Indoor Human–Mobile Robot Encounters: A Transdisciplinary Study on Perceived Safety

RYAN GUPTA, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas, USA
EMILY NORMAN, Department of Communications Studies, The University of Texas at Austin, Austin, Texas, USA
HYONYOUNG SHIN, Department of Electrical and Computer Engineering and Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas, USA
ZHIYUN DENG, Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas, USA
MARIA ESTEVA, Texas Advanced Computing Center, The University of Texas at Austin, Austin, Texas, USA
NANSHU LU, Department of Electrical and Computer Engineering and Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas, USA
KERI K. STEPHENS, Department of Communications Studies, The University of Texas at Austin, Austin, Texas, USA
LUIS SENTIS, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas, USA

Despite the rise of mobile robot deployments in community settings, the perceived safety of cohabitants remains understudied in many domains. To address this gap, we perform a study to identify elements of indoor human–mobile robot encounters that impact perceived safety. This study evaluates the effects of robot movement behavior and the number of robots nearby on perceived safety of participants. Further, this article investigates how the presence of other people impacts perceived safety in such settings. We leverage

This research was supported in part by National Science Foundation (NSF) Award #2219236 (GCR: Community Embedded Robotics: Understanding Sociotechnical Interactions with Long-term Autonomous Deployments) and Living and Working with Robots, a core research project of Good Systems, a UT Grand Challenge. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.
Authors' Contact Information: Ryan Gupta (corresponding author), Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas, USA; e-mail: ryan.gupta@utexas.edu; Emily Norman, Department of Communications Studies, The University of Texas at Austin, Austin, Texas, USA; e-mail: emilynorman@utexas.edu; Hyonyoung Shin, Department of Electrical and Computer Engineering and Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas, USA; e-mail: hyonyoung.shin@utexas.edu; Zhiyun Deng, Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas, USA; e-mail: zdeng@utexas.edu; Maria Esteva, Texas Advanced Computing Center, The University of Texas at Austin, Austin, Texas, USA; e-mail: maria@tacc.utexas.edu; Nanshu Lu, Department of Electrical and Computer Engineering and Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas, USA; e-mail: nanshulu@utexas.edu; Keri K. Stephens, Department of Communications Studies, The University of Texas at Austin, Austin, Texas, USA; e-mail: keristephens@austin.utexas.edu; Luis Sentis, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas, USA; e-mail: lsentis@utexas.edu.

ACM 2573-9522/2026/5-ART70
https://doi.org/10.1145/3803850

methodologies from physiological signal analysis, autonomy, surveys, and qualitative interviews to decode insights into the human experience during such encounters. Particularly, signal analysis yielded that the presence of multiple robots decreases perceived safety and that search behaviors were more comfortable than navigation behaviors. Similarly, interviews with participants demonstrated clear effects on perceived safety in the presence of others, and that sensemaking was a key component involved in their perceptions of safety. When the data were combined, interview data revealed that near collisions between the robots likely confounded the signal analysis findings with respect to the number of robots and their movement behavior. The data types agree that the presence of a robot impacts perceived safety; however, there are also conflicting results that we discuss, which highlight that near-collisions impact perceived safety. In aggregate, the study illustrates the benefits of leveraging eclectic methods to ascertain deeper insights. Overall, the article aims to unlock insights into human perceptions during encounters with community embedded robots, which can be used in the future design of such systems.

CCS Concepts: • **Computer systems organization** → **Robotic autonomy**; • **Human-centered computing** → **User studies**; **HCI theory, concepts and models**;

Additional Key Words and Phrases: Perceived Safety, Community Embedded Robotics, Transdisciplinary Perspective, Physiological Signal Data, Comfort

ACM Reference format:
Ryan Gupta, Emily Norman, Hyonyoung Shin, Zhiyun Deng, Maria Esteva, Nanshu Lu, Keri K. Stephens, and Luis Sentis. 2026. Indoor Human–Mobile Robot Encounters: A Transdisciplinary Study on Perceived Safety. *ACM Trans. Hum.-Robot Interact.* 15, 4, Article 70 (May 2026), 41 pages.
https://doi.org/10.1145/3803850

1 Introduction

Autonomous mobile robot research has been ongoing for decades in a wide range of real-world scenarios, including delivery [Lee et al., 2021], search [Zheng et al., 2023], surveillance [Grocholsky et al., 2006], mapping [Charrow et al., 2015], and inspection [Tan et al., 2021]. These tasks are often performed in dynamic and unpredictable environments, where individual robots may be limited in capability or vulnerable to failure [Rizk et al., 2019]. Multi-robot systems, extensively studied over the years, address these challenges by offering robustness, adaptability, and resilience through distributed coordination and redundancy [Prorok et al., 2021].

As teams of autonomous robots advance in capability and proliferate in communities, designers and engineers must consider the impacts these systems have on the users, onlookers, and the places they encounter. There exists a vast body of research concerned with physically safe **Human–Robot Interaction (HRI)** [Rubagotti et al., 2022], for example, in social navigation [Francis et al., 2023]. While physical safety is a requirement for HRI, perceived safety, synonymous with psychological safety and comfort, is also necessary [Akalin et al., 2023]. Perceived safety reflects people's perceptions and feelings and is found to involve six factors [Akalin et al., 2023]: Context of Robot Use, Comfort, Experience and Familiarity, Predictability and Transparency, Sense of Control, and Trust.

In recent years, perceived safety has grown in significance in the research community as it becomes increasingly necessary to deter negative psychological outcomes in situations where humans and robots coexist in common spaces, even if they are not directly interacting. While notions of encounters and interactions represent shared space and a level of mutual awareness, we recognize a distinction. The distinction follows naturally to the extent a direct or indirect interaction between two humans differ. Direct interactions among humans and robots impart more depth and impact, yet encounters will dominate the total number of exchanges. Particularly, robots deployed by organizations will encounter community members who lack knowledge of the robot's purpose, capability, or mission. In this work, interactions require direct involvement between the human and the robot, characterized by overlapping goals, conversation and/or clearly defined roles. Encounters capture the remainder, for example, when a community member works in a shared space, a maintenance worker completes their work nearby.

Many studies employ a single research method to evaluate peoples' perceptions in the presence of robots. These include self-report questionnaires and qualitative open-form responses like interviews and physiological response. Subjective questionnaires have been successfully employed to measure preference and human mental state during HRI [Bartneck et al., 2009; Hauser et al., 2023; Moesgaard et al., 2022; Petrak et al., 2019; Taylor et al., 2022; Van Waveren et al., 2023]. However, questionnaires are limited by their poor temporal resolution and sources of bias including moderacy response bias and memory effects induced by repetitive administration. Analysis of open-form responses is also shown to offer valuable insights into HRI [Axelsson et al., 2024; Döring and Poeschl, 2019; Fortunati et al., 2021]. For example, **Thematic Analysis (TA)** of interview data offers flexibility, enables both inductive and deductive inquiry, and generates accessible results [Braun and Clarke, 2006; Clarke and Braun, 2017]. However, these results are limited due to their subjectivity. Others leverage more objective physiological measures to ascertain human internal state during HRI [Chen et al., 2011; Dehais et al., 2011; Stark et al., 2018; Willemse et al., 2017]. Non-invasive physiological sensing are hypothesized to enable the tracking of human internal state over time for example to evaluate stress during HRI [Greco et al., 2021; Huh et al., 2023; Turner-Cobb et al., 2019]. However, analysis of biological markers lacks individuals' feelings, contexts, and experiences that are captured through social science methods. Community-embedded robotics is a highly complex field, with decades of research dedicated to a diversity of components, including perceived safety. Given this complexity and the limitations associated with each research method, a transdisciplinary approach is required to fully understand human response to these situations. A fundamental prerequisite of technologies such as robotics is to ensure that it benefits the communities and people it encounters, which requires a combined approach in order to form a more complete picture of peoples' perceptions.

Motivated by encounters with single and multi-robot systems and to address the aforementioned shortcomings associated with each research method, this work employs three complementary methods to provide a deeper understanding of perceived safety. Questionnaires gather subjective measures of the specific robot conditions. Non-invasive physiological sensors enable the quantification of perceived safety during human–robot encounters by decoding acute stress levels from biomarkers found in **Electrocardiography (ECG)** and **Electrodermal Activity (EDA)** data. To the best of our knowledge, this is the first analysis of physiological response to mobile robots navigating freely [Rubagotti et al., 2022]. Our study deploys two quadruped robots in a realistic apartment setting (Figure 1). Interview transcripts are approached from both *deductive* and *inductive* methods, which reveal in depth contextual information, interpretations, and emotions identified by participants. Social science methods like deductive content analysis [Krippendorff, 1989] and inductive TA [Axelsson et al., 2024] enable the study of interdependencies between the participants, environment, and robots. Accordingly, insights regarding perceived safety may be found in these interactions between social and technical factors during human robot encounters. Together, the results provide a comprehensive view of the experiences and reactions of participants during the study. Some of the results from interviews conflict with the results from physiological signal analysis. We highlight this in Section 6.4 and find that near-collisions explain this conflict. Particularly, we analyzed biological markers in individual trials and found spikes when near-collisions occurred between the two quadrupeds. We also ensure reproducibility and facilitate the future use of our data by publishing an extensive, open access dataset including but not limited to all of the raw time series data collected in this study [Gupta et al., 2024]. The dataset unifies insights and vocabulary

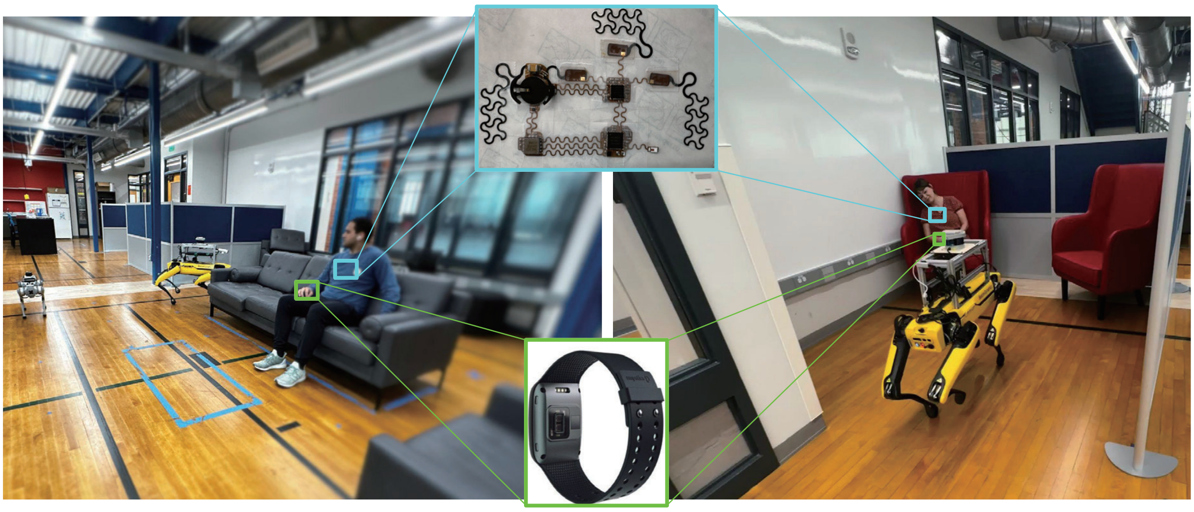

Fig. 1. An overview of the experiments. On the left, one author sits in the living room seating configuration, while on the right, another sits in the small room configuration. The Unitree Go1 and Boston Dynamics Spot perform various mobile behaviors in shared spaces with UT community members. Physiological measures are collected from the E4 wristwatch (green) [Empatica, 2024] and chest e-tattoo (cyan) [Bhattacharya et al., 2023].

between data types using a knowledge graph representation of the multimodal database such that these findings apply beyond any individual discipline. This knowledge graph enables users to query the database for specific data. The full dataset is published open access for validation and reuse.

This article begins with a background on perceived safety in HRI, Social Navigation, human studies, and stress decoding from physiological signal analysis in Section 2. Next, we present our research questions in Section 3 and explain the experiment design and data acquisition methods in Section 4. Section 5 explains the data analysis techniques used for both the quantitative data taken from physiological signals and surveys and the qualitative data taken from interviews with participants. Next, Section 6 demonstrates the results from these methods. Particularly, Section 6.1 investigates the statistical classification of stress with respect to the control variables; Section 6.2 investigates the interview transcripts to ascertain whether the sensors themselves played a role in any discomfort felt by participants; and Section 6.3 presents the qualitative findings. The results conclude with Section 6.4, which takes an interdisciplinary approach to the relationship between the qualitative and quantitative findings. This study deliberately merges multiple perspectives and disciplines, facilitating a richer, more comprehensive interpretation of the findings. Section 7 discusses the key findings before the open source dataset, and the accompanying database is presented in Section 7.1. Future research directions are presented in Section 7.2 before we conclude with Section 8. The findings can inform researchers and roboticists on the key aspects to account for during the design of mobile robots for indoor environments where community members may co-exist and suggest a way forward beyond disciplinary boundaries.

2 Background

2.1 Perceived Safety in HRI

We identify three broad categories of studies concerning perceived safety and mental stress during HRI and human–robot encounters with mobile robots. The first group of research studies human physiological responses to estimate certain features like stress [Staffa and Rossi, 2022; Turner-Cobb et al., 2019] or mental workload [Zhou et al., 2020]. Often these studies integrate robots alongside human activities with known external stressors like seminar-style presentations or cognitive quizzes that evoke discomfort [Akalin et al., 2022]. They ask how replacing humans with robots impacts people's response to situations. However, none are designed to study mobile robots moving freely through shared environments. In particular, Turner-Cobb et al. [2019] study participants during the well-known Trier Social Stress Test. In this test, participants are put into a room with several judges. The test involves a 10-min verbal presentation, followed by a mental arithmetic task in front of the live audience [Kirschbaum et al., 1993]. They replace a human audience with a pre-recorded robot audience and estimate stress from salivary cortisol. Instead, Staffa and Rossi [2022] train a neural network to recognize stress from an EEG device during "friendly" and "authoritarian" behaviors from a Pepper robot during mental tests.

