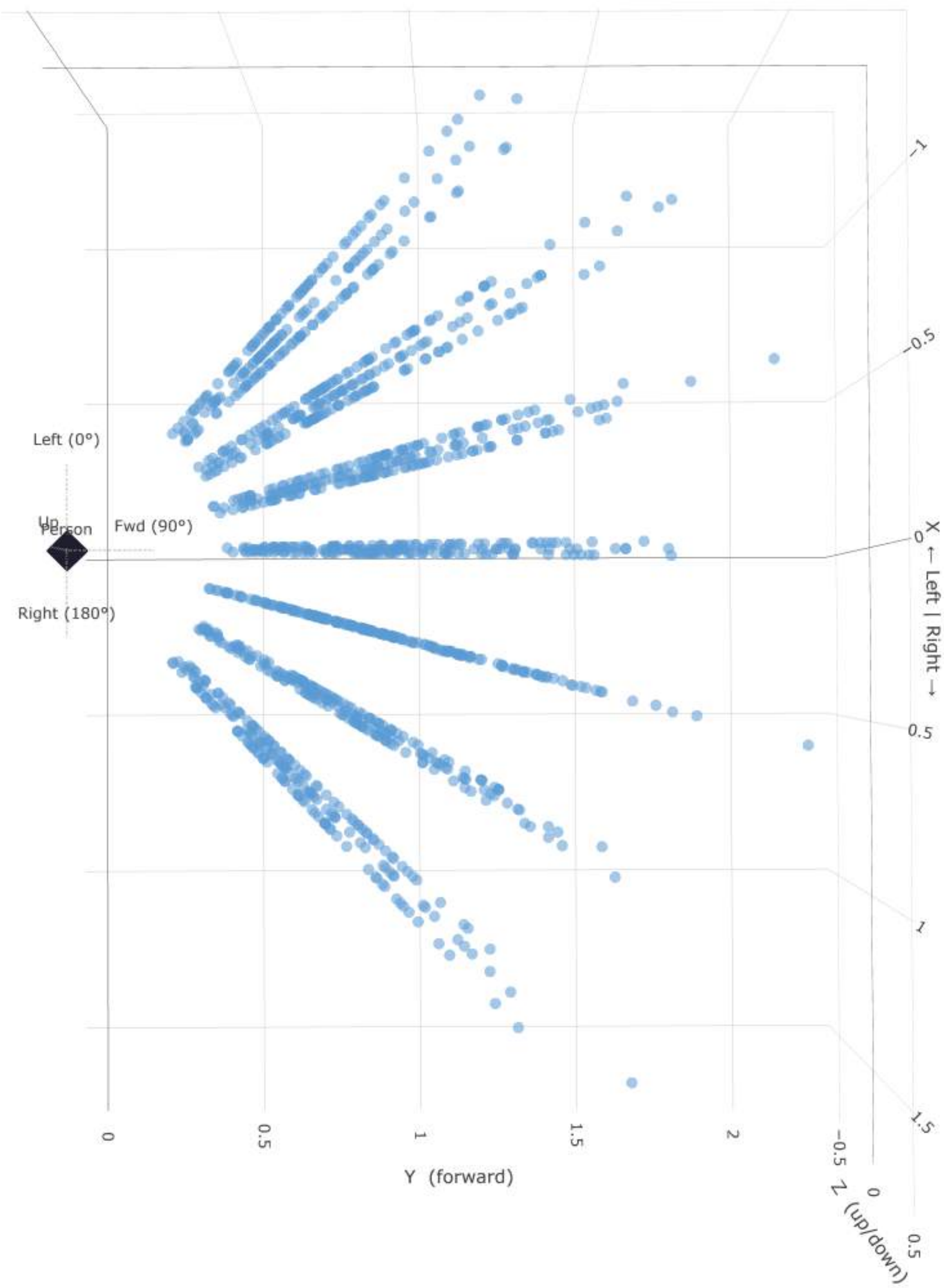


Figure 7.1: All observed stopping points in the combined sample seen from above; visualizing the wide spread of preferred stopping distances.

During the warm-up, the drone moved laterally up to 1.8 m to both sides of the participant. During this lateral movement, the drone's flight path covers the entire horizontal angular plane, all the way to the farthest horizontal angle $\theta = 45°$ and beyond. This farthest point of the horizontal plane is illustrated in Figure 7.2, which corresponds to an approximate distance of 1.4 m. If participants had simply reproduced distances experienced during the warm-up, they might have preferred the 1.4 m distance illustrated in Figure 7.2 at the horizontal approach angle $\theta = 45°$, or at least shown greater variation across the horizontal angular plane that the drone covered in the warm-up, with the closest being a frontal approach. Instead, the results showed relatively stable preferred distances regardless of angle, suggesting that maybe participants were not directly mirroring the drone's distance during the warm-up.

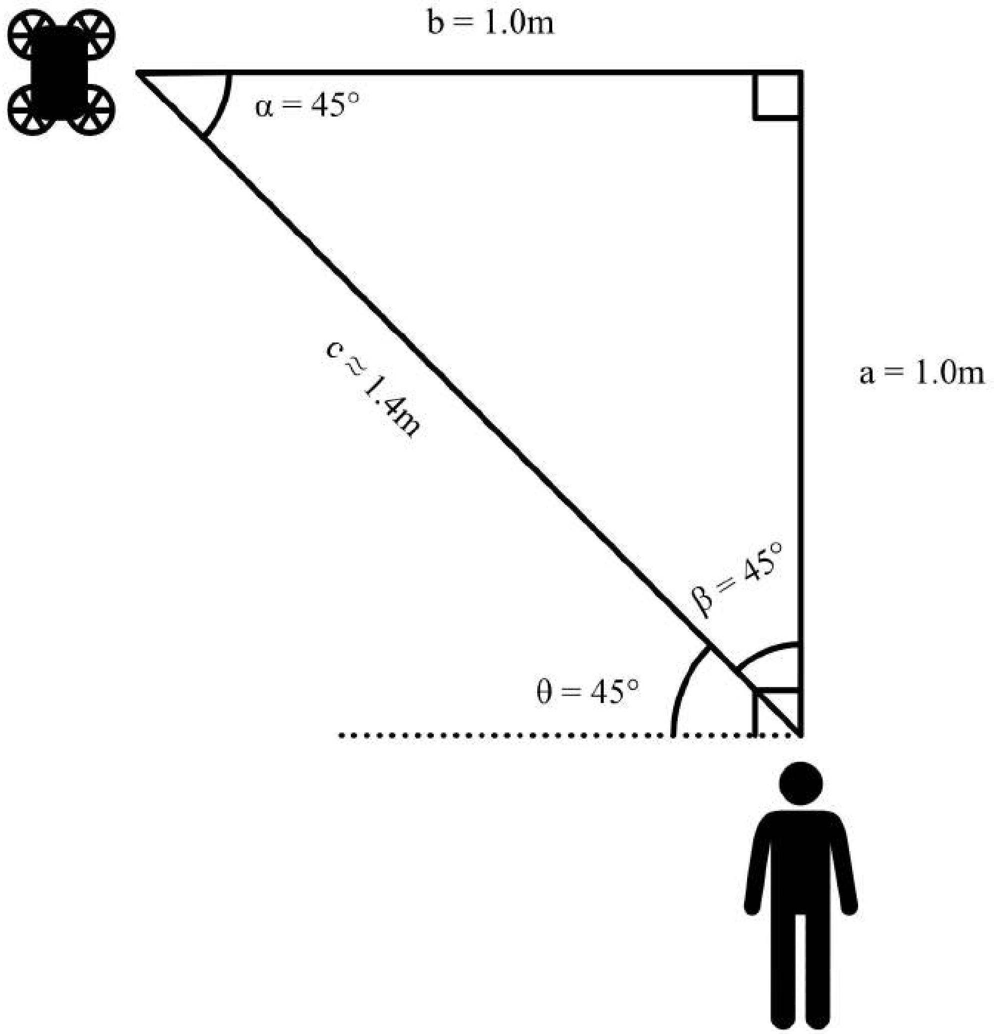


Figure 7.2: Illustration of the farthest point reached in the horizontal angular plane during the warm-up movement.