The second group of research studies leverages more subjective measures, like statistical evaluation of questionnaire responses [Aliasghari et al., 2023; Cohen et al., 1994; Hauser et al., 2023; Mavrogiannis et al., 2022; Van Waveren et al., 2023; Wang et al., 2023] and qualitative analyses on open-form participant responses [Döring and Poeschl, 2019; Fortunati et al., 2021; Joffe, 2011; Terry et al., 2017]. Mobile robotics studies are well represented in this group of work. However, we find no existing literature that particularly studies perceived safety in the types of indoor encounters studied in this work. In particular, questionnaire data have been used to understand perceived safety due to noise from flying robots [Wang et al., 2023]. Instead, Van Waveren et al. [2023] leverage Likert-scale responses to inform a personalized motion planners for small drones indoors. Interviews and the Godspeed Questionnaire are used to evaluate incidental outdoor encounters with mobile ground robots in Hauser et al. [2023]. Their findings suggest that traditional HRI survey instruments that focus on the perception of robot characteristics may not be suitable for incidental human–robot encounters research. Mavrogiannis et al. [2022] propose a new algorithm and investigate legible robot motion generation using statistical survey analysis. They evaluate encounters while both the people and robots move around the environment. They also leverage TA of data collected from open-form responses; however, they note that their study does not provide an objective measure of comfort. Legible robot actions, which are closely related to Predictability and Transparency, have also been studied in the context of multiple robots [Capelli et al., 2024]. Content analysis is also employed in robotics contexts, for example, to understand media portrayals of humanoid robots and intimate robot–human relationships [Döring and Poeschl, 2019; Fortunati et al., 2021]. Usability studies researchers also adopt content analysis when collecting information about user reviews [Koh et al., 2022]. These approaches vary between qualitative and quantitative content analysis methodology, as well as between inductive and deductive approaches. Researchers typically refer to deductive analysis when applying a theory or existing framework as a guide for data analysis, whereas inductive analysis draws from the raw data itself, sometimes generating new theory [Joffe, 2011; Terry et al., 2017]. The final group of studies falls outside the relevant scope and primarily study specially trained workers who collaborate with robots on a regular basis [Marvel et al., 2020; Villani et al., 2018].

2.2 Social Navigation Studies

The first group of works we discuss with respect to social navigation include work in Proxemics, a concept originally defined by Hall et al. [1968]. Hall [1966] conceptualized the space (proxemics) around a person into four subspaces: the intimate zone ranging from 0 to 0.45 m, the personal zone ranging from 0.46 to 1.20 m, the social zone ranging from 1.21 to 3.50 m, and the public zone from above 3.51 m. It has been shown that not only are human–robot proxemics similar to human–human proxemics but also that proxemics in HRI has an impact on participants [Neef et al., 2023]. Their research has shown that humans tend to prefer greater distances in these interactions and have more positive impressions of robots when they are encountered at greater distances [Neef et al., 2023]. However, in may real-world scenarios, ensuring large distances from community members cannot be guaranteed, which motivates the need to study scenarios like the proposed work.

Leaders in social navigation collaborated on a set of guidelines and open challenges facing the field [Francis et al., 2023]. Under their proposed taxonomy, our study focuses on the principles of Comfort and Safety. Relevant studies in social navigation look at path crossing scenarios in hallways or narrow spaces [Reinhardt et al., 2021], elevators [Gallo et al., 2023], and open areas to simulate crowd navigation [Neggers et al., 2024; Shiomi et al., 2014; Vassallo et al., 2017]. Neggers et al. [2024] point out the need for an accurate representation of comfortable distances for crossing paths in crowds. They estimate comfort in their study by measuring the paths that participants take under crossings with the robot under varied speeds. Reinhardt et al. [2021] propose an autonomous back-off behavior when bottlenecks are encountered and their focus is robot legibility. Particularly, the intent of the back-off is to indicate to the human that they are being given the right of way. Further, Mirsky et al. [2024] focus on conflict avoidance, or situations when either the human or robot (or both) are forces that must deviate in order to avoid a collision or near collision. These studies provide interesting and important results; however, their focus is on a specific moment of an encounter. In contrast, our study looks at the entirety of an encounter, where a robot is performing some task in a shared environment with a person. The authors note that most papers on social navigation focus on the quantifiable and technologically focused aspects of navigation rather than on the experience of the human in the interaction. Our research aims to address this identified gap in the literature.

2.3 Stress Decoding from Physiological Signals

Literature demonstrates that physiological signal recordings contain a wealth of information about human physical and mental states. EDA, or galvanic skin response, measures the changes in skin conductance over time on the palm, foot, or wrist. Skin conductance time series consists of two components: tonic and phasic. Tonic refers to low-frequency changes in the baseline signal, while phasic refers to sharp spikes followed by recovery to the baseline signal. The phasic component is known to correspond to psychological arousal events, allowing quantitative determination of mental stress over time. Decoding acute stress from tonic–phasic decompositions of EDA signals has been demonstrated extensively [Greco et al., 2021].

Another signal used in mental stress detection is **Heart Rate Variability (HRV)** [Kim et al., 2018], which can be evaluated from any modality that reliably records heart beat information. HRV features have been linked to a variety of acute and chronic stressors in real life and experimentally induced situations [Immanuel et al., 2023]. This is due to the fact that HRV reflects complex changes in sympathetic and parasympathetic nervous systems which manage the homeostasis of psychological arousal. HRV is often evaluated over long recordings (5 minutes–24 hours); however, it can be reliably estimated from ultra-short-term and short-term recordings (less than 5 minutes) like our study [Lee et al., 2022; Salahuddin et al., 2007]. Acute stress has been associated with decreases in HRV entropy in real-life situations [Dimitriev et al., 2016; Melillo et al., 2011] and experimentally induced situations [Brugnera et al., 2018; Lee et al., 2022]. **Normal-to-Normal (NN)** intervals, which are used to evaluate HRV features, represent the time between successive heartbeats. Common time domain features are **Standard Deviation of NN Intervals (SDNN)**, which represents total variability over the length of the recording, and **Root Mean Square of Successive Differences between Normal Heartbeats (RMSSD)**. Specifically, decreases in SDNN and RMSSD and increases in low-frequency bandpower (0.04–0.15 Hz) compared to high-frequency (0.15–0.40 Hz) are typical responses to acute stress [Immanuel et al., 2023].

Mental stress estimation from EDA and HRV is non-trivial as these signals indirectly measure changes caused by the autonomic nervous system, thereby acting as proxies for mental stress. To maximize success, raw signals from participants must be as high-quality and free of noise and motion artifacts as possible to reduce the loss of information. While this incentivizes the use of clinical grade equipment, such equipment conflicts with goals of this study, focused on realistic human–robot encounters in an apartment setting. Furthermore, care must be taken to ensure physiological sensors are as comfortable and non-disruptive as possible, such that they do not restrict mobility or reduce perceived safety. We conclude that state-of-the-art wearable devices are more appropriate to avoid lengthy sensor setup, skin preparation, and cables that interfere with movement or cause stress on their own.

3 Research Questions and Hypotheses

The core motivating tenet of this work is toward the design of systems that users perceive as comfortable to share spaces with. From this standpoint and based on the surveyed literature, we identified the following research questions:

- Is perceived safety changed near multiple robots versus a single robot?
- Is perceived safety changed when a robot is searching rather than navigating through a space?
- Does the social versus isolated setting of the participant impact perceived safety around robots?

While many existing studies investigate the impact of proxemics on perceived safety, this article instead considers the impact of multiple robots compared to only one. Particularly, real-world tasks like inventory, inspection, or delivery are often performed using teams of robots due to the resilience and robustness of multi robot systems. Furthermore, search and navigation behaviors represent the general movement characteristics of autonomous robots performing these tasks. Therefore, if perceptions significantly change under these two different behaviors, this information can inform the design of community embedded robotics. Specifically, if search behaviors represent scenarios that provoke discomfort from bystanders, then system designers can ensure that tasks like inventory or inspection are completed during low occupancy times. Finally, the impact of other people nearby on perceived safety is relevant in many contexts. For example, a worker in an office might be likely to encounter a robot while alone, compared with a break room where they are more likely to encounter a robot while in the presence of others. Overall, the research questions and control variables offer insights to the designers and engineers behind community embedded robotics on how their systems will be perceived by the members of their community.

4 Experiment Design

To investigate perceived safety and stress response in the context of indoor human–multi-robot encounters, we conducted a study with 24 total participants over 13 sessions. Participants ages ranged from 19 to 48 while the average age was 29.0 years with a standard deviation of 8.5 years. Eighteen of the participants were students at the University of Texas, while the remaining six were staff members. Finally, 11 participants reported as male and 13 as female. The study was independently reviewed and approved by the University of Texas Institutional Review Board (IRB #00004099). Most sessions involved two participants, while two sessions had one participant. The study sessions consisted of participants sitting and observing various mobile robot behaviors in a 20×30 m apartment setting. During trials, the Unitree Go1 and Boston Dynamics Spot quadrupeds performed navigation and visual search behaviors together and separately. System details and open source code used for the autonomous localization and navigation are described further in Appendix A. After experiments, participants took part in an interview. An overview of experiments is shown in Figure 2.

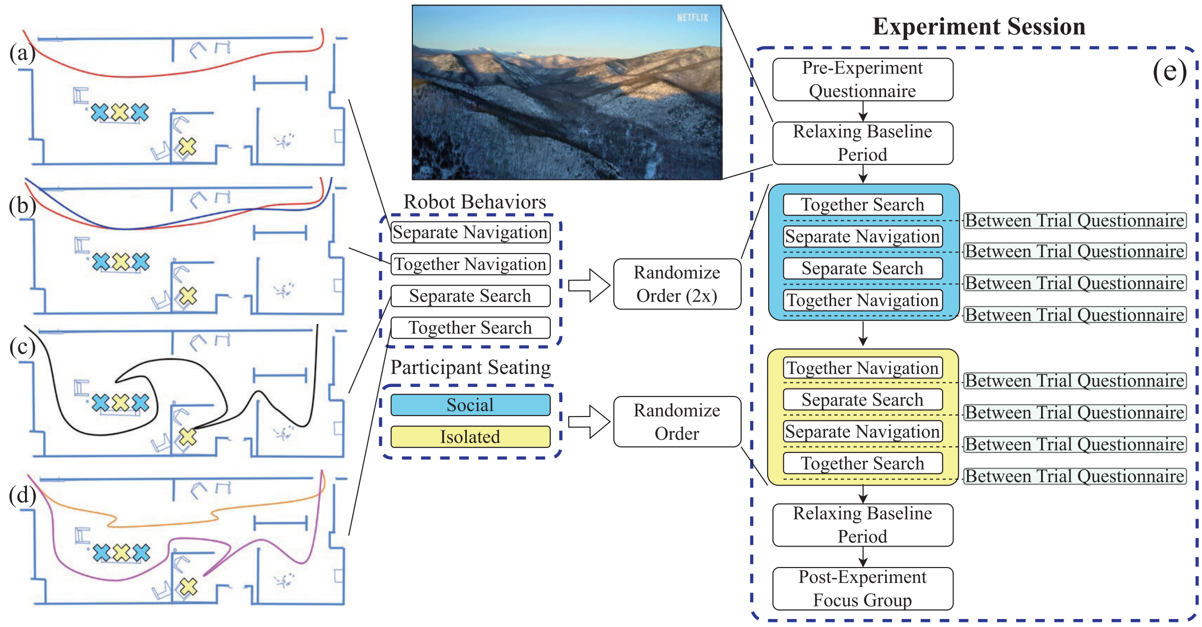

Fig. 2. A visual representation of the three control variables are shown on the left in (a)–(d) including the robot behaviors and participant seating locations, shown as X's. Blue represents social seating and yellow represents isolated seating. (a) The single robot navigation path. (b) The two robot navigation paths. In (c) is the single robot search path and (d) shows the two robot search paths. The order of behaviors and seating locations were randomized and then inserted into the experiment session as overviewed in (e). These experiments were designed to gain insights into human responses to encounter with robots. The first step was to obtain consent from the participants, followed by administering a questionnaire. The nature video was repeated before and after the experimental session to establish a relaxed baseline physiological state. After experiments and sensor removal, participants took part in an interview.

4.1 Conditions

This study controlled three variables: (1) single (separate) vs. multi (together) robot, (2) navigation vs. search behavior, and (3) social vs. isolated participant seating. Figure 2 illustrates these variables on a map of the apartment, showing the four robot behaviors and participant seating locations. The figure shows the separate and together robot navigation paths (Figure 2(a) and (b)) and search paths (Figure 2(c) and (d)). Both the search and the navigation behaviors involve the robots navigating in the shared environment; however, in the search trials, the robot looked around the environment, faced the participant, and checked behind furniture. During navigation, the robots merely pass through the environment without deviating from their path from one end to the other. In social seating, both participants sat together on a living room couch. In isolated seating trials, one remained on the couch and the other moved to a chair in the corner of a small room. Seating positions are shown as colored X's in Figure 2(a)–(d).

4.2 Experiment Procedure

Figure 2(e) illustrates the layout of sessions in this study. First, participants were presented with a consent form. Upon receiving signed consent, researchers applied the sensors to participants. Next, they were seated on the living room couch to fill out a pre-experiment questionnaire. Following the questionnaire, participants watched a 5-minute nature video to establish a relaxed physiological baseline. Their seating locations were updated after the video to isolated if necessary (chosen at random). Next, experimental trials begin, each characterized by (1) a seating position, (2) robots separate or together, and (3) search or navigation behavior. Between trials, participants took a short survey. After four trials, participant seating was switched. The four robot behaviors were randomized and repeated with updated participant seating locations. The blockwise randomization of trial configurations ensured that the effect of confounding factors such as time was removed. The repetition of the same behaviors in single and multi-robot settings under both seating configurations enhances the repeatability of results. After completion of the experimental encounters, participants watched the same 5-minute nature video. This establishes a second relaxed baseline state, confirming that perturbations in the physiological signals during trials were induced by the robot and not a confounding factor such as elapsed time or fatigue. After the nature video, participants removed the sensors and were debriefed. Finally, participants engaged in an interview with one or two research team members.

4.3 Synchronized Data Acquisition

To effectively estimate human physiological response to the robots and varied behaviors, we collected synchronized data from multiple sensors during experiments. First, we used a wireless chest e-tattoo device to record ECG and **Seismocardiography (SCG)** [Bhattacharya et al., 2023], which is minimally intrusive due to its ultrathin and lightweight form factor while accurately recording the signal. For EDA, we used an Empatica E4 wristwatch [Empatica, 2024], which, according to a recent review on wearables used to assess perceived stress [Klimek et al., 2023], in 56.3% of such studies. The device also uses **Photoplethysmography (PPG)** to measure **Blood Volume Pulse (BVP)**. Other data sources include pose information from both robots and audio intensity from microphones placed at participant seating locations. Robot pose includes SE(2) position and orientation, represented as $(x, y, \theta)_t^{Spot}$ and $(x, y, \theta)_t^{Go1}$ recorded at time t. Given these raw data, other time series can be found: (1) Distance from robot(s) to participants, and (2) whether the robot(s) are moving.

Figure 3 illustrates the architecture for real-time synchronized data acquisition. Robot data are sent as timestamped ROS [*Robot Operating System*, 2007] messages via Robofleet [Sikand et al., 2021] to a base station, which is streamed to the local network with timestamps to **LabStreamingLayer (LSL)** [*LabStreamingLayer*, 2012]. Laptops placed in the living room near the couch and the small room near the chair stream data from omnidirectional microphones. The E4 wristwatch and ECG e-tattoo connect via Bluetooth to Windows machines before streaming via LSL. Synchronized data streams are recorded with LabRecorder [2012] into **Extensible Data Format (XDF)** format.

4.4 Questionnaires

The pre-experiment questionnaire collected age, gender, university affiliation, baseline stress information, and personality traits because prior research shows that these factors can affect the acceptability of robotic devices [Rossi et al., 2020]. The personality section of the questionnaire is the 10-Item Personality Inventory, a widely used measure of the Big-Five personality dimensions [Gosling et al., 2003]. The between-trial questionnaire contained 10 questions, each on a 7-point Likert scale. The questions are devised from existing surveys that individually focus on stress, comfort, and perceived safety [Akalin et al., 2022]. Questionnaires can be found in Appendix D and were administered using Qualtrics. Participants responded using their smart phones after scanning a QR code linked to the questionnaire.

4.5 Semi-Structured Interview Protocol

The purpose of two-person interview discussions after the experimental portion of the study was to uncover participants' perceptions, feelings, and narratives about their encounters. Two-person

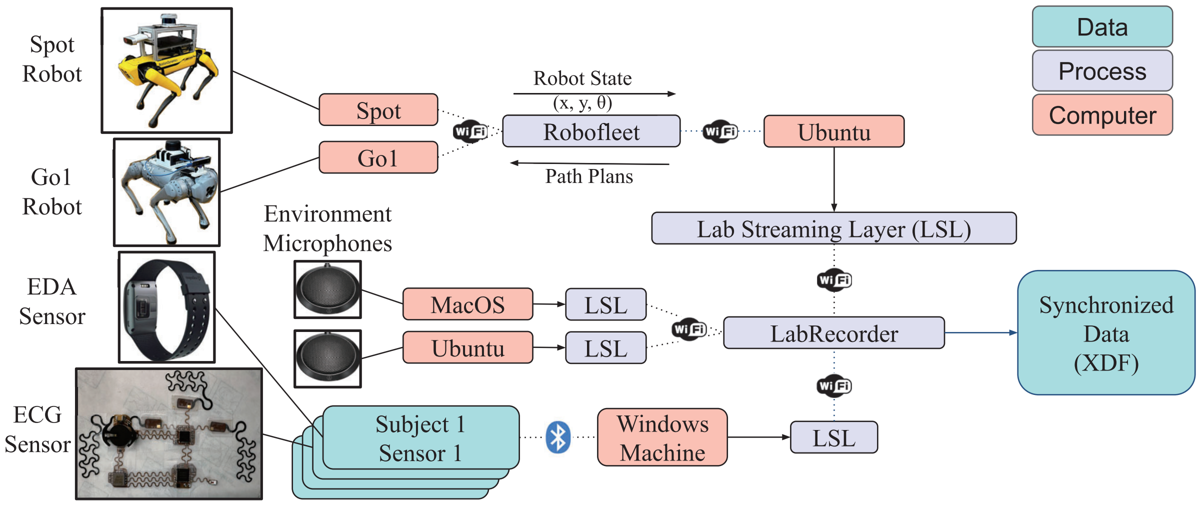

Fig. 3. An overview of the synchronized data acquisition architecture used in this study. The Spot and Go1 robots sent pose information via Robofleet [Sikand et al., 2021] to an Intel NUC base station. Two omnidirectional microphones were connected to laptops. Finally, the E4 wristwatch and ECG e-tattoo from both participants were connected via Bluetooth to designated Android and Windows machines. LabStreamingLayer [2012] was used to push timestamped sensor data from all different sources to the network. LabRecorder [2012] recorded all streams into a single synchronized XDF file.

interviews present benefits compared with larger focus groups by offering fewer opportunities for off-topic conversations and increased participant comfort and honesty [Lobe and Morgan, 2021]. The interview protocol was designed as a semi-structured, flexible set of questions promoting discussion related specifically to participants' experiences. The protocol featured four different sections: questions about participants' affective states, social and isolated seating arrangements, general opinions and experiences with robots, and personality questions. Questions included: "Were there any moments during today's experiment that stood out to you, or not?" or "During the experiment, the robot got closer to you. How did you feel about that?" Other questions were more closed including: "Did anything happen that made you feel nervous or stressed, or not so much?" or "Was there anything that made you feel particularly comfortable or happy?" Questions about general experiences included: "Where have you been exposed to robots before today?" and "What is your overall opinions on robots?" The full protocol can be found in Appendix D.

5 Data Analysis Methods

5.1 Questionnaire Analysis

Between-trial questionnaire responses were analyzed using the Mann–Whitney–Wilcoxon two-sided test with Bonferroni correction for multiple comparisons. Answers to questions which reflected discomfort were flipped in sign before analysis. No statistically significant correlation was reported between the control conditions from the questionnaire responses. The authors speculate as to why no statistically significant correlations were observed further in 7 and posit the relevant consequences for transdisciplinary study of human–robot encounters.

5.2 Physiological Decoding

This study concentrated on mental stress, particularly acute stress from short recordings of EDA and ECG (HRV). Although the e-tattoo device records both ECG and SCG, only the former was used for analysis in this study. Additionally, the wristwatch collects BVP, which was not used in this study due to the fact that ECG more directly measures heart activity. The trial recordings in this study are characterized as ultra-short term due to the fact that each trial lasts substantially less than 5 minutes. Because the goal of this analysis is to analyze the difference in stress response by participants caused by the independent variables tested (i.e., Baseline-Robot, Search-Navigation, Social-Isolated, Separate-Together), the analysis considers features evaluated over each individual recording in its entirety.

5.2.1 Signal Processing and Feature Extraction. ECG signals were first validated by visually examining the waveform of the average R peak grand average for a clear QRS complex and then by ensuring that the heart rate was reasonable. HRV features were extracted from peak information using Neurokit2 [Makowski et al., 2021], which returns a total of 91 features from both time and frequency domains and non-linear features. From these, 26 features cannot be reliably evaluated from ultra-short-term recordings and were removed. This consisted of HRV bandpower in the ULF and VLF bands, detrended fluctuation analysis-related indices, and all time domain features requiring recordings longer than 5 minutes.

We found that EDA signals had a level of noise mixed into them in some sessions, possibly due to poor skin–electrode contact during subject motion and the low sampling rate of the device. To mitigate this, a 2nd order Savitzky–Golay filter was used to act as a low-pass filter and smooth the noise before the cvxEDA algorithm is applied [Greco et al., 2015]. This algorithm decomposed the 1D EDA signal into phasic and tonic components via convex optimization. EDA features were calculated according to the methodology in Greco et al. [2015]. We computed additional features including EDA autocorrelation [Halem et al., 2020], normalized EDA sympathetic index, and median and maximum amplitudes of the phasic **Sudomotor Nerve Activity (SMNA)** peaks for a total of 15 EDA features. Features were normalized by the recording length wherever necessary. For either HRV or EDA, if the signal acquired in a given trial was deemed too noisy or incomplete (due to issues such as device disconnection), the signal was removed from the analysis. EDA and HRV features were then Z-scored and combined to result in a total feature matrix of 17 subjects $\times$ 10 trials $\times$ 80 features.

5.2.2 Exploratory Data Analysis. Next, the Chi-square test was used to examine the 20 most statistically significant features found in the Baseline-Robot condition. We used the Chi-square test over other feature transformative selection techniques such as principal component analysis as we prioritize retaining original physiological features for improved model interpretability. The Chi-square test feature selection was implemented through the fscchi2 function on Matlab which utilizes a ranking of p-values from individual predictor-response Chi-square tests. We note that EDA and HRV features were shortlisted in an approximately even proportion (9 EDA, 11 HRV) by the algorithm despite more candidate HRV features evaluated. This suggested that the multimodal approach, combining cardiac and EDA, successfully estimated robot-induced stress. Since exposure to robots is not a well-established stressor in contrast to established psychological or neuroscientific tasks such as Stroop and Speech tests, an important step was to: (1) further examine the shortlisted features based on whether they have reasonable physiological background for being predictive of acute stress and arousal, and (2) examine the evolution of these features in the encounter conditions of interest. These steps established that the predictive performance of the automatic human–robot encounter classifier in Section 5.2.3 was not a result of overfitting.

We plotted physiological changes induced by robot encounters over each of the robot encounter trials and baseline periods before and after experiments in Figure 4. These physiological changes were represented by two principal features: HRV composite multiscale entropy (HRV_{CMSEn}) and the number of EDA phasic SMNA peaks normalized by length of the recording ($nsEDRfreq$). HRV_{CMSEn} was calculated from the chest e-tattoo's ECG signals using the methodology in Wu et al. [2013], and $nsEDRfreq$ was calculated from the E4's EDA signals using the methodology in Greco et al. [2021].

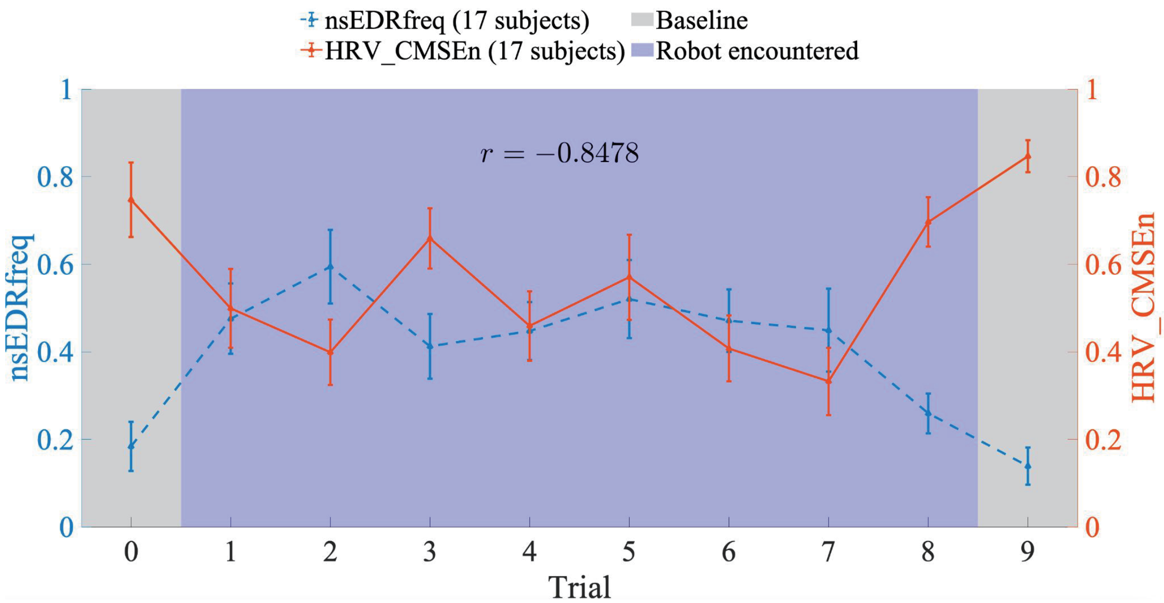

Fig. 4. Normalized features HRV_{CMSEn} and $nsEDRfreq$ over trials, i.e., experiment time. Plot shows averaged data from 17 subjects and their standard error. The trends show negative correlation as expected ($r = -0.8478$, see also Section 5.2.2) and show a clear U-shape as values representing the baseline relaxed state are perturbed by the introduction of robots into the environment in trials 1 through 8 before returning to normal (see further analysis in Figure 5).

HRV_{CMSEn} and $nsEDRfreq$ showed high negative correlation ($r = -0.8478$) in their values over experiment time, where time was represented on a trial level basis. This validated the successful estimation of acute stress over time wherein multiple devices with different sensing modalities observed similar stress-indicative signal properties across the experiment. This result agrees with the literature that reports increases in EDA phasic activity (i.e., $nsEDRfreq$) and decreases in HRV entropy (i.e., HRV_{CMSEn}) are indicative of increased acute stress (see Section 2.3). For most subjects, both features had some baseline value established in trial 0 and returned to trial 9, as demonstrated by the average over all participants in Figure 4. Individual data in trials 1 through 8 (robot encounter trials) showed that different individuals reacted with different intensities and temporal profiles to repeated human–robot encounters.

5.2.3 Classification of Human–Robot Encounters. To confirm the insights from the exploratory analysis, we built a machine learning-based classifier of human–robot encounters using features from EDA and ECG signals. To ensure fair evaluation, we report **Leave-One-Subject-Out (LOSO)** cross-validated results. In each iteration of LOSO cross-validation, the model was trained on data from all subjects except one, and the prediction was evaluated on data from the excluded subject. This process was repeated the same number of times as the number of subjects to cover all possible train–test splits. LOSO cross-validation therefore ensured that the reported performance accounted for the classifier's ability to predict the type of human–robot encounter experienced by any potential unseen person wearing the same sensors. We decided to use random undersampling boosting (RusBoost) [Seiffert et al., 2008] for the binary classification and multi-class adaptive boosting (AdaBoost) [Hastie et al., 2009] for the multi-class classification. For the former, RusBoost was chosen for its known robustness to class imbalances resulting from our study design where subjects spent more time encountering robot(s) rather than not. This was not an issue for the multi-class classification, as robot encounters were further categorized into specific types. The maximum number of splits was fixed at 20.

5.3 Deductive and Inductive Coding

Two different types of analysis provided insights to data collected from interviews. First, a deductive content analysis was employed [Krippendorff, 1989; Spearing et al., 2022], applying taxonomy by Akalin et al. [2023] of human comfort and perceived safety during HRI. Deductive analysis asked how well the taxonomy describes the experience of participants during this study. Second, TA [Braun and Clarke, 2006, 2012; Clarke and Braun, 2017] inductively explored emergent themes within the data. Particularly, this analysis enables the determination of what patterns and themes the participants felt was important regarding their encounters with the robots. We followed suggestions from Nowell et al. [2017] to produce a trustworthy and reproducible TA. In Section 7, we investigate the relationship of the resulting models from these two analyses.