Anecdotally, several participants stated that they aimed to stop the drone at approximately "an arm's reach." Because the measurement pivot point was located 11 cm behind the front of the drone, the mean stopping distance to the closest edge of the drone was approximately 0.86 m. This closely aligns with adult arm lengths. For females ($N = 1530$), the mean shoulder-to-fingertip length is 0.729 m, with a mean height of 1.489 m, while for males ($N = 1191$), the mean is 0.808 m, with a mean height of 1.639 m [6]. While 0.86 m slightly exceeds these reaches, it matches a realistic interaction zone when accounting for various factors such as natural torso lean, shoulder extension, and minor safety clearances. This suggests that arm's reach could serve as an implicit reference point for comfortable interaction distance with a social drone.

The warm-up may still have influenced participants for all angles by establishing the drone as safe and non-threatening at approximately 1 m distance. Having already observed the drone at that range, participants may have perceived distances around 1 m as acceptable throughout the experiment. Such an effect may have been subconscious to participants, which could also explain why not a single participant mentioned the warm-up as influencing their decisions.

7.7.2 Framing

Another possible factor in the limited predictive power of the studied variables is the ambiguity surrounding the drone's intended purpose during the interaction. Although participants were instructed to perceive the drone as social, the framing remained intentionally broad and open to interpretation, with the intention of creating as general a model as possible. Qualitative observations suggested that participants frequently projected their own assumptions, experiences and professional contexts onto the drone. Several participants explicitly expressed uncertainty regarding how the drone should be interpreted. For example, a participant asked whether the drone should be viewed as belonging to a friend or as a random drone encountered "out on the street." Other participants anthropomorphized the drone and described it as "their friend." A participant working as a nurse interpreted it as a medical delivery drone and therefore allowed it to approach more closely, since such a delivery drone would need to come close. This interpretive variability may have introduced variance into the dataset, reducing the effects related to the studied variables. In the present study, the relatively open-ended framing may therefore have weakened the consistency of participants' preferred distances, as in [9], and contributed to the difficulty of constructing a predictive model.

7.7.3 AR Environment

Another factor that may have influenced the findings is the use of an AR environment rather than a physical drone. The post-study questionnaire showed consistently high ratings for comfort with AR, with nearly all participants rating their comfort between 5 and 7 on the Likert scale, with only a single participant expressing discomfort across all 69 participants (see Figures 6.14 and 6.15). This suggests that factors such as simulator discomfort, nausea, or unfamiliarity with the AR system had minimal impact on the results.

However, despite participants being comfortable with the technology itself, the AR setting may still have reduced the perceived physical risk associated with the drone. Although participants generally rating being comfortable with the drone (see Figures 6.14 and 6.15), several participants allowed the drone to approach closer than would realistically be expected in real-world interactions, which led to the data-cleanup described in Section 6.1. One possible explanation is a limited suspension of disbelief: although participants visually perceived the drone in their environment, they remained consciously aware that the drone was virtual and incapable of causing physical harm. Looking back, it would have been valuable to include a post-study question assessing how realistically participants perceived the AR environment and interaction. Anecdotally, several participants commented that the experience felt realistic. However, collecting systematic data on perceived realism would have strengthened the interpretation of the findings and provided a clearer understanding of the level of immersion achieved during the experiment. Furthermore, the limited suspension of disbelief could also have been subconscious, which would not have been reflected in a subjective post-study response.

This reduced sense of physical consequence may have weakened natural avoidance behavior inherent in certain studied variables, and thereby compressed interpersonal distances across participants. This could help explain the difficulty in identifying predictive relationships between variables and proxemic behavior. In a real-world interaction involving an actual flying drone with more realistic noise, airflow and a collision risk, participants may have had more cautious and differentiated responses.

7.7.4 Language Barriers and Good Subjects

In Vietnam the written instructions were used to convey the message of the drone being a social drone and that it is coming to interact with the participant. The framing of the social aspect was conveyed further during the verbal aspect and it was noted that the drone is coming to "talk" to the user. However, misunderstandings in the objective may have lead to participants letting the drone come unrealistically close or keeping it further away than they would if the drone was not social. Although, this framing could have little effect in the end as stated by [9].

There is a risk that some participants tried to be "good subjects." This means they might have let the drone come closer than they actually liked because they thought that was the goal of the test [51]. In Vietnam, this could be due to language barriers and Vietnam's high "power distance" score compared to Sweden [30]. In cultures with high power distance, people are often more likely to try and please the researcher or follow perceived rules. Interestingly, this wasn't just a potential issue in Vietnam as one Swedish participant even mentioned that they sometimes "challenged" themselves to let the drone get closer than they normally would. One possible contributing factor in the Swedish sample is the use of convenience sampling, where many participants were personally acquainted with the researchers. Familiarity with the researchers may have increased participants' tendency to act as “good subjects,” potentially influencing their behavior and responses during the study.

7.8 Future Work

Future research should validate the present findings using physical drones. While the AR methodology allowed for precise experimental control and eliminated the safety risks and regulatory hurdles associated with human-drone experiments, virtual models cannot fully reproduce the physical profile of an actual drone. Factors such as rotor blade airflow, real sound, and the perceived physical risk of a real collision may influence proxemic behavior in ways that are difficult to replicate in AR. For example, several participants in this study allowed the front of the virtual drone to approach within approximately 20-30 cm of their face, a threshold unlikely to be observed in real-world interactions due to the risk of collision. A comparative study between AR-based and real-drone experiments would establish the bounds of the "simulation-to-real" gap in human-drone proxemics.

Future studies should also investigate whether the uniform stopping distances observed across the frontal zone were artifacts of unconstrained head movement. In this design, the approach angle was defined relative to the participant's chest, allowing users to freely track the drone with their head. This tracking frequently kept the drone within the central field of view, which likely minimized the perceived spatial differences between approach trajectories.

Testing the correlation between a user's preferences in different zones is also of interest. This will answer the question of generalizing over zones, not having to train a model on every zone for every type of user. It will answer the question of: if a user prefers a specific stopping distance in the frontal zone, how does this translate into the other zones?

Future work should explore the impact of explicitly defining the drone's operational role. In this study, the functional intent of the drone was left open to interpretation, and qualitative observations suggested that participants projected specific use cases onto the drone based on their professional backgrounds. Providing a standardized social framing or use case in future iterations would minimize the variance in interpretation.

The impact of the drone's physical embodiment and visual aesthetics requires further systematic variation. In this study, the size, shape and visual architecture of the drone were kept static to isolate the effects of the approach angles and the objective parameters of the participants. However, post-test feedback indicated that visual design might modulate comfort levels. As shown in the post-study questionnaire results (see Figures 6.14 and 6.15), over 50% of the Swedish participants rated the drone's size with the highest possible comfort score (7/7), and about 85% of the Vietnamese participants rated it between 5 to 7. This highly favorable baseline perception may have actively shortened their preferred stopping distances. Future iterations should explicitly vary design factors such as size, color schemes, exposed vs. shrouded rotor blades, and anthropomorphic features to determine the extent to which proxemic boundaries are shaped by the perceived physical threat of the drone's visual form.

Future work should examine if there are better models than standard Linear Regression to deploy in autonomous drones. In this study the LMM and Bayesian models are mostly used for diagnostics and not considered for drone deployment, but the subject specific random intercepts of LMM could make for better personalized models, and the probabilistic nature of Bayesian could better capture the inherent uncertainty in all included parameters for more honest predictions.

In order to get the best rubber-bubble models for deployment, future validation would need to be done. Future research should deploy a rubber-bubble model in a user study and have the participants evaluate how they feel about the stop distances.

8

Conclusion

This thesis investigated wether a two dimensional human-robot rubber-sheet proxemics model could be adapted to a three dimensional human-drone proxemics model called the "rubber-bubble" model. It also investigated the predictive power of demographics such as age, gender, height, and culture for human-drone proxemics. A user study was carried out to collect data on the preferred stopping distances for various demographics which could then be used for training and evaluating proxemics models and for investigating the predictive power of the chosen parameters through various analytical frameworks. The study was performed in an AR environment where a virtual drone approached participants at various angles in front of them, an area we call the frontal zone, and having the participants say "stop" when it reached their preferred interaction distance at that angle. The distance was then recorded and used for the study.

The rubber-bubble was evaluated through the perspectives of varying degrees of personalization, finding that it personalized very well within the frontal zones and establishing the use for a rubber-bubble model for autonomous drones. When evaluated on unseen subjects via LOSOCV, the rubber-bubble models failed to outperform simply guessing the population mean, with negative R^2 scores for all models tested in this study. When limited to the specific demographic parameters included in this study, the models are not generalizable, making them unsuited for predicting the boundaries of unseen individuals to a degree that is subjectively appropriate for that person. Future research utilizing larger, more diverse datasets is required to establish whether generalizable patterns can be found for human-drone demographic proxemics modeling.