5.3.1 Interview Data. The full transcripts from interviews were analyzed, totaling 13 unique transcripts and 159 pages of single-spaced text. Conversations were transcribed via the university's library web site, through an AI-assisted process on Amazon Web Services. The accuracy rate of the transcripts averaged around 70% accuracy. Corrections were made to transcripts that substantially deviated from the original recording. Access to recordings and identifiable information from interviews remains restricted to IRB-approved researchers.

5.3.2 Quantitative Content Analysis. The deductive content analysis [Krippendorff, 1989; Spearing et al., 2022] examined the relative frequency of specific topics appearing within interview data. The quantitative results, presented as frequencies of code occurrence in data, provide insights into patterns and topics important to participants. Particularly, content analysis explored the raw data (transcripts) to determine how well they fit the existing model.

The taxonomy by Akalin et al. [2023] of perceived safety in HRI defines the following factors: (1) Context of Use, (2) Predictability and Transparency, (3) Comfort, (4) Experience and Familiarity, (5) Trust, and (6) Sense of Control. These six factors influencing perceived safety represent the codes in the codebook. The primary coder created a codebook from these factors which can be found in the open access dataset [Gupta et al., 2024].

The codebook includes the definition for each code as given from the taxonomy, a description from the coders understanding of the definitions, inclusion and exclusion criteria, and relevant examples from the data. For example, in the codebook, Context of Use is defined as "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves; combination of users, goals and tasks, resources, and environment." Context of Use is described as "How environment, actors, action, or robot might paint a picture of a particular situation/outcome/use," while inclusion criteria are listed as "Person describing the lab in relationship to their feelings about the robot; What the future use of robots might look like; Discussing sitting together versus separate."

The first step of the analysis required familiarization with the data, noting emergent theoretical ideas and similarities between independent interviews. The primary coder applied definitions of deductive codes to the data by highlighting relevant phrases and organizing them. Codes were cross referenced by the definitions and changes were made to any codes as appropriate. Following the first round of content analysis, the primary coder trained a secondary coder in deductive content analysis, discussed code definitions, and negotiated coding rules.

To reduce bias and increase replicability of the coding schema, each coder independently performed a content analysis on three interviews. Code definitions were discussed to form an agreed set of coding criteria. Both researchers independently coded three additional interviews and computed **Inter-Coder Reliability (ICR)** using Cohen's Kappa [Cohen, 1960], a measure of reliability that takes random chance of similarities into account. Results showed an 84% chance of similarity between the two coders ($\kappa = .84$), which can be categorized as near perfect agreement in this model, measured as $\kappa > .81$. After ICR confirmation, the primary coder re-coded the full dataset using the finalized schema.

5.3.3 Qualitative TA. The researchers used TA, taking an inductive approach to the interview data, in addition to the deductive content analysis. Inductive approaches to data analysis enable researchers to study transcripts through the systematic identification of insights relevant to the research problem at hand, but not directly identified during initial study design.

Because TA excels when the researchers' goals include capturing experiences, feelings, and meaning in-context and is [Braun and Clarke, 2006; Clarke and Braun, 2017] also useful when developing theory based on collected data [Riger and Sigurvinsdottir, 2016], the researchers chose to utilize the method in addition to content analysis. Furthermore, TA is a cross-disciplinary method used for research in a wide range of fields of study including psychology [Selvam and Collicutt, 2012], medical research [Roberts et al., 2019], tourism studies [Heldt Cassel and Maureira, 2017], and HRI [Axelsson et al., 2024], making it a reliable choice for qualitative data analysis. This study primarily followed Braun and Clarke's [Braun and Clarke, 2006, 2012; Clarke and Braun, 2017] approach to qualitative TA, common in social psychological research. Care was taken to ensure replicability, therefore a codebook was produced for all analytical activities.

The first step of TA involved a complete read of the interview transcripts. Open coding began during the second read of the data, which involved highlighting sections of data and assigning a code—a relevant word or phrase, for example, a short sentence or reflection about a snippet of text. When codes appeared in multiple instances, the researcher noted potential connections between codes, which often suggest the presence of a higher-order category. After initial code generation, the entire list of codes were examined and then sorted into potential categories by collapsing them into singular more relevant words or phrases. Each category was also associated with relevant in vivo quotes generated by participants and placed in a spreadsheet in the open access dataset. The similarities across categories became clear as they took shape. The next step required experimentation with potential themes. All themes consisted of full sentences, characterized by the appropriate phenomena illustrated in the data, and served an explanatory purpose. To generate the resulting themes, the researcher returned to the data to further investigate participant quotes, sometimes sorting categorized quotations into other more appropriate themes or collapsing initial themes altogether. The final themes are reported in Section 6.3.2.

6 Results

This section begins with results from physiological signal decoding. Particularly, we investigate the overall stress response, represented by physiological arousal, of participants over individual conditions. Next, we inspect participants' descriptions of the sensors to ensure that these sensors did not induce stress or discomfort. We then present findings from content and thematic analyses of interview transcripts. The section ends by triangulating the findings from these distinct methodologies to give an overall perspective of participants' experiences during indoor human–robot encounters. Detailed information about the open source dataset from this study can be found in Section 7.1.

6.1 Physiological Signal Decoding

For data from N = 17 subjects combined, the Mann–Whitney–Wilcoxon two-sided test is used with Bonferroni correction for multiple comparisons. We chose to conduct a non-parametric Mann–Whitney–Wilcoxon test to compare the features between experimental conditions due to

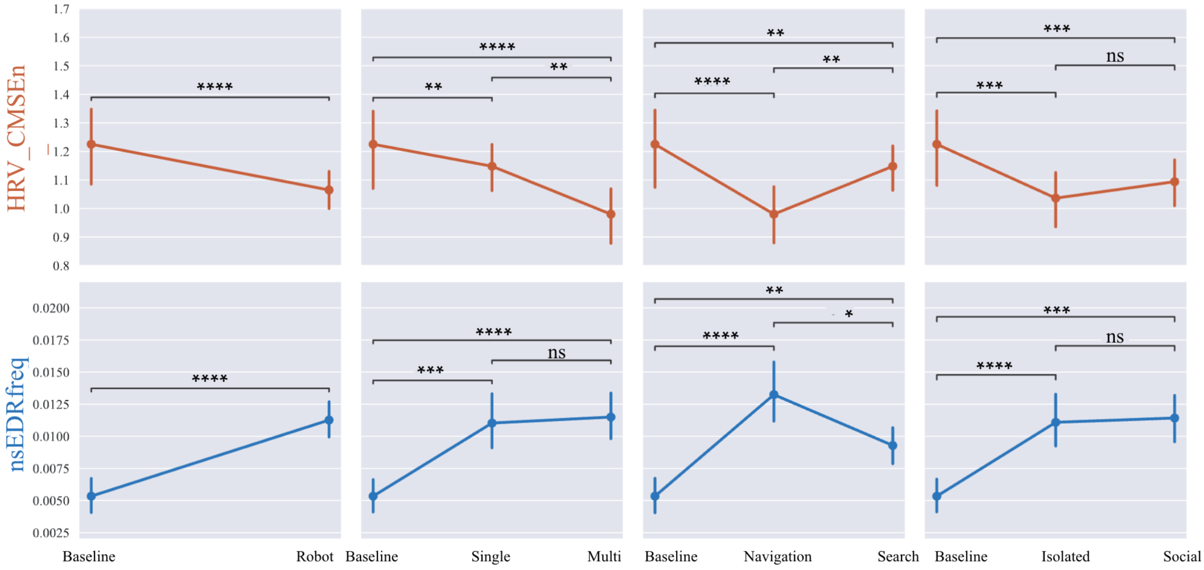

Fig. 5. Statistical comparison for the principal features HRV_{CMSEn} and $nsEDRfreq$ in each possible experimental condition, with all subject data. Error bars show the 95% confidence interval. Asterisks indicate statistical significance after two-sided Mann–Whitney–Wilcoxon tests with Bonferroni correction for multiple comparisons—ns: 0.05 < p; *: 0.01 < p ≤ 0.05; **: 0.001< p ≤ 0.01; ***: 0.0001< p ≤ 0.001; ****: p ≤ 0.0001.

the non-normality of the features HRV_{CMSEn} and $nsEDRfreq$ across all subjects (Shapiro–Wilk test, p-value < 0.001). For the full results of the Mann–Whitney–Wilcoxon tests including the adjusted p-values after multiple comparisons correction wherever applicable and the U statistics, please see Appendix C. Although the study included 24 total participants, the EDA data from the E4 wristwatch for 7 participants included severe noise. Therefore, data from these 7 participants are excluded from the analysis. We plotted average feature changes over all participants according to the encounter conditions and conducted pairwise comparisons (Figure 5) to investigate if the controlled encounter conditions induced physiological changes compared to baseline across all subjects. On average, we found that subjects were: more nervous whenever the robots were in the shared environment compared to when they were not (Baseline-Robot, Figure 5(a)); more nervous when the robots were traveling together rather than alone (Single-Multi Robot, Figure 5(b)); and more nervous when the robots were navigating compared to searching (Search-Navigation, Figure 5(c)). Whether the subjects were located alone or together did not induce significant changes in the stress features examined (Social-Isolated, Figure 5(d)). Note that reduced HRV_{CMSEn} and increased $nsEDRfreq$ reflect an increase in physiological response.

To control for inter-subject variability, repeated-measures ANOVA was used to investigate the effect of experiment conditions on the trial-level EDA response ($nsEDRfreq$). Table 1 suggests that EDA response within subjects was significantly altered across the 10 trials. Specifically, individual subjects had their EDA response change significantly depending on the presence of a robot (Baseline-Robot) and robot behavior. The predictive value of the HRV and EDA features was validated by classifying human–robot encounters via machine learning according to Section 5.2.3. The LOSO-validated confusion matrices are presented in Tables 2–5.

Figure 5 and Tables 2–5 suggest that for the average participant, sharing indoor spaces with mobile quadrupedal robots was associated with statistically significant physiological changes indicative of acute stress. Further, it was possible to distinguish, with higher than chance-level accuracy, specific characteristics (single vs. multi robot and navigation vs. search) about human–robot encounters

Table 1. Results of the Repeated-Measures ANOVA for the Number of Phasic EDA Responses during Each Encounter Condition

	F value	p-value
Trials ($df = 9$)	2.763	0.00588**
Baseline-Robot ($df = 1$)	13.477	0.000365***
Participant Seating ($df = 1$)	0.463	0.498
Robot Behavior ($df = 1$)	5.986	0.0159*
Single–Multi-Robot ($df = 1$)	2.215	0.139

***p-value ≤ 0.001; **p-value ≤ 0.01; *p-value ≤ 0.05.

Table 2. Accuracy of the Acute Stress Classifier of with Respect to Robot Encounters vs. Baseline

		Predicted	
		Baseline	Robot
Actual	Baseline	79.4%	20.6%
	Robot	13.2%	86.8%

The blue shading highlights the values when the model correctly predicts the robot condition based on the physiological response of participants.

Table 3. Accuracy of the Acute Stress Classifier of with Respect to Single–Multi-Robot Encounters

		Predicted		
		Baseline	Single	Multi
Actual	Baseline	70.6%	20.6%	8.8%
	Single	16.2%	54.4%	29.4%
	Multi	0.0%	30.9%	69.1%

The blue shading highlights the values when the model correctly predicts the robot condition based on the physiological response of participants.

based on the ECG and EDA signals. These observations should be critically examined for potential insights for designers of robots and HRI. As hinted by the statistical significance of differences in principal features between the encounter conditions (Figure 5), it appears that the binary and 3-class classifiers used to identify robot encounter conditions were the most effective. The classifier for the Isolated-Social condition was the least effective, not reaching statistical significance. Figure 4 and Table 1 suggest that these changes were consistent when examined on a per-participant basis.

Table 4. Accuracy of the Acute Stress Classifier of with Respect to Search-Navigation Behavior

Actual \ Predicted	Baseline	Navigation	Search
Baseline	67.6%	14.7%	17.6%
Navigation	2.9%	55.9%	41.2%
Search	11.8%	30.9%	57.4%

The blue shading highlights the values when the model correctly predicts the robot condition based on the physiological response of participants.

Table 5. Accuracy of the Acute Stress Classifier of with Respect to Encounters in Isolated-Social Seating

Actual \ Predicted	Baseline	Isolated	Social
Baseline	56.7%	33.3%	10.0%
Isolated	26.8%	46.4%	26.8%
Social	7.8%	46.9%	45.3%

The blue shading highlights the values when the model correctly predicts the robot condition based on the physiological response of participants.

These promising findings with respect to the ability to classify the Search-Navigation and Single-Multi robot suggests the potential for longer-term tracking of individuals' perceived safety around mobile robots. Notably, Section 6.4 investigates these statistical correlations from the combined lens of the interview findings. The conclusion of that analysis finds that near-collisions between robots likely confounded the findings here regarding these two conditions (i.e., Search-Navigation and Single-Multi Robot). The significance of these findings are discussed there.

6.2 Physiological Sensor Comfort

An important step to validate that the statistical results found in the previous section is to ensure the sensors themselves did not induce a strong response from participants. Effort was taken during the study design to ensure that physiological sensors were as comfortable and non-disruptive as possible, such that they did not restrict mobility or reduce perceived safety. This step further validates that physiological responses are indeed caused by the robot encounters. To this end, interview discussions were inspected to determine whether the physiological sensors induced any level of stress or discomfort that may have impacted participant response to encounters.

Subject ID11 stated, "I think the sensor technology is very cool and it's really neat, that the sensors are so sophisticated and that it's not just a watch, which is great." ID1 disclosed, "I think they were OK. I mean [...] they asked us if it was too uncomfortable or not, no, it was fine. [...]

Table 6. Frequency of Factors of Perceived Safety in Interview Data from Deductive Content Analysis

Code	**Number of Interviewees**	**Relative Freq. of Interviewees (%)**	**Number of References**	**Relative Freq. of References (%)**
Context of Use	24	100	104	25
Predictability and Transparency	23	95.83	91	22
Comfort	23	95.83	85	20
Experience and Familiarity	22	91.67	64	15
Trust	21	87.50	46	11
Sense of Control	18	75	33	8

after some time, I feel like you don't even remember that you have that you're wearing that. So, yeah, it was good. Not uncomfortable for me." Participant ID4 shared, "I forgot they were there mostly unless I had an itch, but I didn't really think of them too much. I mostly feel comfortable." Another participant enjoyed the concept of the sensors, ID5 stated, "the fact that you can sort of measure people's feelings about things based on other data that is a little bit less, sort of subjective with your choices, I think that's really interesting. I like the idea of, being able to, either control or, inform robot decision making with people's bio data that's coming out. That sort of stuff."