Approach angles within the frontal zone did not contribute meaningfully to prediction, as stopping distances remained largely uniform across the tested angles, suggesting a zone-based rubber-bubble formulation may offer a less complex model for deployment.

In addressing the demographic influences, no measured variable emerged as a statistically reliable predictor of stopping distance although patterns did emerge. Gender produced the most consistent directional signal, with the linear mixed-effects model estimating a negative association ($\beta = -0.17$, $p = 0.069$) and the Bayesian model estimating female stopping distances to be approximately 19% shorter than male distances. Height also showed initial promise within both the Bayesian analysis and non-parametric statistical tests as a potential physical predictor. However, these individual demographic effects ultimately failed to improve out-of-sample prediction. Age, pet ownership, and student status showed no meaningful associations under any framework. The LMM ICC of 0.76 confirmed that three-quarters of all variance was due to stable between-subject differences not captured by objective demographics, indicating that broad physical traits do not strongly dictate these boundaries.

Bibliography

[1] Hervé Abdi. Coefficient of variation. *Encyclopedia of research design*, 1(5):169–171, 2010.

[2] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. 2010.

[3] Ali Ayub, Zachary De Francesco, Patrick Holthaus, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. Continual learning through human-robot interaction: Human perceptions of a continual learning robot in repeated interactions. *International Journal of Social Robotics*, 17(2):277–296, February 2025.

[4] Harald Baayen. *Mixed-effects models*, pages 668–677. Oxford University Press, 01 2012.

[5] Christoph Bartneck, Tatsuya Nomura, Takayuki Kanda, Tomohiro Suzuki, and Kennichi Kato. Cultural differences in attitudes towards robots. In *Proceedings of the Symposium on Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction*, pages 1–9, 2005.

[6] G. Ö. Başıbüyük, Z. Ö. Güler, B. Kılıç, N. K. Yaylagül, F. Ay, O. Bekdaş, Ö. Özgür, and T. Gültekin. Designing ergonomic toilets and bathrooms for older adults: A study on anthropometric dimensions and recommendations. *Geriatric Nursing*, 55:263–269, Jan-Feb 2024. Epub 2023 Dec 12.

[7] Mehmet Aydin Baytas, Damla Çay, Yuchong Zhang, Mohammad Obaid, Asim Evren Yantaç, and Morten Fjeld. The design of social drones: A review of studies on autonomous flyers in inhabited environments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.

[8] Robin Bretin, Mohamed Khamis, Emily Cross, and Mohammad Obaid. The role of drone's digital facial emotions and gaze in shaping individuals' social proxemics and interpretation. *J. Hum.-Robot Interact.*, 14(3), May 2025.

[9] Rémy Bretin, Emily S. Cross, and Mohamed Khamis. Co-existing with drones: A virtual exploration of proxemic behaviours and users' insights on social

drones. *International Journal of Social Robotics*, 16:547–567, 2024.

[10] Fanta Camara and Charles Fox. Space invaders: Pedestrian proxemic utility functions and trust zones for autonomous vehicle interactions. *International Journal of Social Robotics*, 13(8):1929–1949, Dec 2021.

[11] Fanta Camara and Charles Fox. Modelling proxemic zone shapes from analytic and simulated motion strategies. *International Journal of Social Robotics*, 17(12):3027–3045, 12 2025.

[12] Fanta Camara, Adam El Jabaoui, Ramakrishnan Mukundan, and Mohammad Obaid. A rubber-sheet transformation model for personalized human-robot proxemics. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9210–9216, 2025.

[13] Davide Chicco, Andrea Sichenze, and Giuseppe Jurman. A simple guide to the use of student's t-test, mann-whitney u test, chi-squared test, and kruskal-wallis test in biostatistics. *BioData Mining*, 18(1):56, 2025.

[14] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ, 2nd edition, 1988.

[15] Brittany A. Duncan and Robin R. Murphy. Comfortable approach distance with small unmanned aerial vehicles. In *2013 IEEE RO-MAN*, pages 786–792, 2013.

[16] Roald Eiselen and Gerhard B. van Huyssteen. A comparison of statistical tests for likert-type data: The case of swearwords. *Journal of Open Humanities Data*, 2023.

[17] Ghadeer Eresha, Markus Häring, Birgit Endrass, Elisabeth André, and Mohammad Obaid. Investigating the influence of culture on proxemic behaviors for humanoid robots. In *2013 IEEE RO-MAN: The 22nd IEEE International Symposium on Robot and Human Interactive Communication*, pages 430–435. IEEE, 2013.

[18] Ghadeer Eresha, Markus Häring, Birgit Endrass, Elisabeth André, and Mohammad Obaid. Investigating the influence of culture on proxemic behaviors for humanoid robots. In *2013 IEEE RO-MAN*, pages 430–435, 2013.

[19] Fernanda Fiel Peres. Effect sizes for nonparametric tests. *Biochemia Medica*, 36(1):5–16, 2026.

[20] Yuan Gao, Sebastian Wallkötter, Mohammad Obaid, and Ginevra Castellano. Investigating deep learning approaches for human-robot proxemics. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1093–1098, 2018.

[21] Edward T Hall. The hidden dimension. *Garden City*, 1966.

[22] Edward T. Hall, Ray L. Birdwhistell, Bernhard Bock, Paul Bohannan, A. Richard Diebold, Marshall Durbin, Munro S. Edmonson, J. L. Fischer, Dell Hymes, Solon T. Kimball, Weston La Barre, S. J. Frank Lynch, J. E. McClellan, Donald S. Marshall, G. B. Milner, Harvey B. Sarles, George L Trager, and Andrew P. Vayda. Proxemics [and comments and replies]. *Current Anthropology*, 9(2/3):83–108, 1968.

[23] Béatrice S. Hasler and Doron A. Friedman. Sociocultural conventions in avatar-mediated nonverbal communication: A cross-cultural analysis of virtual proxemics. *Journal of Intercultural Communication Research*, 41(3):238–259, 2012.

[24] Trevor Hastie. The elements of statistical learning: data mining, inference, and prediction, 2009.

[25] Heiko Hecht, Robin Welsch, Jana Viehoff, and Matthew R. Longo. The shape of personal space. *Acta Psychologica*, 193:113–122, 2019.

[26] Olivia Herzog, Simone Nertinger, Katharina Wenzel, Abdeldjallil Naceri, Sami Haddadin, and Klaus Bengler. Assessing perceived discomfort and proxemic behavior towards robots: A comparative study between real and augmented reality presentations. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 959–965, 2023.

[27] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 42(1):80–86, 2000.

[28] Geert Hofstede. Culture and organizations. *International studies of management & organization*, 10(4):15–41, 1980.

[29] Geert Hofstede. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations.* Sage Publications, Thousand Oaks, CA, 2nd edition, 2001.

[30] Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. *Cultures and Organizations: Software of the Mind.* McGraw-Hill Education, 3rd edition, 2010.

[31] Gert Jan Hofstede, Paul B. Pedersen, and Geert Hofstede. *Exploring Culture: Exercises, Stories and Synthetic Cultures.* Intercultural Press, Yarmouth, United States, 2002.

[32] S.N.; Leonardi-Bee J.; Rosset S.; Shomron N.; Couturier D.-L.; Gran S. Hollestein, L.M.; Lo. Multiple ways to correct for multiple comparisons in multiple types of studies. *British Journal of Dermatology*, 185(6):1081–1083, 2021.

[33] Ann Huang, Pascal Knierim, Francesco Chiossi, Lewis L Chuang, and Robin

Welsch. Proxemics for human-agent interaction in augmented reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[34] Michiel Joosse. Lost in proxemics: Spatial behavior for cross-cultural hri, 05 2016.

[35] Sumin Kang, Sungwoo Yang, Daewon Kwak, Jargalbaatar Yura, and Donghan Kim. Social type-aware navigation framework for mobile robots in human-shared environments. *Sensors*, 24:4862, 07 2024.

[36] Piotr Kardasz, Jacek Doskocz, Mateusz Hejduk, Paweł Wiejkut, and Hubert Zarzycki. Drones and possibilities of their using. *J. Civ. Environ. Eng*, 6(3):1–7, 2016.

[37] Kari Daniel Karjalainen, Anna Elisabeth Sofia Romell, Photchara Ratsamee, Asim Evren Yantac, Morten Fjeld, and Mohammad Obaid. Social drone companion for the home environment: a user-centric exploration. In *Proceedings of the 5th International Conference on Human Agent Interaction*, HAI '17, page 89–96, New York, NY, USA, 2017. Association for Computing Machinery.

[38] Yunkyung Kim and Bilge Mutlu. How social distance shapes human–robot interaction. *International Journal of Human-Computer Studies*, 72(12):783–795, 2014.

[39] Rachel Kirby, Reid Simmons, and Jodi Forlizzi. Companion: A constraint-optimizing method for person-acceptable navigation. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 607 – 612, 11 2009.

[40] Michael H Kutner, Christopher J Nachtsheim, John Neter, and William Li. *Applied Linear Statistical Models*, volume 5. McGraw-Hill Irwin Boston, 2005.

[41] J. A. Landay, L. E. Jane, L. E. Ilene, and J. R. Cauchard. Drone & wo: Cultural influences on human-drone interaction techniques. In *Proceedings of the 2017 ACM CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 6794–6799, New York, NY, USA, 2017. ACM.

[42] Giuseppe Lanza, Francesco Fisicaro, Cinzia D'Agate, Raffaele Ferri, Mariagiovanna Cantone, Luca Falzone, Giovanni Pennisi, Rita Bella, Marios Hadjivassiliou, and Manuela Pennisi. Preserved central cholinergic functioning to transcranial magnetic stimulation in de novo patients with celiac disease. *PLOS ONE*, 16:e0261373, 12 2021.

[43] Rui Li, Marc van Almkerk, Sanne van Waveren, Elizabeth Carter, and Iolanda Leite. Comparing human-robot proxemics between virtual reality and the real world. In *2019 14th ACM/IEEE International Conference on Human-Robot*

Interaction (HRI), pages 431–439, 2019.

[44] Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, Boca Raton, FL, 2 edition, 2020.

[45] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis*. John Wiley & Sons, Hoboken, New Jersey, 5th edition, 2012.

[46] Jonathan Mumm and Bilge Mutlu. Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 331–338, 2011.

[47] Shinichi Nakagawa and Holger Schielzeth. A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142, 2013.

[48] Nadja E. Neef, Sören Zabel, Matthias Lauckner, et al. What is appropriate? on the assessment of human-robot proxemics for casual encounters in closed environments. *International Journal of Social Robotics*, 15:953–967, 2023.

[49] Margot M. E. Neggers, Raymond H. Cuijpers, Peter A. M. Ruijten, and Wijnand A. IJsselsteijn. Determining shape and size of personal space of a human when passed by a robot. *International Journal of Social Robotics*, 14(2):561–572, mar 2022.

[50] Margot M. E. Neggers, Raymond H. Cuijpers, Peter A. M. Ruijten, and Wijnand A. IJsselsteijn. The effect of robot speed on comfortable passing distances. *Frontiers in Robotics and AI*, Volume 9 - 2022, 2022.

[51] Austin Nichols and Jon Maner. The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, 135:151–65, 05 2008.

[52] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, Jiyoung Han, Naho Shin, Jennifer Burke, and Kennichi Kato. What people assume about humanoid and animal-type robots: cross-cultural analysis between japan, korea, and the united states. *International Journal of Humanoid Robotics*, 5(01):25–46, 2008.

[53] Geoffrey Norman. Likert scales, levels of measurement adn the “laws” of statistics. *Advances in health sciences education : theory and practice*, 15:625–32, 02 2010.

[54] Mohammad Obaid, Wafa Johal, and Omar Mubin. Domestic drones: Context of use in research literature. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, HAI ’20, page 196–203, New York, NY, USA,

2020. Association for Computing Machinery.

[55] Vineeth Rajamohan, Connor Scully-Allison, Sergiu Dascalu, and David Feil-Seifer. Factors influencing the human preferred interaction distance. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7, 2019.

[56] David Roberts, Volker Bahn, Simone Ciuti, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, Jose Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David Warton, Brendan Wintle, Florian Hartig, and Carsten Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 12 2016.

[57] Judy Robertson and Maurits Kaptein. *Modern Statistical Methods for HCI*. Springer, Cham, Switzerland, 2016.

[58] S. M. Bhagya P. Samarakoon, M. A. Viraj J. Muthugala, and A. G. Buddhika P. Jayasekara. A review on human–robot proxemics. *Electronics*, 11(16), 2022.

[59] Aziez Sardar, Michiel Joosse, Astrid Weiss, and Vanessa Evers. Don't stand so close to me: users' attitudinal and behavioral responses to personal space invasion by robots. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, page 229–230, New York, NY, USA, 2012. Association for Computing Machinery.

[60] P. Schober, C. Boer, and L. Schwarte. Correlation coefficients: Appropriate use and interpretation. *Anesthesia Analgesia*, 126:1763–1768, 2018.

[61] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[62] Leila Takayama and Caroline Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5495–5502, 2009.

[63] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 01 1996.

[64] Peter Trautman and Andreas Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 797–803. IEEE, 2010.

[65] Jacqueline Urakami and Katie Seaborn. Nonverbal cues in human–robot interaction: A communication studies perspective. *ACM Transactions on Human-Robot Interaction*, 12(2):1–21, March 2023.

[66] Arunas Petras Verbyla. A note on model selection using information criteria for

general linear models estimated using reml. *Australian & New Zealand Journal of Statistics*, 61(1):39–50, 2019.

[67] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert-Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T. Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, Yoshiki Vázquez-Baeza, and SciPy 1.0 Contributors. Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, 2020.

[68] Max Weber. *Economy and society: An outline of interpretive sociology*, volume 2. University of California press, 1978.

[69] Timothy D. Wilson, Christopher E. Houston, Kathryn M. Etling, and Nancy Brekke. A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, 125(4):387–402, December 1996.

[70] Elisabeth Maria Wögerbauer, Christoph von Castell, Robin Welsch, and Heiko Hecht. Preferred distance in human–drone interaction. *Vision*, 8(4), 2024.

[71] Alexander Yeh, Photchara Ratsamee, Kiyoshi Kiyokawa, Yuki Uranishi, Tomohiro Mashita, Haruo Takemura, Morten Fjeld, and Mohammad Obaid. Exploring proxemics for human-drone interaction. In *Proceedings of the 5th International Conference on Human Agent Interaction*, HAI '17, page 81–88, New York, NY, USA, 2017. Association for Computing Machinery.

A

Example: Calculating A Standardized 2^{nd} Degree Polynomial

The polynomial regression model is fitted using a scikit-learn `Pipeline` consisting of three sequential steps:

1. **Polynomial feature expansion**: the raw input features are expanded into all polynomial terms up to the specified degree (see Eq. (5.2)).
2. **StandardScaler**: each expanded feature is standardized independently to zero mean and unit variance using statistics computed from the training data only.
3. **Linear regression**: the model learns one coefficient per standardized feature plus a single intercept term.

Because standardization is applied *after* polynomial expansion, the published coefficients act on the standardized polynomial features, not on standardized raw inputs. This section shows how to reproduce a model prediction by hand.

A.1 Model Parameters

The example below uses the 2^{nd} female model with *height* (cm) and *age* (years) as predictors and stop distance (m) as the target (see Table 6.4 for the model chosen). The five polynomial features generated from two raw inputs h and a are:

$$\mathbf{z} = \Big(h,\ a,\ h^2,\ h \cdot a,\ a^2\Big). \tag{A.1}$$

Each element z_i is then standardized using the mean μ_i and standard deviation σ_i estimated from the training set (Table A.1).

Table A.1: Fitted coefficients and training-set scaler statistics for the degree-2 female model. The standardized feature entering the linear predictor is $\tilde{z}_i = (z_i - \mu_i)/\sigma_i$.

Feature z_i	**Coefficient** $\hat{\beta}_i$	**Mean** μ_i	**Std** σ_i
Intercept	0.9733	—	—
h	−5.6307	165.5299	5.6403
a	1.6118	28.0983	9.4744
h^2	5.6906	27431.9658	1881.7693
$h \cdot a$	−0.8438	4656.5085	1617.6120
a^2	−0.8572	879.2778	751.4655

A.2 Step-by-step Procedure

Given a new observation with raw height h (cm) and raw age a (years), the predicted stop distance $\hat{y}$ (m) is computed as follows.