Transcripts lack statements that sensors were uncomfortable, itchy, tight, or irritating. In particular, findings suggest that the sensors have negligible impact and do not induce unwanted stress or distract from experiments. Therefore, we conclude that these non-invasive sensors are an effective measurement tool for human–robot encounters.

6.3 Content and TA

6.3.1 Content Analysis. Deductive codes were the factors by Akalin et al. [2023] of perceived safety, namely (1) Context of Use, (2) Predictability and Transparency, (3) Comfort, (4) Experience and Familiarity, (5) Trust, and (6) Sense of Control. Table 6 presents the results of the final deductive coding as a series of frequencies, percentage of relative frequency per interviewee (study participant), and relative frequency of all codes. Participants (N = 24) mentioned the importance of Context of Use 104 times, representing 25% of all codes. This agrees with Raue et al. [2019], which suggests the specific tasks a robot performed altered participants' perception of context and their individual risk perception. Many participants envisioned hypothetical contexts and scenarios, discussing their comfort around robots in diverse settings. According to these findings, context of interactions between humans and robots plays a clear role in participants' notions of perceived safety.

Predictability and Transparency and Comfort appeared next most often in the data. While both of these factors appeared in 95% of interviews, Predictability and Transparency appeared slightly more frequently among all codes, representing 22% compared with Comfort's 20%. Predictability is the ability to understand future actions, while Transparency is defined as being able to understand current actions [Akalin et al., 2023]. Comfort is synonymous with perceived safety, composed of both positively and negatively valenced feelings, for example, the stress response as estimated in the previous subsection. Notably, participants in this study were not informed of the tasks the robots would perform until after completing the experimental portion of the study. Interview analyses yielded several instances of participants making predictions about various aspects of the robot encounters or guessing what the robots were doing, in order to assess safety within their current context.

The remaining three codes, Experience and Familiarity, Trust, and Sense of Control, composed of 15%, 11%, and 8% of all codes. Despite appearing less frequently within the data, the findings suggested a strong connection among these three features. Understanding how Experience and Familiarity relates to Trust, and subsequently to Sense of Control enhanced researcher's understanding of perceived safety. Akalin et al. [2023] define Trust as "a multidimensional latent variable that mediates the relationship between events in the past and the former agent's subsequent choice of relying on the latter in an uncertain environment." This definition alludes to former experiences of robots and individuals' tendency to rely on such experiences during a present encounter. However, this definition was ambiguous with respect to what "multidimensional" factors contributed. After the initial round of ICR testing, the researchers more concretely defined Trust as "situations in which participants bank on their experience, during these experiments or from previous interactions with robots, to gain a Sense of Control and ultimately increase their comfort around robots." This definition ties together Experience and Familiarity and Sense of Control as the dimensions of Trust. Both factors often appeared together in participant interviews. On the other hand, some participants predicted future robot behaviors in order to increase their Sense of Control during encounters, representing the joint appearance of Predictability and Transparency and Sense of Control. This suggested that Predictability and Transparency may have some relationship to Sense of Control, as a mediating or moderating variable. Participants also reported the impact of robot-related factors on the participants' Sense of Control over the interaction, for example, robot size and speed.

6.3.2 Thematic Analysis. Five primary themes emerged as a result of the TA. In particular, the final themes include: (1) sense-making, (2) familiarity of "robot dogs," (3) human impact on perception, (4) excitement and novelty, and (5) future possibilities of robotics.

Sense-Making. Participants discussed their thoughts and interpretations of the robot encounters during interviews primarily motivated by uncertainty and sense-making. They related their experience to describe the robot encounters using senses, primarily referencing sight and hearing, with some mention of touch involved. The structure of information processing reflected participants uncertainty about their relationship to the robot, attempts to *assess* their safety, and their determinations about safety.

Specific instances that startled participants included the moment robots appeared in the first trial and the potential for robot collisions between robots or with obstacles in the environment. The introduction of one or both robots at the onset created a sense of fear and discomfort, as one participant compared the robot to a "jump scare" (ID8). They elaborated: "we were just watching TV, and then I heard these big loud noises and I was kind of scared," which suggests that merely hearing a nearby robot can be cause for alarm, even without visual contact. Participants also described their perceptions of malfunction by the robots. General notions of equivocality characterized feelings of some participants who logically understood that—based on the experimental setting—they would be safe from harm of the robots; however, they were not able to interpret exactly what the robot was doing based on just watching and listening to them.

Another source of uncertainty involved whether participants could see the robots. Participants in the isolated seating described uncertainty about when the robot may appear, although they anticipated it eventually would come around the corner. ID4 used prediction to reduce uncertainty and gain a Sense of Control, despite their inability to see the robot: "I think as the trial went on whenever they went out of my eyesight or went behind the wall. I started anticipating. Oh, what is it time for them to come out in front of me?" Particularly, participants who began the study in the isolation of the small room, who had no prior visual contact with either robot, described their predictions of the robots' appearance. They considered how the robot may navigate the space and whether or not the robots had wheels or legs. Surprisingly, some participants did not attribute the sound of the robots to a legged design.

Participants described a lack of understanding of the robot's abilities and considered how this impacted their own perceptions of safety. Some questioned the capabilities of the robot and whether it had the capacity to do harm. Participant ID22 compared robots to their mental model of overall advancements in AI, "I don't know the capabilities of all of the things that they can do. Does it have the power to hurt me? Maybe, I don't know." Others made sense of the robots over time, particularly when they realized that "well, it's not doing anything to me" (ID18). *Increased exposure* to the robots eased their feelings of anxiety about robots, and in some cases, led to increased enjoyment. ID22 also noted a more developed sense of comfort and was "able to figure out what is was they were supposed to be doing," and even "loved watching them turn corners."

In a few instances, participants facilitated their own information gathering through experimentation, discovered as a category of sense-making. ID5 described pushing their foot toward the robot to test its response. Another subject disclosed, "I [reached] out to pet one once. I'm not sure if I was allowed to do that, but I did anyway" (ID14). Interestingly, they acknowledged the robot was a "metal object," yet referred to the haptic interaction as a "pet." This calls attention to the action of "petting" as both a method of gathering information related to the safety of the robot itself, but also an action based on the inferral of safety in the given situation. This symbolizes a connection between the participants' mental models of similar creatures in comparison to the quadrupeds, robot characteristics that facilitate interactive and safe behavior, and the resulting impact of these connections on perceived safety.

TA also identified the familiarity of "robot dogs" as a factor influencing participants sense-making. That is, they found that robot dogs were familiar and *disarming*. This sub-theme of sense-making contributed to the reactions of some participants. The dog-like appearance of the two robots appealed to those who favor animals. This similarity was likely drawn as a comparison due to the Spot's quadrupedal design and is supported by participants trying to "pet" the robot (ID2 and ID14).

Sense-making is shown through these findings to play a key role in participants' perceptions and feelings of safety. They mediated their uncertainty by assessing their environment through experimentation. Eventually, they made determinations about their safety through these various information gathering activities. Additionally, some participants found the familiarity of the dog-like structure of the robots as a disarming factor, in some cases likening them to pets.

Humans Nearby Affect Perception. Participants reflected that fellow participants affected their perception of the robot encounter. ID4 described that "the isolation was more [...] I analyzed how I felt, seeing the robots or just knowing they were nearby and not knowing exactly what's going on, or what *they* know." Interestingly, they not only described increased awareness to their own feelings but also to the robot's knowledge, even when not in their line-of-sight. Reactions of fellow participants also impacted some: "if they were reacting in some way or, I don't know, visibly nervous, like shifting around or something else [...] that might change my reaction" (ID15). Similarly, ID21 watched their counterpart's reaction, noticing "he didn't really care about [the robots], so I was like, yeah it's fine." ID1 felt more comfortable when sitting together because "we could talk about what was going on." ID12 described their "feeling was more empathetic. I wonder what she's thinking and I wonder what her reactions are." Well summarized by ID13, "whenever you're in a private space, you find a different identity [...] I'm just noticing that." Thus, the data strongly suggest that having other participants nearby during the encounters increased people's feelings of safety, which may also explain the physiological signal findings between social and isolated seating. We note that this differs from the lack of statistical significance found from the

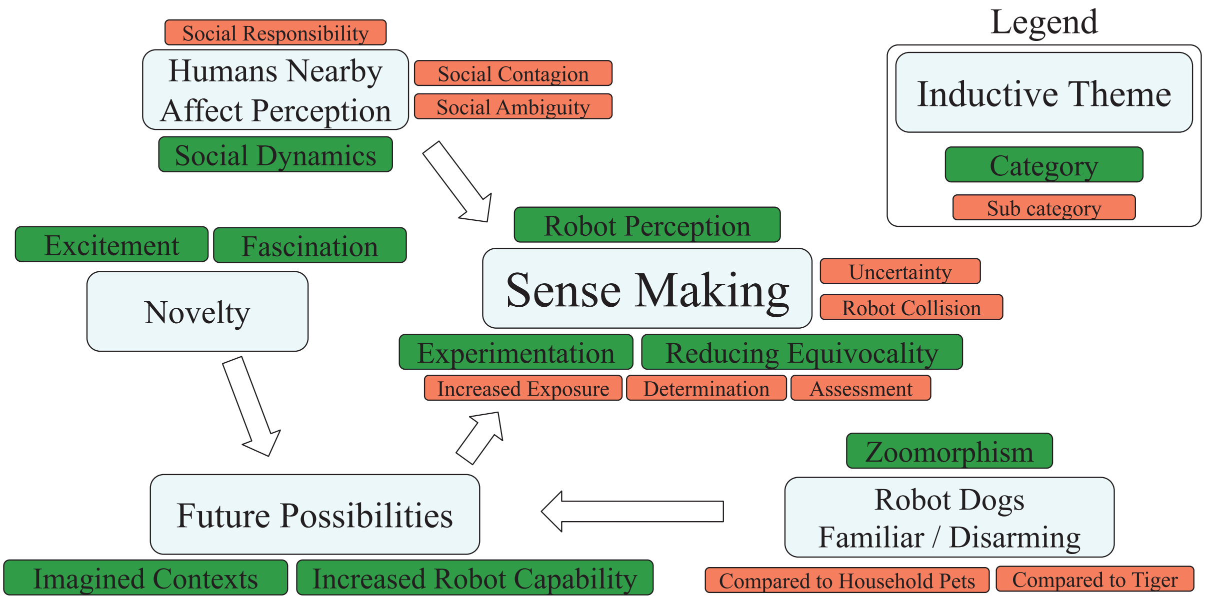

Fig. 6. A legend of themes, categories, and sub-categories is found in the upper right. A representation of the categories and sub-categories found related to inductive themes from TA of interviews.

physiological signal data with respect to the seating condition. The next section addresses and discusses this conflict.

Excitement and Novelty, Future Possibilities, and the Familiarity of Robot Dogs. Closely interrelated, TA identified Excitement and Novelty of robots and Future Possibilities of robotics as sub-themes of the core sense-making. In some cases, participants' feelings of excitement and imagination may have overwhelmed any sense of uncertainty or discomfort. For instance, ID22, who assigned the Spot the name "digi-dog" exclaimed, "I saw the digi-dog coming—I got so excited!" Experimental encounters, especially with participants doing so for the first time, transported them to a future in which robots are part of everyday life. Various sentiments included hopefulness, excitement, and openness, but also concern, curiosity, and uncertainty about the future of robotics. Increased robot capability included helping around the home and even homes themselves becoming automated and "somewhat sentient" (ID5). These themes played a positive role on participants perceptions of safety and comfort in the dataset.

6.3.3 Inspecting Relationships between Inductive Themes and Factors Influencing Perceived Safety. In Section 6.3.1 the results from participant interviews were inspected under the existing taxonomy of perceived safety. Section 6.3.2 instead approached the interview transcripts from an inductive standpoint and identified several themes or patterns in the data. In this section, we investigate the interdependence between these inductive themes, categories, and sub-categories with the deductive codes (i.e., the factors influencing perceived safety) in order to identify any relevant connections between the two analyses.

The directional relationships among inductive themes as well as categories, sub-categories, and deductive codes are illustrated in Figure 6. The inductive themes are the key results (patterns in the data) from the interview transcripts. Themes are shown as hierarchically related to categories and sub-categories in green and salmon, which were used during the open (primary) and secondary coding phases of TA. Themes resulted after combining and collapsing these categories, grouping them together with higher-order ideas. For example, participants described imagined contexts in which they may interact with robots, and also imagined scenarios in which robots may have an increased capacity to perform various actions. Because these instances were discussed in relationship to the future, the theme “Future Possibilities” then linked those categories together. The rest of this section makes comparisons between the perceived safety taxonomy and the inductive themes identified during TA, further contextualizing the potential usefulness of the model.

Context of Use was shown in the previous section to be closely related to participants’ ability to make sense of the robots. ID8 reflected on their increased Comfort during experiments as their Experience and Familiarity increased, resulting in a level of trust in the robots: “I just got less anxious and more used to it cause I was like, well, it’s not doing anything to me.” Participants also made predictions about the robot’s knowledge: “it’s aware of me and I’m aware of it. So it’s kind of a mutual perception” (ID9), while ID1 could not make sense of the robots “I don’t know what’s, what these guys are doing.” Thus, Comfort and Experience and Familiarity were also closely tied to the familiarity of robot dogs category. Trust and Sense of Control also played a role for some in their relationship to this idea that the robots are familiar to dogs or other living beings. ID10 rationalized how they could categorize this technology, considering their trust that these systems are not harmful, stating that “they’re not necessarily predators, they’re not servitude, they’re not equal to those necessarily. They’re just another category of autonomous beings.”

Participants’ Sense of Control also played a role in identifying encounters: “I’m sitting there, I see it. There it is. I’m interacting well, I mean, I really didn’t interact with it. I sat there and it came up to me, I didn’t touch it. I didn’t press any buttons. I didn’t operate it in any kind of way. I didn’t, you know, alter his behavior in any kind of way. So I wouldn’t really call it an interaction” (ID23).