Step 1 — Compute the five raw polynomial features.

$$z_1 = h, \qquad z_2 = a, \qquad z_3 = h^2, \qquad z_4 = h \cdot a, \qquad z_5 = a^2. \tag{A.2}$$

Step 2 — Standardize each feature using Table A.1.

$$\tilde{z}_i = \frac{z_i - \mu_i}{\sigma_i}, \qquad i = 1, \ldots, 5. \tag{A.3}$$

Step 3 — Calculate the stop distance.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \tilde{z}_1 + \hat{\beta}_2 \tilde{z}_2 + \hat{\beta}_3 \tilde{z}_3 + \hat{\beta}_4 \tilde{z}_4 + \hat{\beta}_5 \tilde{z}_5. \tag{A.4}$$

A.3 Practical Example

Let $h = 165$ cm and $a = 30$ years.

Step 1: Raw polynomial features:

$$z_1 = 165, \quad z_2 = 30, \quad z_3 = 165^2 = 27225, \quad z_4 = 165 \times 30 = 4950, \quad z_5 = 30^2 = 900. \tag{A.5}$$

Step 2: standardize:

$$\tilde{z}_1 = \frac{165 - 165.5299}{5.6403} = -0.094, \tag{A.6}$$

$$\tilde{z}_2 = \frac{30 - 28.0983}{9.4744} = \quad 0.201, \tag{A.7}$$

$$\tilde{z}_3 = \frac{27225 - 27431.9658}{1881.7693} = -0.110, \tag{A.8}$$

$$\tilde{z}_4 = \frac{4950 - 4656.5085}{1617.6120} = \quad 0.181, \tag{A.9}$$

$$\tilde{z}_5 = \frac{900 - 879.2778}{751.4655} = \quad 0.028. \tag{A.10}$$

Step 3: Predicted stop distance:

$$\begin{aligned} \hat{y} &= 0.9733 + (-5.6307) \times (-0.094) + 1.6118 \times 0.201 \\ &\quad + 5.6906 \times (-0.110) + (-0.8438) \times 0.181 + (-0.8572) \times 0.028 \\ &= 0.9733 + 0.5293 + 0.3240 - 0.6257 - 0.1527 - 0.0240 \\ &\approx \mathbf{1.02\ m}. \end{aligned} \tag{A.11}$$

B

ANOVA Analyses

This appendix reports the parametric ANOVA analyzes. As described in Section 5.1, non-parametric tests were used as the primary inferential approach because the normality assumption was not satisfied for all groups. The ANOVA results are included here for transparency.

Effect sizes are reported as eta-squared (η^2) for one-way ANOVA and partial eta-squared (η_p^2) for two-way ANOVA. Effect sizes were interpreted as negligible ($< .01$), small (.01-.05), medium (.06-.13) and large ($\geq .14$).

B.1 Assumption Checks

In addition to the assumptions described in 5.1, ANOVA also assumes homogeneity of variance, which was checked using Levene's test.

Levene's test did not indicate unequal variances for the one-way ANOVA tests reported below.

B.2 One-Way ANOVA: Main Effects

Table B.1 reports the descriptive statistics for each grouping variable. Table B.2 reports the corresponding one-way ANOVA results.

None of the one-way ANOVA tests reached statistical significance at the conventional $\alpha = .05$ threshold, and therefore none survived the Bonferroni-corrected threshold 5.1. The age-group ANOVA was not run because the Gen X group contained fewer than five participants. Overall, the ANOVA results do not provide evidence that mean stopping distance differed significantly by culture, gender, pet ownership, student status or height group.

Table B.1: Descriptive statistics for one-way ANOVA main effects.

Variable	Group	n	Mean	SD	Median
Culture	Swedish	37	0.929	0.254	0.885
Culture	Vietnamese	32	0.969	0.319	0.948
Gender	Male	43	0.984	0.288	0.912
Gender	Female	26	0.888	0.274	0.908
Pet ownership	No pets	25	0.903	0.245	0.885
Pet ownership	Has pets	44	0.974	0.304	0.919
Student status	Non-student	24	0.913	0.294	0.902
Student status	Student	45	0.967	0.281	0.928
Age group	Gen Z ($\leq$ 29)	60	0.969	0.281	0.926
Age group	Millennial (30–45)	5	0.733	0.254	0.634
Age group	Gen X (46+)	4	0.905	0.325	0.783
Height group	Short ($<$ 168 cm)	23	0.918	0.301	0.924
Height group	Medium (168–175 cm)	24	1.007	0.302	0.936
Height group	Tall ($\geq$ 175 cm)	22	0.915	0.248	0.880

Table B.2: One-way ANOVA main effects for mean stopping distance.

Variable	F	df	p	η^2
Culture	0.337	$(1, 67)$	.5638	.005
Gender	1.859	$(1, 67)$	.1773	.027
Pet ownership	0.986	$(1, 67)$	.3243	.015
Student status	0.554	$(1, 67)$	.4595	.008
Height group	0.778	$(2, 66)$	.4637	.023

B.3 Two-Way ANOVA: Culture by Demographic Variables

Exploratory two-way ANOVAs were conducted to test whether culture moderated the effect of each demographic variable on mean stopping distance. The interaction term was the primary term of interest in each model. Results are shown in Table B.3.

Table B.3: Two-way ANOVA interactions between culture and demographic variables.

Interaction	F	df	p	η_p^2
Gender × Culture	3.605	$(1, 65)$	.0620	.053
Pet ownership × Culture	1.590	$(1, 65)$	.2119	.024
Student status × Culture	1.818	$(1, 65)$	.1823	.027
Age group × Culture	0.700	$(2, 64)$	.5002	.021
Height group × Culture	1.238	$(2, 63)$	.2968	.038

No interaction reached the conventional statistical significance. However, Gender × Culture interaction ($F(1, 65) = 3.605$, $p = .0620$, $\eta_p^2 = .053$) borderline did. But given $\alpha = .05$, these results do not provide significant evidence that the demographic effects on stopping distance differed reliably between the Swedish and Vietnamese samples.

B.4 Two-Way ANOVA: Pairwise Demographic Interactions

Additional two-way ANOVAs were conducted for all pairwise combinations of the five demographic variables: gender, pet ownership, student status, age group and height group. Results are shown in Table B.4.

Table B.4: Exploratory pairwise demographic interactions from two-way ANOVA.

Interaction	F	df	p	η_p^2
Gender × Pet ownership	3.389	$(1, 65)$	.0702	.050
Gender × Student status	1.363	$(1, 65)$	.2472	.021
Gender × Age group	0.650	$(2, 63)$	.5257	.020
Gender × Height group	2.118	$(2, 64)$	.1286	.062
Pet ownership × Student status	0.035	$(1, 65)$	.8512	.001
Pet ownership × Height group	0.348	$(2, 63)$	.7077	.011
Student status × Age group	0.010	$(2, 65)$	.9213	.000
Student status × Height group	0.394	$(2, 63)$	.6758	.012

None of the pairwise interaction terms reached conventional statistical significance. The largest interaction effect was Gender × Height group, $F(2, 64) = 2.118$, $p = .1286$, $\eta_p^2 = .062$. The Gender × Pet ownership interaction almost reached $\alpha = .05$ significance, $F(1, 65) = 3.389$, $p = .0702$, $\eta_p^2 = .050$.

Two interactions were skipped because sparse subgroup combinations made the models unreliable or unidentified: Pet ownership × Age group and Age group × Height group.

B.5 Summary

Across the one-way ANOVA analyses, none of the tested grouping variables significantly predicted mean stopping distance. Culture, gender, pet ownership, student status and height group all produced non-significant effects, and the age-group ANOVA was skipped due to a small Gen X subgroup. The results therefore do not support a reliable main effect of these demographic variables under the parametric ANOVA framework.

The two-way ANOVA analyzes did not identify significant interaction effects under conventional $\alpha = 0.5$, but Gender × Culture almost did. Pairwise demographic interaction tests also produced no significant effects.

Overall, the ANOVA analyzes are consistent with the non-parametric approach in showing no conventionally significant evidence that mean stopping distance differed systematically across the tested groups.

C

Non-Parametric Statistical Analyses

This appendix reports the non-parametric statistical analyzes conducted on mean stopping distance. The outcome variable was each participant's mean stopping distance to the drone, measured in metres.

C.1 Main Effects

Table C.1 reports the descriptive statistics for the main group comparisons. Table C.2 reports the corresponding non-parametric test results.

Table C.1: Descriptive statistics for non-parametric main-effect analyzes.

Variable	Group	n	Median	Mean	SD
Culture	Swedish	37	0.885	0.929	0.254
Culture	Vietnamese	32	0.948	0.969	0.319
Gender	Male	43	0.912	0.984	0.288
Gender	Female	26	0.908	0.888	0.274
Pet ownership	No pets	25	0.885	0.903	0.245
Pet ownership	Has pets	44	0.919	0.974	0.304
Student status	Non-student	24	0.902	0.913	0.294
Student status	Student	45	0.928	0.967	0.281
Age group	Gen Z (≤ 29)	60	0.926	0.969	0.281
Age group	Millennial (30–45)	5	0.634	0.733	0.254
Age group	Gen X (46+)	4	0.783	0.905	0.325
Height group	Short ($<$ 168 cm)	23	0.924	0.918	0.301
Height group	Medium (168–175 cm)	24	0.936	1.007	0.302
Height group	Tall ($\geq$ 175 cm)	22	0.880	0.915	0.248

None of the main effects reached statistical significance at the conventional $\alpha = .05$ threshold. Consequently, none survived the Bonferroni-corrected threshold 5.1. The results therefore do not provide evidence that mean stopping distance differed reliably by culture, gender, pet ownership, student status, age group or height group.

Table C.2: Non-parametric main-effect tests for mean stopping distance.

Variable	Test statistic	p	Effect size
Culture	$U = 547.0$	.5923	$r = -0.076$
Gender	$U = 652.0$	.2520	$r = +0.166$
Pet ownership	$U = 481.0$	.3925	$r = -0.125$
Student status	$U = 484.0$	.4844	$r = -0.104$
Age group	$H = 3.288$	.1932	$\eta^2 = 0.020$
Height group	$H = 1.202$	.5483	$\eta^2 = -0.012$

The age-group result should be interpreted cautiously because the Gen X group contained only four participants.

C.2 Culture-Stratified Analyses

To explore whether demographic effects differed between the Swedish and Vietnamese samples, each demographic variable was tested separately within each cultural group.

Table C.3: Culture-stratified non-parametric analyses.

Variable	Culture	Test statistic	p	Effect size
Gender	Swedish	$U = 164.0$	.8669	$r = -0.035$
Gender	Vietnamese	$U = 151.0$	.0488	$r = +0.459$
Pet ownership	Swedish	$U = 158.0$	.8077	$r = +0.053$
Pet ownership	Vietnamese	$U = 86.0$	.1557	$r = -0.304$
Student status	Swedish	$U = 173.0$	.9393	$r = +0.018$
Student status	Vietnamese	$U = 63.0$	.2809	$r = -0.280$
Age group	Swedish	$H = 0.566$	.7536	$\eta^2 = -0.042$
Age group	Vietnamese	$H = 3.491$	.0617	$\eta^2 = 0.083$
Height group	Swedish	$H = 0.598$	.7416	$\eta^2 = -0.041$
Height group	Vietnamese	$H = 2.388$	.3030	$\eta^2 = 0.013$

Gender within the Vietnamese sample, $U = 151.0$, $p = .0488$, $r = +0.459$, was significant under conventional $\alpha = .05$. Vietnamese male participants had a higher mean stopping distance than Vietnamese female participants. However, based on an imbalanced subgroup comparison and not satisfying the Bonferroni-corrected threshold 5.1, this result should be viewed with caution.

No corresponding gender effect was observed in the Swedish sample. Pet ownership, student status, age group and height group were not significant within either cultural group. The Vietnamese age-group comparison approached conventional significance $p = .0617$, with a medium effect size, but the Millennial subgroup contained only two participants and the Gen X subgroup was absent, making this comparison unreliable for inference.

C.3 Pairwise Demographic Stratified Analyses

Additional analyzes tested whether the effect of one demographic variable changed across levels of another demographic variable. This was done using stratified non-parametric tests.

Table C.4: Pairwise demographic stratified non-parametric analyses.

Effect tested	Stratifier	Stratum	Test statistic	p	Effect size
Pet ownership	Gender	Male	$U = 124.0$	.0959	$r = -0.333$
Pet ownership	Gender	Female	$U = 109.0$	.2184	$r = +0.290$
Student status	Gender	Male	$U = 162.0$	.3901	$r = -0.169$
Student status	Gender	Female	$U = 87.0$	.8355	$r = +0.055$
Age group	Gender	Male	$H = 1.184$	.5533	$\eta^2 = -0.020$
Age group	Gender	Female	$H = 2.374$	.3051	$\eta^2 = 0.016$
Height group	Gender	Male	$H = 3.801$	.1495	$\eta^2 = 0.045$
Height group	Gender	Female	$H = 0.803$	.3702	$\eta^2 = -0.008$
Student status	Pet ownership	No pets	$U = 66.0$	.5654	$r = -0.143$
Student status	Pet ownership	Has pets	$U = 189.0$	.7576	$r = -0.062$
Age group	Pet ownership	No pets	$H = 2.419$	.2983	$\eta^2 = 0.019$
Age group	Pet ownership	Has pets	$H = 1.546$	.4617	$\eta^2 = -0.011$
Height group	Pet ownership	No pets	$H = 1.616$	.4457	$\eta^2 = -0.017$
Height group	Pet ownership	Has pets	$H = 0.479$	.7868	$\eta^2 = -0.037$
Age group	Student status	Non-student	$H = 2.796$	.2471	$\eta^2 = 0.038$
Height group	Student status	Non-student	$H = 0.010$	.9950	$\eta^2 = -0.095$
Height group	Student status	Student	$H = 1.346$	.5103	$\eta^2 = -0.016$
Height group	Age group	Gen Z	$H = 1.260$	.5326	$\eta^2 = -0.013$
Height group	Age group	Millennial	$H = 2.133$	.3442	$\eta^2 = 0.067$
Height group	Age group	Gen X	$H = 1.800$	.4066	$\eta^2 = -0.200$

None of the pairwise demographic stratified analyzes reached conventional statistical significance. Several of the relevant strata contained very small sample sizes; some too small to conduct the test. These results should therefore be treated with caution.

C.4 Summary

The non-parametric analyzes found no statistically significant main effects of culture, gender, pet ownership, student status, age group or height group on mean stopping distance. None of the main effects survived the Bonferroni-corrected threshold 5.1.

The culture-stratified analyzes revealed one conventionally significant result: within the Vietnamese sample, male participants stopped the drone farther away than female participants. This effect was medium-sized but should be interpreted cautiously because it was based on a smaller subgroup comparison and did not survive the Bonferroni-corrected threshold 5.1. No equivalent gender effect was observed in the Swedish sample.

The remaining stratified analyzes did not provide reliable evidence of demographic interactions. Several subgroup comparisons were limited by sparse cells, especially those involving age group and height group. Overall, the non-parametric results suggest that the measured demographic variables did not reliably explain differences in mean stopping distance in the full sample. This supports the broader interpretation that individual variation, rather than the variables examined in the study, accounted for much of the variability in human-drone stopping-distance preferences.

D

Bayesian Model Diagnostics

The Bayesian model was fit to the full dataset using four chains and 4000 iterations per chain, using 2000 warmup and 2000 post-warmup sampling iterations per chain, giving 8000 posterior draws in total. Diagnostics were evaluated using $\hat{R} < 1.01$ and $n_{\text{eff}} > 400$, corresponding to four chains with 2000 post-warmup iterations per chain.