Excitement and Novelty and Future Possibilities themes were related to several of the components. Context of Use was found to be important, with ID1 considering the possibility of robots “doing tasks around me. [...] I imagine what we have in the future, which is just [...], robots doing stuff.” ID5 considered “whole house intelligence [...] you walk in and it says, hello, how was your day? And, you know, blinds down thermostat up, et cetera.” Subject ID5 considered their comfort level and trust in smart homes. ID10 also reflected on their possible lack of Trust and Sense of Control with respect to robots in the future, “we will definitely be out paced and we are gonna get into a weird position in relation to robotics.” Some also made predictions of the robots: “They’re kind of amusing. They’re entertaining. They often look a little lost or a little bit, like they’re not fully sure what they’re doing.” (ID5) The theme regarding other humans nearby is shown to be related to Context of Use and Trust in Section 6.3.2. Participants described the impact of people and location on their feelings during encounters. Furthermore participants paid more attention to their trust in the robots without other people around.

Overall, our results agreed with prior research that Context of Use plays the most critical role in the perceived safety of participants. Contextual effects on perceived safety were shown to be mediated by Predictability and Transparency, Comfort, or Experience and Familiarity. In contrast to the user study presented alongside the taxonomy, Trust and Sense of Control appeared rather infrequently within the interview data. This suggests that factors affecting perceived safety during encounters differs from factors driving perceived safety during interactions. Thus, Figure 6 represents a potential starting point from which to design a taxonomy of factors influencing perceived safety during human robot encounters. Future iterations of the taxonomy may account for the difference between interactions and encounters or consider organizing elements of perceived safety hierarchically, perhaps by exploring which element may precede or proceed one another.

6.4 Mixed-Methods Findings

Sections 6.1 and 6.3 investigate the data from independent methodologies and disciplines with respect to the control conditions—i.e. Baseline-Robot, Single-Multi Robot, Navigation-Search Behavior, and Social-Isolated seating. Here, we inspect those findings from a combined perspective. We also identify and address conflicting results among the datasets and analyses.

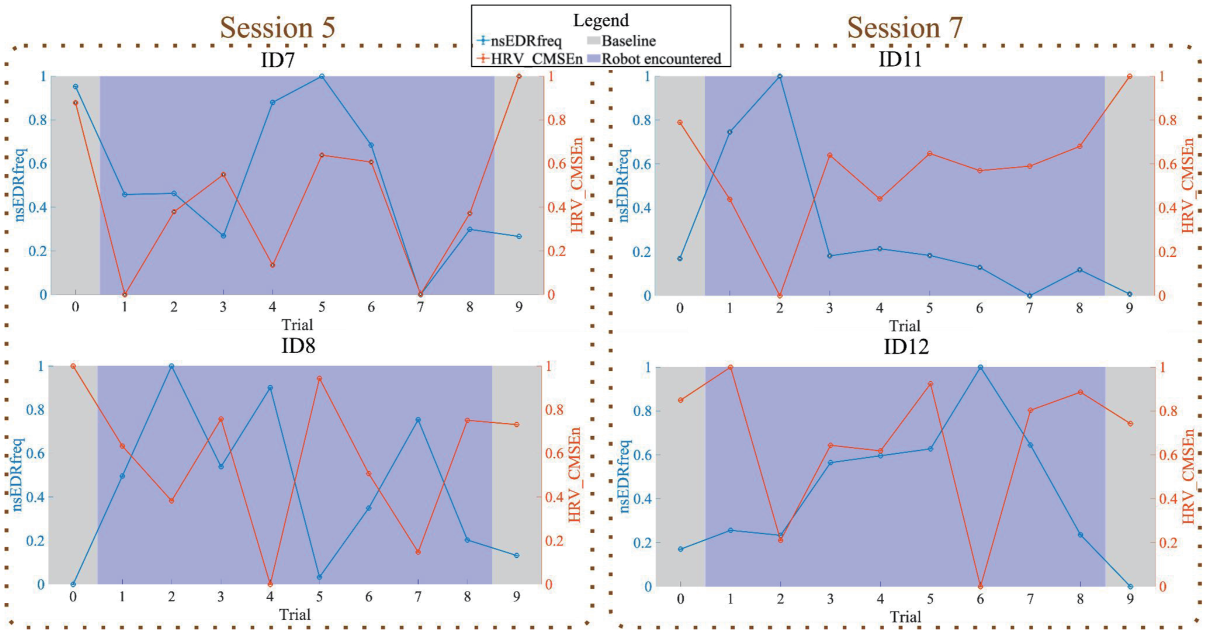

Fig. 7. Normalized features HRV_{CMSEn} and $nsEDRfreq$ over trials. Data show individuals from sessions 5 and 7. In particular, we see a strong arousal from participants in the together navigation robot behavior when participants are seated in the living room. Session 5 trial 4 results in high arousal from subjects ID7 and ID8. In session 7, both participants show a strong response during their first encounter with this behavior in trial 2 for subject ID11 and trial 6 for subject ID12.

6.4.1 Repeated Exposure. We begin this section by asking whether repeated exposure to the study conditions is justified. Specifically, we consider the physiological responses and the interview transcripts for instances that indicate repeated exposure to the robots made a clear impact on participants feelings and responses. First, we recall Figure 4, which plots average values (for N = 17 participants) for HRV_{CMSEn} and $nsEDRfreq$ over the trials of the experiment session. The relaxing baseline is shown in Trials 0 and 9 in the figure, with clear disturbance to both values during trials 1–8 (i.e., during the robot encounters). This plot shows no indication that participants' physiological response trended toward the relaxed baseline during any of the trials. A visual inspection of all individual participant plots (see Appendix B in Figure B1) of HRV_{CMSEn} and $nsEDRfreq$ over trials reveals that only a single participant, ID11, displays this trend.

The plot for ID11 is shown in Figure 7. This plot illustrates that participant ID11 experienced reduced stress during the final 3 trials. Inspection of the interview transcript involving ID11 reveals that they "got more used to it. I was like kind of able to figure out what it was they were supposed to be doing." Therefore, both datasets indicate that for this individual participant, repeated exposures likely impacted their feelings and perceptions. Next, we inspected the transcripts for mentions by participants that they felt more comfortable or more familiar with the experiments as trials continued. In total, besides ID11, we find only four other times that repeated exposure impacted participants' feelings and perceptions. Therefore, we believe that the combined datasets indicate that the repeated measures of the independent variables tested were justified.

6.4.2 Baseline–Robot Condition. Despite some participants' lackluster sentiments toward the encounter itself and generalized feelings of safety due to the lab environment, physiological sensor data and interview findings demonstrate that robot encounters have measurable effects on cohabitants. Subject ID6 stated, "it's more just that initial entrance of the robots." In other cases, participants documented stronger reactions, like ID19: "For me, if it got [...] very close to me, yeah, I feel kind of afraid [...]I think they can act more, more fast [...]faster than the human. And if they got close to me, I don't think I can, react actively to get to [safety]." These results agree with the strong statistical findings in the Baseline-Robot condition in Section 6.1.

6.4.3 Conflicting Results. When jointly considered, findings yield a conflict in the Social-Isolated seating condition. While analysis of EDA and HRV data identifies no statistically significant relationship based on seating condition, many participant interviews depict meaningful changes in experience between the two seating arrangements. Recall that a key theme identified during TA was that nearby humans affect participant perception of the robots. Participant ID18 explains, "sitting together makes me much more comfortable. And when I was alone in the small room space, then I felt more stressed when they are coming." ID21 describes, "Being in a backed up corner and having a robot that I wasn't anticipating coming at me [...] Not having anywhere to go and just being kind of in that close quarters that made me kind of uncomfortable [...] If I were on the couch by myself, I wouldn't be as stressed out cause I'd be like, well, this is an open space." Similarly, ID19 discusses discomfort with having robots behind them, while seated on the couch: "that makes me a little bit scared cause I can't see [...] where they are going and the sounds that they are approaching to me."

These results suggest an effect of social dynamics, room geometry, and/or seating conditions on perceived safety. These results strongly disagree, which presents little surprise given the disparity in data type and the shortcomings associated with single-discipline approaches. Particularly, this exemplifies that one method alone (i.e., signal analysis) misses a key finding while subjective, open form responses identify relevance. This finding further substantiates the need for research conducted from several perspectives and the use of disparate data types.

6.4.4 Near Collisions. While findings with respect to seating condition reveal clear conflict, next steps included identifying underlying findings within the combined data, with respect to the Navigation-Search behavior and Single-Multi robot conditions. These conditions demand deeper inspection. Specifically, the Search-Navigation condition identifies that participants showed higher acute stress during Navigation compared with Search. The authors were surprised at this finding, given the expectation that participants would present more anxiety during search instead of navigation, when the robot spent more time with and walked behind the participant(s). However, once asked about moments in which participants felt less safe, researchers observed that near collisions between robots occurred during navigation trials.

To investigate, we review individuals' physiological signal responses for possible outliers or trends. Specifically, near-collisions correspond to outliers identified in participants' physiological signal responses. The following discussion establishes that near-collisions meaningfully impact perceived safety, acting as a confounding factor within these two control conditions.

Figure 7 displays normalized features HRV_{CMSEn} and $nsEDRfreq$ found from individuals' HRV and EDA for subjects ID7 and ID8 from session 5 and for subjects ID11 and ID12 from session 7. Participants ID7 and ID8 show strong physiological responses during trial 4, characterized by the drop in HRV_{CMSEn} and rise in the $nsEDRfreq$. A similar response is seen from ID11 in trial 2 and ID12 in trial 6. During all three of these trials, the robots performed "together navigation" behavior, moving from the kitchen area toward the living room. Session 5 trial 4 involved the participants seated socially in the living room, as was the case in session 7 trial 6. In session 7 trial 2, only participant ID11 sits in the living room, while ID12 instead is isolated in the small room. Closer inspection of the together robot behavior in these three trials reveals the two robots nearly collided when the Spot passed the Go1 in the living room. These moments are shown in Figure 8.

The physiological sensor results demonstrate that all four of these subjects reacted strongly to witnessing a near collision between the two robots, despite ID11 describing the near collision as

Fig. 8. Moments taken during together navigation behaviors in sessions 5 and 7 where the two robots nearly collide. The Spot moves faster than the Go1 and passes the smaller robot in the living room in these scenarios. Participants seated in the living room individually show strong physiological responses during these trials (Figure 7).

"not scary." When cross-examining with interview responses, results reveal that ID11 recalled "[the big robot] kind of almost hit the little one, like, they were both in the same space at the same time and kind of had to stop and readjust, and the big one kicked." Participants report changes in their feelings with respect to the Navigation-Search condition. In particular, ID19 noted that "in some trial they are walking behind me. So even though we were on the couch [...] that makes me a little bit scared cause I can't see what, where they are going and the sounds that they are approaching to me." Their statement reflects the uncertainty they felt when the robots moved behind them, something that only occurred during search conditions. Furthermore, ID17 recognized the change between navigation and search behaviors, stating that "when the robot came in front of us and we went in a circle. So that was different than it had done."

Further investigation of interviews yielded several instances of participants noting near collisions or collisions. ID0 recalled, "I know that they collided or something, I couldn't see that because the couch was in between [...] I heard a noise like a sound that I was not expecting. I was used to them walking around by then, but that was like a crashing sound or something." During session 1 trial 8, the robots performed together search behavior and participants sat isolated. The Spot and Go1 nearly collided, causing the Go1 to divert into a wall. The concurrence among the varied data types indicates a level of anxiety associated with near collisions, which (1) suggests near collisions as a possible confounding factor among the remaining two conditions and (2) supports the efficacy of the physiological measures used to estimate human stress responses.

To further investigate near collisions as a confounding factor, we considered the synchronized XDF file playbacks using Python to search for occurrences of near collisions between the two robots in the living room, near the participants. Relevant trial candidates are those with together navigation and together search behaviors when the robots move from the kitchen toward the living room. No other collisions or near misses occurred outside of these conditions. Inspection of the robot animations made from synchronized XDF files reveals ID6 witnessed a collision during session 4 trial 4, where robots perform the together navigation behavior. During their interview, subject ID6 mentioned "what made me most anxious though was when they collided with each other, which I don't know if it was done on purpose by you guys or not. But it was like, oh, no. Is there gonna be a crash or are they gonna panic now?"

In total, we find collisions or near collisions occurred 12 times throughout the dataset and were witnessed by a total of 14 different participants. The data inspected in this subsection account for 9 of those participants. Unfortunately, the remaining occurrences involved participants whose physiological data were not included in the analysis due to the noisy EDA signal. Individual plots

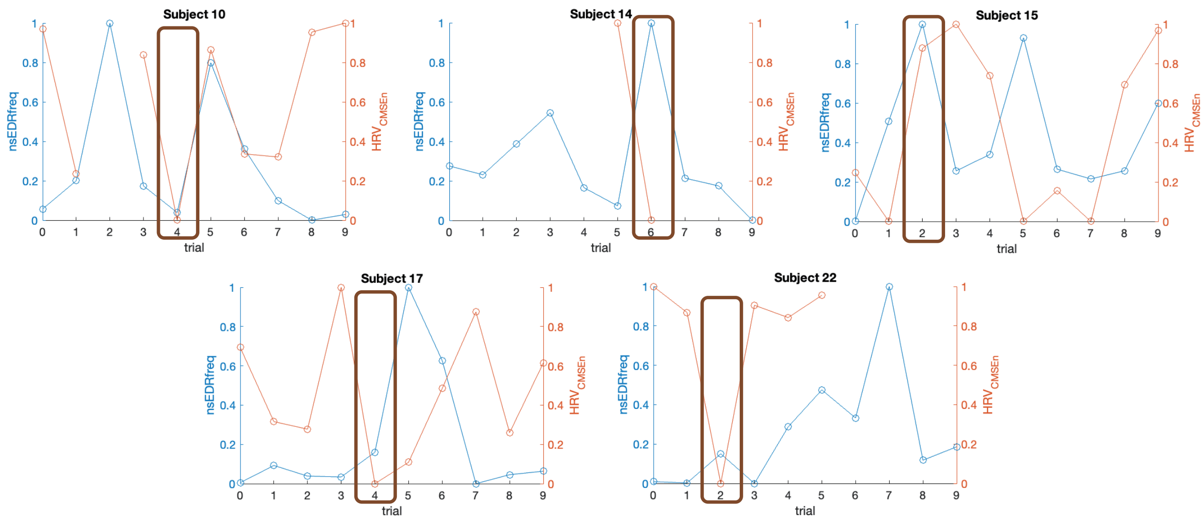

Fig. 9. Normalized features HRV_{CMSEn} and $nsEDRfreq$ over trials for labeled individual responses. Boxes label individual trials wherein the robots perform together navigation behavior and cross paths near the participant seating location. In these marked trials, we notice that overall, participant HRV_{CMSEn} tends to decrease and $nsEDRfreq$ tends to increase, representing a stronger physiological response.