D.1 Population-level parameter diagnostics

Parameter	Mean	SD	5.5%	94.5%	$\hat{R}$	n_{eff}
a	4.6326	0.0730	4.5131	4.7495	1.0056	502.4
bG	-0.2057	0.0962	-0.3509	-0.0454	1.0069	477.0
bC	-0.0935	0.0880	-0.2310	0.0484	1.0042	480.5
bA	-0.0378	0.0356	-0.0957	0.0189	1.0051	1138.8
bH	-0.0811	0.0520	-0.1604	0.0054	1.0040	651.2
sigma_subject	0.2910	0.0259	0.2525	0.3342	1.0016	3107.7
scale	2.3906	0.0927	2.2491	2.5445	1.0027	3384.8

D.2 Per-subject intercept diagnostics

Parameter	Mean	SD	5.5%	94.5%	$\hat{R}$	n_{eff}
a_subject[1]	4.6810	0.1672	4.4158	4.9457	1.0048	716.7
a_subject[2]	5.1811	0.0784	5.0537	5.3032	1.0076	617.2
a_subject[3]	5.1081	0.1439	4.8830	5.3396	1.0050	1096.7
a_subject[4]	4.3552	0.0756	4.2327	4.4735	1.0030	875.5
a_subject[5]	4.9086	0.0764	4.7825	5.0274	1.0032	739.8
a_subject[6]	4.1839	0.0978	4.0216	4.3342	1.0025	815.5
a_subject[7]	4.8244	0.1026	4.6572	4.9851	1.0031	709.7
a_subject[8]	4.4559	0.1539	4.2127	4.7046	1.0027	902.9
a_subject[9]	4.6233	0.1469	4.3957	4.8604	1.0046	1035.9
a_subject[10]	4.5889	0.0659	4.4822	4.6912	1.0022	908.6

Parameter	Mean	SD	5.5%	94.5%	$\hat{R}$	n_{eff}
a_subject[11]	4.8474	0.1066	4.6712	5.0111	1.0066	550.1
a_subject[12]	4.8276	0.0802	4.6996	4.9539	1.0057	719.3
a_subject[13]	4.6809	0.0929	4.5313	4.8263	1.0035	772.4
a_subject[14]	4.5775	0.0810	4.4474	4.7060	1.0077	756.1
a_subject[15]	4.6558	0.0811	4.5258	4.7825	1.0068	739.1
a_subject[16]	4.8135	0.0523	4.7267	4.8948	1.0028	964.8
a_subject[17]	4.4639	0.0394	4.4002	4.5269	1.0010	2603.2
a_subject[18]	4.7877	0.0494	4.7080	4.8661	1.0014	1118.0
a_subject[19]	4.6266	0.1141	4.4392	4.8013	1.0061	537.2
a_subject[20]	5.0071	0.1129	4.8194	5.1803	1.0056	539.5
a_subject[21]	4.6307	0.1143	4.4450	4.8084	1.0038	720.5
a_subject[22]	4.1008	0.0893	3.9571	4.2420	1.0052	748.7
a_subject[23]	4.9367	0.0554	4.8451	5.0225	1.0024	860.1
a_subject[24]	4.3395	0.0462	4.2652	4.4129	1.0015	1661.3
a_subject[25]	4.8377	0.0868	4.6941	4.9693	1.0061	546.3
a_subject[26]	4.5329	0.0352	4.4759	4.5885	1.0014	3688.6
a_subject[27]	4.4246	0.0824	4.2935	4.5529	1.0075	723.6
a_subject[28]	4.2191	0.1202	4.0171	4.4081	1.0034	786.7
a_subject[29]	4.1355	0.1058	3.9617	4.2999	1.0035	799.2
a_subject[30]	4.5212	0.0632	4.4174	4.6197	1.0026	917.1
a_subject[31]	4.5841	0.0811	4.4549	4.7117	1.0076	804.7
a_subject[32]	4.6438	0.0904	4.4957	4.7837	1.0069	551.3
a_subject[33]	4.4782	0.1102	4.2980	4.6507	1.0035	717.7
a_subject[34]	4.4642	0.0610	4.3643	4.5591	1.0040	958.6
a_subject[35]	5.0781	0.0802	4.9502	5.2039	1.0063	819.8
a_subject[36]	4.7176	0.0834	4.5815	4.8454	1.0067	635.2
a_subject[37]	4.5850	0.0727	4.4658	4.6966	1.0029	831.8
a_subject[38]	4.9632	0.0833	4.8303	5.0936	1.0016	602.8
a_subject[39]	4.8410	0.1206	4.6431	5.0307	1.0041	527.4
a_subject[40]	4.5232	0.0841	4.3898	4.6535	1.0016	664.8
a_subject[41]	4.4830	0.0882	4.3412	4.6220	1.0025	635.1
a_subject[42]	4.2946	0.1283	4.0914	4.4964	1.0034	459.4
a_subject[43]	4.7067	0.1163	4.5219	4.8914	1.0033	461.8
a_subject[44]	4.3005	0.1721	4.0231	4.5695	1.0044	467.4
a_subject[45]	4.6228	0.0908	4.4751	4.7650	1.0033	535.0
a_subject[46]	4.2936	0.1538	4.0476	4.5385	1.0045	551.1
a_subject[47]	4.7239	0.1108	4.5415	4.8951	1.0044	480.8
a_subject[48]	4.7367	0.1267	4.5348	4.9383	1.0034	413.7
a_subject[49]	4.5834	0.0933	4.4331	4.7302	1.0026	593.5
a_subject[50]	4.7357	0.0905	4.5904	4.8776	1.0024	563.7
a_subject[51]	5.1546	0.1176	4.9617	5.3353	1.0049	515.0
a_subject[52]	4.4733	0.0800	4.3443	4.5997	1.0015	772.7
a_subject[53]	5.0273	0.0771	4.9043	5.1491	1.0018	737.6
a_subject[54]	4.2645	0.1085	4.0904	4.4422	1.0031	761.6
a_subject[55]	4.2584	0.1070	4.0833	4.4262	1.0027	562.2

Parameter	Mean	SD	5.5%	94.5%	$\hat{R}$	n_{eff}
a_subject[56]	5.1977	0.0834	5.0627	5.3273	1.0025	559.2
a_subject[57]	4.9838	0.0846	4.8479	5.1154	1.0024	565.5
a_subject[58]	4.7182	0.0937	4.5670	4.8627	1.0045	540.0
a_subject[59]	4.8622	0.0847	4.7260	4.9952	1.0016	557.7
a_subject[60]	4.7998	0.1136	4.6188	4.9807	1.0032	501.1
a_subject[61]	4.1595	0.1177	3.9715	4.3437	1.0029	655.2
a_subject[62]	4.7655	0.1321	4.5527	4.9716	1.0040	425.6
a_subject[63]	4.9121	0.0982	4.7548	5.0634	1.0031	502.1
a_subject[64]	4.8859	0.0848	4.7493	5.0192	1.0018	562.1
a_subject[65]	4.2118	0.1217	4.0106	4.4016	1.0041	542.3
a_subject[66]	4.1866	0.0853	4.0523	4.3214	1.0032	920.9
a_subject[67]	4.2463	0.0932	4.0964	4.3945	1.0017	713.5
a_subject[68]	4.4687	0.1084	4.2938	4.6387	1.0035	565.9
a_subject[69]	4.7283	0.1200	4.5374	4.9151	1.0030	433.8

D.3 Convergence summary

Across all population-level and subject-level parameters, the maximum $\hat{R}$ was 1.0077 and the minimum effective sample size was 413.7. Thus, all reported parameters satisfied the convergence threshold $\hat{R} < 1.01$ and the effective sample-size threshold $n_{\text{eff}} > 400$.

E

Leave-One-Subject-Out Model Comparison Full Report

The LOSO comparison evaluated a null intercept-only model against a full demographic model. Each model was fit using four chains and 4000 iterations per chain, using 2000 warmup and 2000 post-warmup sampling iterations per chain, giving 8000 posterior draws per fit.

For each held-out subject, both models were trained on the remaining subjects and evaluated on the held-out subject. The full model was favored when $\Delta\text{LPD} = \text{LPD}_{\text{full}} - \text{LPD}_{\text{null}} > 0$, equivalently when $\Delta\text{deviance} = -2\Delta\text{LPD} < 0$.

E.1 Per-subject predictive diagnostics

Subject	LPD null	LPD full	Delta LPD	Delta dev.
1	-78.639	-78.552	0.088	-0.175
2	-89.289	-89.478	-0.189	0.379
3	-85.880	-86.853	-0.973	1.945
4	-74.903	-75.008	-0.106	0.211
5	-85.153	-85.163	-0.010	0.020
6	-72.859	-72.883	-0.024	0.048
7	-87.858	-87.958	-0.100	0.199
8	-61.489	-61.243	0.246	-0.493
9	-78.077	-78.130	-0.053	0.106
10	-96.658	-96.678	-0.020	0.041
11	-87.828	-88.034	-0.206	0.412
12	-82.849	-82.932	-0.083	0.167
13	-80.294	-80.312	-0.018	0.037
14	-81.302	-81.330	-0.029	0.057
15	-88.502	-88.500	0.002	-0.004
16	-82.393	-82.246	0.147	-0.295
17	-84.530	-84.694	-0.164	0.328
18	-84.878	-84.698	0.180	-0.360
19	-81.698	-81.642	0.056	-0.112

Subject	LPD null	LPD full	Delta LPD	Delta dev.
20	-89.270	-89.783	-0.512	1.025
21	-89.981	-89.992	-0.011	0.022
22	-58.877	-59.083	-0.206	0.412
23	-85.937	-85.797	0.140	-0.281
24	-80.892	-81.185	-0.292	0.585
25	-83.967	-84.069	-0.102	0.204
26	-82.619	-82.675	-0.056	0.113
27	-80.138	-80.214	-0.076	0.152
28	-52.183	-52.022	0.161	-0.323
29	-40.299	-39.939	0.359	-0.719
30	-81.236	-81.285	-0.049	0.098
31	-83.925	-83.950	-0.025	0.050
32	-85.898	-85.904	-0.007	0.013
33	-78.346	-78.301	0.045	-0.090
34	-81.175	-81.309	-0.134	0.268
35	-86.698	-86.658	0.040	-0.081
36	-86.726	-86.758	-0.032	0.064
37	-83.426	-83.433	-0.007	0.013
38	-88.445	-88.196	0.250	-0.499
39	-91.120	-91.237	-0.117	0.233
40	-82.686	-82.739	-0.053	0.106
41	-96.123	-96.221	-0.098	0.196
42	-75.291	-75.012	0.278	-0.556
43	-83.516	-83.565	-0.050	0.099
44	-37.553	-36.821	0.732	-1.464
45	-89.341	-89.323	0.018	-0.035
46	-41.069	-40.434	0.635	-1.270
47	-84.245	-84.256	-0.010	0.021
48	-80.992	-81.063	-0.071	0.142
49	-83.749	-83.752	-0.003	0.006
50	-201.803	-204.294	-2.491	4.983
51	-83.264	-83.491	-0.228	0.456
52	-87.121	-87.269	-0.149	0.297
53	-119.461	-118.902	0.559	-1.118
54	-73.947	-74.169	-0.223	0.445
55	-57.347	-57.462	-0.115	0.231
56	-93.710	-93.418	0.292	-0.585
57	-87.506	-87.303	0.203	-0.407
58	-126.152	-126.121	0.031	-0.061
59	-101.724	-101.561	0.163	-0.325
60	-95.895	-95.969	-0.074	0.148
61	-69.979	-70.137	-0.159	0.317
62	-91.425	-91.535	-0.110	0.220
63	-120.448	-120.392	0.056	-0.112
64	-88.356	-88.176	0.180	-0.360

Subject	LPD null	LPD full	Delta LPD	Delta dev.
65	-70.929	-70.907	0.023	-0.045
66	-81.495	-82.340	-0.844	1.689
67	-50.825	-51.186	-0.360	0.721
68	-84.717	-84.709	0.008	-0.016
69	-107.593	-107.634	-0.041	0.083

E.2 Predictive-performance summary

Table E.2: LOSO predictive-performance summary for the null and full Bayesian models.

Quantity	Value
Total LPD, null	-5804.50
Total LPD, full	-5808.28
Total deviance, null	11608.99
Total deviance, full	11616.57
Delta deviance, full–null	7.57
Mean per-subject delta deviance	0.1098
SD of per-subject delta deviance	0.7844
SE of mean delta deviance	0.0944
95% CI for mean delta deviance	[-0.0753, 0.2948]

The 95% interval for the mean per-subject delta deviance included zero, indicating no clear predictive advantage for either model. The positive total delta deviance indicates that the null model had slightly better aggregate predictive performance, but the difference was small relative to the subject-level uncertainty.

F

Research Consent Form

Research Consent Form

Title of study: A Rubber-Sheet Transformation Model for Human–Drone Proxemics in the Context of Social Drones

Purpose of the Study

You are invited to participate in a research study conducted at Chalmers University of Technology. The purpose of this study is to investigate how individuals determine preferred distances when approached by an autonomous drone in a three-dimensional augmented reality environment. The results will be used to develop predictive models of human-drone interaction distances.

Your participation will take approximately 30 minutes.

Procedure

If you agree to participate, you will:

- Wear augmented reality equipment
- Observe a virtual drone approaching from different directions and altitudes
- Indicate when the drone reaches a distance that you consider acceptable
- Complete a short demographic questionnaire

No physical drone will approach you. All drone movements are simulated in augmented reality.

Voluntary Participation

Participation is voluntary. You may withdraw at any time without giving a reason and without negative consequences. If you choose to withdraw, your collected data will be removed upon request.

Risks

The study does not involve risks beyond those encountered in everyday life. Some participants may experience mild discomfort from wearing AR equipment. You may pause or stop the session at any time. Access to your data will be limited to the research team.

Data Collection and Confidentiality

The following data will be collected:

- Preferred distance measurements

- Drone approach parameters

- Demographic information such as age, gender, and height
- Short post-study questionnaire

No identifying personal information will be linked to your responses. Each participant will be assigned a numerical ID.

All data will be anonymized and stored securely. The data will be used only for academic research purposes, including a master's thesis and potential scientific publications. Results will be reported in aggregated form, and no individual participant will be identifiable.

Contact Information

Researchers:
Hugo Raimer (raimer@chalmers.se)
Vilmer Malm (vilmerm@chalmers.se)

Supervisor:
Dr. Mohammad Obaid (mobaid@chalmers.se)

Dr. Fanta Camara (University of York)

If you have questions about the study, please contact the researcher or supervisor using the information above.

Consent

By signing below, you confirm that:

- You have read and understood the information above.

- You understand that participation is voluntary and that you are free to withdraw the use of your data without giving any reason. If withdrawn the data will be removed and destroyed.

- You consent to the use of your anonymized data for educational purposes, academic research and publication.

I agree to the above statements:

Participant Name (Printed): ______________________________

Signature: ____

Date: ______________________

G

Written Vietnamese Instructions (with translation)

Hướng dẫn:

Quá trình này sẽ mất khoảng 15 phút.

Một thiết bị bay không người lái (drone) xã hội đang tiếp cận bạn, nó ở đây để giúp đỡ bạn. Đây KHÔNG phải là thiết bị quân sự hay giám sát. Hãy coi nó như một người bạn, xuất hiện để nói điều gì đó hoặc hỗ trợ bạn một việc gì đó.

- Hãy đứng vào dấu X trên mặt đất, bạn chỉ được phép cử động đầu, không được di chuyển chân hay cơ thể.
- Drone sẽ bay đến vị trí bắt đầu và đối diện với bạn.
- Khi bạn đã sẵn sàng để drone tiếp cận, hãy nói “start” (bắt đầu).
- Khi bạn cảm thấy không muốn drone lại gần hơn nữa, hãy nói “stop” (dừng lại). Hãy cân nhắc khoảng cách mà bạn mong muốn đối với một drone đang đến để tương tác với mình trong đời thực.
- Luôn nhìn vào drone khi nó đang tiến về phía bạn.

Sau khi bạn nói “stop”, bạn sẽ nghe thấy một tiếng bíp, điều này có nghĩa là bạn cần nhìn xuống dấu X trên mặt đất ngay trước mặt mình. Quan trọng: Trong khi nhìn xuống đất, bạn không được nhìn drone khi nó đang điều chỉnh lại vị trí.

Khi bạn nghe thấy tiếng bíp thứ hai, bạn có thể nhìn lên drone một lần nữa và nói “start”. Lặp lại việc này một số lần cho đến khi chúng tôi báo bạn đã hoàn thành.

Đầu tiên sẽ có một đợt chạy thử nhỏ, drone sẽ bay xung quanh để bạn làm quen với môi trường thực tế tăng cường (AR).

Sau đó, sẽ có một số câu hỏi để bạn trả lời.

Cảm ơn bạn đã tham gia!

Instructions:

This process will take approximately 15 minutes.

A social unmanned aerial vehicle (drone) is approaching you. It is here to help you. This is NOT a military or surveillance device. Think of it as a friend, appearing to say something or assist you with something.

- Stand on the X mark on the ground. You are only allowed to move your head; do not move your feet or body.
- The drone will fly to its starting position and face you.
- When you are ready for the drone to approach, say "start."
- When you feel that you do not want the drone to come any closer, say "stop." Please consider the distance you would prefer for a drone approaching you to interact with you in real life.
- Always look at the drone while it is moving toward you.

After you say "stop," you will hear a beep. This means you need to look down at the X mark on the ground directly in front of you. Important: While looking down at the ground, you must not look at the drone while it is repositioning itself.

When you hear the second beep, you may look back up at the drone and say "start" again. Repeat this several times until we inform you that you are finished.

First, there will be a short practice run in which the drone flies around so that you can become familiar with the augmented reality (AR) environment.

After that, there will be several questions for you to answer.

Thank you for participating!