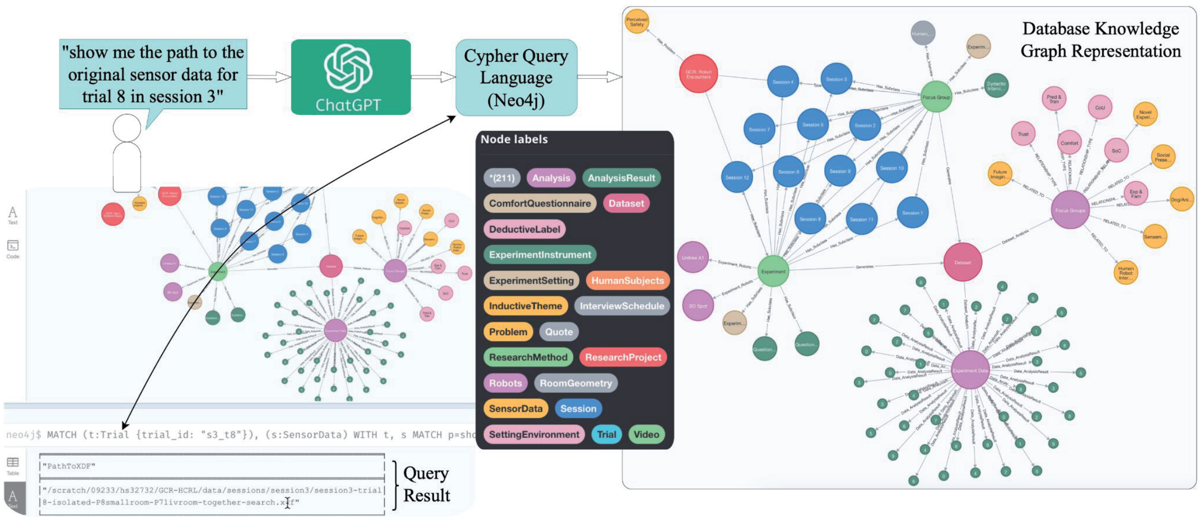

Fig. 10. A demonstration query against the Neo4j knowledge graph database. The legend in the middle marks node types for the graphic representation on the right. The blue circles, for example, represent the 12 total experiment sessions. Natural language queries are input by a user, which are experimentally translated to Cypher query language using ChatGPT. The translated queries are input into the Neo4j prompt, as shown on the lower left. In this example, the output from the database is the synchronized XDF file from session 3 trial 8.

of normalized features HRV_{CMSEn} and $nsEDRfreq$ for participants over several trials meeting this criteria are shown in Figure 9. The trials in question are marked with brown boxes, which are all together navigation behavior. Subjects ID10, ID14, ID17, and ID22 show strong nervous system response as characterized using HRV_{CMSEn}. Participant ID10 shared "I guess general sense[...]a little nervous[...] when one of them was on the trajectory but then the other one got in their way." Subjects ID14 and ID15 also show high galvanic skin response, characterized by $nsEDRfreq$ found from EDA data.

These data further suggest that near collisions and collisions during experiments act as confounding variables in the statistically significant correlation found between the Search-Navigation condition and Single–Multi-robot condition. The multidisciplinary data clearly illustrate that near collisions and collisions should be avoided in multi-robot encounter scenarios.

7 Discussion

In the previous section, we applied an Interdisciplinary mixed-methods approach, which analyzes, synthesizes, and harmonizes links between disciplines into a coordinated and coherent whole [Choi and Pak, 2006]. Transdisciplinarity, on the other hand, integrates the natural, social, and health sciences in a humanities context, and transcends their traditional boundaries [Choi and Pak, 2006]. This section aims to triangulate the results of the survey, physiological signal, and interview data in the interest of transcending disciplinary barriers. Human robot encounters will continue to proliferate in daily life in public, community and campus settings. This research aims to contribute knowledge about human perception and perceived safety to the institutions, corporations, engineers, and designers who implement technology in these places. With human robot encounters relatively understudied in many possible future applications, our approach looks to provide a complete picture of human response.

A key contribution is the demonstrated capability to classify human perceived safety via acute mental stress under varied robot behavior. Particularly, this includes situations when participants are equipped with the appropriate wearable sensing technology and wireless data acquisition hardware. We outlined an objective approach to recording human–robot encounters and investigated factors affecting perceived safety of mobile quadrupeds in indoor environments by analyzing human physiological markers of acute stress. Interviews with participants revealed strong agreement among the varied data types with respect to the Baseline-Robot condition, suggesting that generalized human robot encounters will generate a reaction in nearby community members. Some participants described their increased comfort as the experiments continued, due to their increased familiarity with the two robots considered. This suggests that for future technology integration, community members may benefit from merely increased exposure to such unfamiliar technology when conditions are deemed safe. Further, it may be beneficial to provide community members with the opportunity to engage with and learn about advanced technologies prior to their deployment.

In contrast, the data disagree regarding the Isolated-Social seating condition. While physiological signal analysis finds no statistically significant correlation, interview analysis reported an impact of social dynamics and contextual factors, like the shape and size of the room. This suggests that a more focused study of these variables (e.g., social dynamics; room geometry; seating visibility) may yield significant physiological changes among participants, perhaps using another sensing modality like EEG. Designers are recommended to critically evaluate the environmental factors before deploying robots in shared spaces. Analysis of EDA and HRV among Navigation-Search and Single–Multi-robot conditions showed statistically significant trends. However, the interdisciplinary approach used indicates that these findings were inconclusive, possibly confounded by real and near collisions. These findings necessitate the future study of the effects of movement behavior and the number of robots nearby. Community embedded robots are strongly recommended to place safeguards such that robots do not collide or nearly collide with each other or obstacles in the environment, as it is clearly shown to negatively impact perceived safety.

Another key contribution was the identification of human sense-making during robot encounters. The social seating arrangement provided an opportunity for participants to make sense of robot encounters socially, in agreement with a core component of Weick's conceptualization of sense-making [Weick et al., 2005]. This suggests that sense-making changes based on contextual factors related to individuals' environmental conditions and the amount of information available.

Whether or not changes in the emotional states of participants in response to the robots were the product of social contagion [Burt, 1987], mere coincidence, or sense-making, it appears participants' relationship to the robots is at least experientially affected by the presence of another person. This provides further evidence that privacy with robots and social dynamics impact perceived safety and require further inquiry in many types of interactional spaces. Participants' attempts to understand the robots involved touching, talking to, thinking about, or imagining what the robots could do in future contexts. Such information gathering activities are closely related to participants' attempts to confirm the safety of their current situation. Overall, participants' acute stress responses, triggered by the presence of a robot, likely resulted in attempts to make sense of the encounter via information gathering. This aligns with existing literature demonstrating that near the end of an interaction with a robot, curiosity incites information gathering behavior [Eshed et al., 2021].

Researchers in HRI suggest that organizational studies' models are useful to understand mental models of people near robots [Rueben et al., 2021]. A theory particularly adept at describing the effects of technology in relationship to context and structure is **Adaptive Structuration Theory (AST)**. AST posits that technology, social life, and organizational structures are mutually informed. Further, when individuals interact with "actionable technology," they also interact with various social or organizational structures via structural features of the technology. This proposition applies to human–robot encounters because, as our data suggest, individuals seek to understand both the capabilities and motives of the robot in tandem with the current setting. These factors also lead back to the primary theme identified in Figure 6 of sense-making, the central component identified during TA, and is closely related to factors influencing perceived safety. When lacking previous experience around robots, not only do people seek information through experimentation with their surroundings, but they also gain increasing trust for authoritative figures. Ultimately, the behavior of robots alters perception, resulting in changes to people's level of Comfort, Sense of Control, Trust in the situation, and Predictability of the robot. Future research may explore this notion using existing scales or theories that explain technology acceptance, such as the Godspeed Questionnaire [Bartneck et al., 2009] or Diffusion of Innovations Theory [Sahin, 2006].

7.1 Dataset

The result of the study is a multimodal dataset from the self-report questionnaires, physiological sensors, and interviews. It also includes data from the robots, static environment cameras, and environment microphones. This is a unique dataset, not only due to the theme that it addresses but also because of the different perspectives that it introduces. The datasets contain synchronized humans' physiological, robots' location/position, and experimental data linked by participant ID. These diverse datasets are then linked to themes deductively and inductively coded from the participants' responses about their experience and feeling related to mobile robots and safety. The dataset is published for open access in the Texas Data Repository and can be found at Gupta et al. [2024]. The dataset contains the following items:

(1) Results from pre-experiment and between-trial questionnaires.
(2) Copies of both questionnaires and the semi-structured interview protocol.
(3) Synchronized XDF files from every trial with raw data from EDA, ECG, PPG, and seismocardiography (SCG). These synchronized files also contain robot pose data and microphone data.
(4) Results from analysis of two important features found from HRV and EDA. Specifically, HRV_{CMSEn} and $nsEDRfreq$ are computed for each participant over each trial. These results also include Robot Confidence, which is a classification score representing the confidence

that the 80 physiological features considered originate from a subject in a robot encounter. The higher the score, the higher the confidence.

(5) Anonymized videos captured using two static cameras placed in the environment. They are located in the living room and small room, respectively.
(6) Animations visualized from XDF files that show participant location, robot behaviors, and additional characteristics like relative audio volume.
(7) Quotes associated with inductive themes taken from interview data. These quotes demonstrate and justify the results of the TA. Raw text from interviews is *not* included to protect participant concerns. All of the quotes associated participant ID numbers so they can be analyzed in relation to the sensor information.
(8) Quantitative results from interviews demonstrate the findings from deductive content analysis. The deductive codebook is also included.

Existing multimodal datasets designed to investigate modes of human–robot engagement in different contexts embed indicators of human emotions around robots. Data are collected using diverse sensing technologies which may involve physiological sensors and participant surveys and various data analysis methods such as observations or researcher annotations. Among those including some kind of biometric data are datasets that record human's responses to robots' facial expressions [Maitreyee and Ravi, 2021] or record self-reported and observed personality traits during dyadic and triadic HRIs [Celiktutan et al., 2017]. Other datasets without physiological data include datasets that register kids' social constructs and dynamics while playing with robots [Lemaignan et al., 2019] and tracking decreased user engagements with robots [Youssef et al., 2019]. The dataset introduced in this article was created to explore perceived safety. It was designed by roboticists and social scientists to include multidisciplinary study data gathered both during experimental sessions and during the follow-up dyadic qualitative interviews.

To enable data contextualization and access to the dataset, we built a knowledge graph of the database. Knowledge graphs are semantic representations of complex systems, used to discover, analyze, and retrieve data in many domains, including healthcare [Abu-Salih et al., 2023], materials science [Qiu et al., 2023], and robotics [Kwak et al., 2022]. The knowledge graph expresses the relations between the different streams of data and the functions, tools, and processes that generated them (i.e., data provenance). At the same time, it enables meaningful multimodal data integration and analyses. Implemented as a Neo4j database [n.d.], this system allows data to be retrieved in context with provenance information so that others understand the details of the experimental and social science research components, discover convergence between different points of data and facilitate inference making. The data model acts as the underlying semantic representation of the key research process components as a system of classes, their corresponding metadata, and the relationships between them. We created a data dictionary that includes the definition for each class and metadata element. We define several relevant queries for the knowledge graph related to the study, data, and results. Example queries include general study information like "How many human subjects participated in the study?" or "What data is included in this dataset?" Others focus on specific aspects, for example, interview related queries like "Where can I find the deductive codebook?" and "What themes are identified from thematic analysis?" Queries also give insights into interview data, for example "What results can I find from physiological signal analysis?" or "Show me the signal analysis results from all navigation behavior trials." Additional queries enable researchers to locate files containing results, research materials (e.g., questionnaires, interview protocol), or raw signal data from trials characterized by specific control conditions.

Because the Neo4j database understands Cypher language queries rather than natural language, we implemented a chatbot based on ChatGPT to help generate Cypher queries from natural language questions, thereby making the dataset more accessible (Figure 10). We prompted a custom chatbot using a detailed data model and schema that reflects the structure and terminology of our knowledge graph. This incorporates the domain-specific terms from our data dictionary and is supplemented with sample queries and annotated conversations that demonstrate how natural language can be accurately translated into Cypher queries. Nonetheless, one notable challenge with using ChatGPT is its tendency to generate confident yet potentially inaccurate responses, also known as hallucinations. To mitigate this, we do not use ChatGPT to directly retrieve data from the database. Instead, we employ a two-step process: first, we restrict ChatGPT to only generate a Cypher command in response to a user's natural language input; next, the user manually executes this returned Cypher statement in the Neo4j database. This approach not only ensures data accuracy by leveraging the precision of Cypher queries but also incorporates robust security measures to protect data integrity and privacy, especially critical for handling restricted and sensitive parts of our dataset.

7.2 Future Directions

This transdisciplinary study identifies a myriad of interesting problems for future investigation. First, several important contextual aspects for future study are identified. For example, findings suggest the importance of social dynamics, room geometry, and visibility conditions on perceived safety. Therefore, the identification of human response to these specific factors may provide important information to designers and engineers to consider. Furthermore, robot movement behavior during different tasks like navigation, search, or inspection that may occur in shared indoor environments is recommended to understand human responses to them. The impact of the number of robots nearby may also provide valuable information on human comfort in shared environments. Our results were inconclusive with respect to the single–multi-robot condition, but the impact of collisions and near-misses was clear. This also provides an interesting condition to study, in particular, people's feelings of safety in response to near collisions between robots. Inquiry along these lines may determine how close robots can move to each other without inducing stress or concern in community members. Further, in our experiment, participants remained seated, which limits the extent to which perceived safety is reflected in their actions. It would be interesting to study situations where humans encounter robots while engaging in different mobility patterns (e.g., standing and walking). Additionally, one could study the impact of the robot performing a certain task under different mobility constraints (e.g., moving at different speeds) or even studying the impact of a stationary robot. Analyzing their action selection (e.g., changing paths due to low perceived safety) could provide insights into the extent of perceived safety.

Second, we recommend the extended investigation of the highly multimodal dataset collected and published for open access [Gupta et al., 2024]. The knowledge graph provides researchers with the ability to find data in context with the processes and tools by which it was obtained, to find convergence between the multi-modal data. Specifically, the synchronized data files contain a wealth of valuable time series data, including four physiological signals (ECG, EDA, SCG, BVP), position and orientation of robots in the environment, and audio volume near participant seating locations. These signals also contain other information. For example, one may find the distance between the robot(s) and participants in the database and observe when the robots are moving versus when they are not, or observe when the robot is "looking" at participants or facing away. These time series data allow for the study of various robot and environment-related conditions impacting perceived safety. Proxemics and noise from robots have been widely researched in user-studies. Insights identified from quantitative physiological response may pave the way for an online pipeline whereby signals from a user's sensors could be decoded and used to inform robot behavior in real time. A quantitative model of perceived safety may lead to mitigation of psychological harm in nearby humans [Akalin et al., 2023]. For example, a quantitative model may act as an implicit communication tool that robots can use to guide their behavior. These data thus provide a novel context (i.e., indoor encounters with mobile quadrupeds) for exploring these features of perceived safety.

Finally, the findings from the inductive TA and the relationships among the themes and the known factors influencing perceived safety offer insight to future iterations of a taxonomy of perceived safety. Specifically, Figure 6 offers a starting point for a future taxonomy. Future hypothesis-driven research along this avenue offers genuine benefits to the stakeholders in community-embedded robotics. Furthermore, further investigation into the differences between HRIs and encounters is required. Overall, as mobile robots proliferate in our communities, it remains an open challenge in the HRI literature to study and understand human response and preference to encounters in a wide range of environments and scenarios.

8 Conclusion

This article presents an in-depth, transdisciplinary study of perceived safety during indoor encounters with mobile robots. Two autonomously navigating quadrupedal robots performed navigation and search behaviors in a shared apartment environment with two participants at a time. We investigated the impacts of participants seated together versus isolated, and also studied the impact of single versus multiple robots in the shared environment. These control conditions reflect our research questions, which seek to consider the impact that robots might have on pedestrians and bystanders in shared environments. These questions consider important aspects of mobile robots, and they reflect realistic situations that are likely to occur upon their widespread deployment in our communities. We utilized sociotechnical approaches from several disciplines, including physiological signal analysis and qualitative analysis of interview data, in order to learn about human perceived safety under these various study conditions tested. Findings come from (1) disciplinary analyses of physiological signals and interviews with participants to investigate the impact of robot behavior, number of robots, and presence of other humans on participants' perceived safety, and (2) the interdisciplinary investigation of these conditions to more thoroughly study the disciplinary results generated from each method. Results from the physiological signal analysis suggest that (a) the presence of robots induced some level of stress compared with baseline settings, (b) participants were more nervous under multi-robot compared to single robot conditions, and (c) participants were more nervous under navigation behavior compared to search behavior. The interview data demonstrated that the presence of others provides a level of comfort, contrasting the physiological results that demonstrated no correlation. Quantitative Content Analysis also demonstrated the close relationship of the existing taxonomy of perceived safety [Akalin et al., 2023] with the interview data. TA identified sense-making as a key theme in the interview data. Particularly, it suggested that the sense-making process enables participants to reduce feelings of uncertainty and gain a sense of control during experiments. When the data were considered together, we found that near collisions between robots likely confounded the physiological signal findings (b) and (c). The deeper understanding offered by the complementary methods represent a key component of this work. This both justifies the transdisciplinary approach and supports the combination of disparate methods in future studies to inspect the complex research problem of human perceptions of autonomous robots. We also investigated the impacts of repeated exposure to the robot encounters on the responses from participants to determine if the within-subject design was justified for this study. The analysis supported the study design and indicated that the repeated exposure had little effect on participants feelings, perceptions, and physiological responses. Finally, we detailed our open source dataset and propose several new research directions related to human–robot encounters. Particularly, the confounding factor of near collisions and the impact these have on bystanders represent an interesting future research direction, as both the physiological signal and interview data relate these situations to a stress response. Clearly, community embedded robotics necessitate the research of many complex problems including social navigation, exploration, mapping, locomotion, physically safe HRI, and perceived safety. Ultimately, the future study of this complex research problem *requires* transdisciplinarity (i.e., the meshing of multiple disciplines to create new processes of design, engineering, and deployment) in order to paint a complete picture of humans' perceptions, feelings, and preferences.

References

Bilal Abu-Salih, Muhammad Al-Qurishi, Mohammed Alweshah, Mohammad Al-Smadi, Reem Alfayez, and Heba Saadeh. 2023. Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. *Journal of Big Data* 10, 1 (2023), 81.

Neziha Akalin, Andrey Kiselev, Annica Kristoffersson, and Amy Loutfi. 2023. A taxonomy of factors influencing perceived safety in human–robot interaction. *International Journal of Social Robotics* 15, 12 (2023), 1993–2004.

Neziha Akalin, Annica Kristoffersson, and Amy Loutfi. 2022. Do you feel safe with your robot? Factors influencing perceived safety in human-robot interaction based on subjective and objective measures. *International Journal of Human-Computer Studies* 158 (2022), 102744.

Pourya Aliasghari, Moojan Ghafurian, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. 2023. How do we perceive our trainee robots? Exploring the impact of robot errors and appearance when performing domestic physical tasks on teachers' trust and evaluations. *ACM Transactions on Human-Robot Interaction* 12, 3 (2023), 1–41.

Minja Axelsson, Micol Spitale, and Hatice Gunes. 2024. Robots as Mental Well-Being Coaches: Design and Ethical Recommendations. Retrieved from https://doi-org.ezproxy.lib.utexas.edu/10.1145/3643457

Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1 (2009), 71–81.

Sarnab Bhattacharya, Mohammad Nikbakht, Alec Alden, Philip Tan, Jieting Wang, Taha A. Alhalimi, Sangjun Kim, Pulin Wang, Hirofumi Tanaka, Animesh Tandon, et al. 2023. A chest-conformable, wireless electro-mechanical E-tattoo for measuring multiple cardiac time intervals. *Advanced Electronic Materials* 9, 9 (2023), 2201284.

Joydeep Biswas and Manuela M. Veloso. 2017. Episodic non-Markov localization. *Robotics and Autonomous Systems* 87 (2017), 162–176.

Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

Virginia Braun and Victoria Clarke. 2012. *Thematic Analysis.* American Psychological Association.

Agostino Brugnera, Cristina Zarbo, Mika P. Tarvainen, Paolo Marchettini, Roberta Adorni, and Angelo Compare. 2018. Heart rate variability during acute psychosocial stress: A randomized cross-over trial of verbal and non-verbal laboratory stressors. *International Journal of Psychophysiology* 127 (2018), 17–25.

Ronald S. Burt. 1987. Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology* 92, 6 (1987), 1287–1335.

Beatrice Capelli, María Santos, and Lorenzo Sabattini. 2024. Towards the legibility of multi-robot systems. *ACM Transactions on Human-Robot Interaction* 13, 2 (2024), 1–32.

Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. 2017. Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing* 10, 4 (2017), 484–497.

Benjamin Charrow, Sikang Liu, Vijay Kumar, and Nathan Michael. 2015. Information-theoretic mapping using cauchy-schwarz quadratic mutual information. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4791–4798.

Tiffany L. Chen, Chih-Hung King, Andrea L. Thomaz, and Charles C. Kemp. 2011. Touched by a robot: An investigation of subjective responses to robot-initiated touch. In *Proceedings of the 6th International Conference on Human-Robot Interaction*, 457–464.

Bernard C. K. Choi and Anita W. P. Pak. 2006. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical & Investigative Medicine* 29 (2006), 6.

Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The Journal of Positive Psychology* 12, 3 (2017), 297–298.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 1 (1960), 37–46.

Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1994. Perceived stress scale. *Measuring Stress: A Guide for Health and Social Scientists* 10, 2 (1994), 1–2.

Frédéric Dehais, Emrah Akin Sisbot, Rachid Alami, and Mickaël Causse. 2011. Physiological and subjective evaluation of a human–robot object hand-over task. *Applied Ergonomics* 42, 6 (2011), 785–791.

Dimitriy A. Dimitriev, Elena V. Saperova, and Aleksey D. Dimitriev. 2016. State anxiety and nonlinear dynamics of heart rate variability in students. *PloS One* 11, 1 (Jan. 2016).

Nicola Döring and Sandra Poeschl. 2019. Love and sex with robots: A content analysis of media representations. *International Journal of Social Robotics* 11, 4 (2019), 665–677.

Empatica. 2024. *Empatica E4 Smart Watch*. Retrieved from https://www.empatica.com/research/e4/

Matan Eshed, Matan Epstein, Ada H. Zohar, and Goren Gordon. 2021. Studying dynamics of human information gathering behaviors using social robots. *Frontiers in Psychology* 12 (2021), 669198.

Leopoldina Fortunati, Alessandra Sorrentino, Laura Fiorini, and Filippo Cavallo. 2021. The rise of the roboid. *International Journal of Social Robotics* 13 (2021), 1457–1471.

Anthony Francis, Claudia Pérez-D'Arpino, Chengshu Li, Fei Xia, Alexandre Alahi, Rachid Alami, Aniket Bera, Abhijat Biswas, Joydeep Biswas, Rohan Chandra, et al. 2023. Principles and guidelines for evaluating social robot navigation algorithms. arXiv:2306.16740. Retrieved from https://arxiv.org/abs/2306.16740

Danilo Gallo, Prescillia Leslie Bioche, Jutta Katharina Willamowski, Tommaso Colombino, Shreepriya Gonzalez- Jimenez, Herve Poirier, and Cecile Boulard. 2023. Investigating the integration of human-like and machine-like robot behaviors in a shared elevator scenario. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 192–201.

Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. Jr. 2003. A very brief measure of the big-five personality domains. *Journal of Research in Personality* 37, 6 (2003), 504–528.

Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. 2015. cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Bio-Medical Engineering* 63, 4 (2015), 797–804.

Alberto Greco, Gaetano Valenza, Jesus Lazaro, Jorge Mario Garzon-Rey, Jordi Aguilo, Concepcion De-la-Camara, Raquel Bailon, and Enzo Pasquale Scilingo. 2021. Acute stress state classification based on electrodermal activity modeling. *IEEE Transactions on Affective Computing* 14, 1 (2021), 788–799.

Ben Grocholsky, James Keller, Vijay Kumar, and George Pappas. 2006. Cooperative air and ground surveillance. *IEEE Robotics and Automation Magazine* 13, 3 (2006), 16–25.

Ryan Gupta, Hyonyoung Shin, Emily Norman, Zhiyun Deng, Maria Esteva, Nanshu Lu, Keri K. Stephens, and Luis Sentis. 2024. *Community Embedded Robotics: A Multimodal Dataset on Perceived Safety during Indoor Mobile Robot Encounters*. Version V1. Retrieved from https://doi.org/10.18738/T8/FT9VYS

Sjoerd van Halem, Eeske van Roekel, Lara Kroencke, Niclas Kuper, and Jaap Denissen. 2020. Moments that matter? On the complexity of using triggers based on skin conductance to sample arousing events within an experience sampling framework. *European Journal of Personality* 34, 5 (2020), 794–807.

Edward T. Hall. 1966. The hidden dimension. In *Garden City*, Vol. 609, Anchor.

Edward T. Hall, Ray L. Birdwhistell, Bernhard Bock, Paul Bohannan, A. Richard Diebold, Jr., Marshall Durbin, Munro S. Edmonson, J. L. Fischer, Dell Hymes, Solon T. Kimball, et al. 1968. Proxemics [and comments and replies]. *Current Anthropology* 9, 2/3 (1968), 83–108.

Trevor Hastie, Saharon Rosset, Ji Zhu, and Hui Zou. 2009. Multi-class adaboost. *Statistics and Its Interface* 2, 3 (2009), 349–360.

Elliott Hauser, Yao-Cheng Chan, Parth Chonkar, Geethika Hemkumar, Huihai Wang, Daksh Dua, Shikhar Gupta, Efren Mendoza Enriquez, Tiffany KaoJustin Hart, et al. 2023. "What's that robot doing here?": Perceptions of incidental encounters with autonomous quadruped robots. In *Proceedings of the 1st International Symposium on Trustworthy Autonomous Systems*, 1–15.

Susanna Heldt Cassel and Teresa Miranda Maureira. 2017. Performing identity and culture in indigenous tourism–a study of indigenous communities in Québec, Canada. *Journal of Tourism and Cultural Change* 15, 1 (2017), 1–14.

Heeyong Huh, Xiangxing Yang, Hyonyoung Shin, and Nanshu Lu. 2023. A multi-daywearable surface EMG E-tattoo for fatigue monitoring. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1–4.

Sarah Immanuel, Meseret N. Teferra, Mathias Baumert, and Niranjan Bidargaddi. 2023. Heart rate variability for evaluating psychological stress changes in healthy adults: A scoping review. *Neuropsychobiology* 82, 4 (Jun. 2023), 187–202.

Helene Joffe. 2012. Thematic analysis. In *Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners*. D. Harper and A. R. Thompson (Eds.), Wiley, Chichester, 209–223.

Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation* 15, 3 (2018), 235.

Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H. Hellhammer. 1993. The ‘trier social stress test’—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 1-2 (1993), 76–81.

Agata Klimek, Ittay Mannheim, Gerard Schouten, Eveline J. M. Wouters, and Manon W. H. Peeters. 2023. Wearables measuring electrodermal activity to assess perceived stress in care: A scoping review. *Acta Neuropsychiatrica* 37 (2023), 1–11.

Wei Koh, Qi Sally Whelan, Pascale Heins, Dympna Casey, Elaine Toomey, and Rose-Marie Dröes. 2022. The usability and impact of a low-cost pet robot for older adults and people with dementia: Qualitative content analysis of user experiences and perceptions on consumer websites. *JMIR Aging* 5, 1 (2022), e29224.

Klaus Krippendorff. 1989. Content analysis. *International Encyclopedia of Communication* 1, 1 (1989), 403–407.

Ji Ho Kwak, Jaejun Lee, Joyce Jiyoung Whang, and Sungho Jo. 2022. Semantic grasping via a knowledge graph of robotic manipulation: A graph representation learning approach. *IEEE Robotics and Automation Letters* 7, 4 (2022), 9397–9404.

LabRecorder. 2012. Retrieved from https://github.com/labstreaminglayer/App-LabRecorder

LabStreamingLayer. 2012. Retrieved from https://github.com/sccn/labstreaminglayer

Daegyu Lee, Gyuree Kang, Boseong Kim, and D. Hyunchul Shim. 2021. Assistive delivery robot application for real-world postal services. *IEEE Access* 9 (2021), 141981–141998.

Seungjae Lee, Ho Bin Hwang, Seongryul Park, Sanghag Kim, Jung Hee Ha, Yoojin Jang, Sejin Hwang, Hoon-Ki Park, Jongshill Lee, In Young Kim, et al. 2022. Mental stress assessment using ultra short term HRV analysis based on non-linear method. *Biosensors* 12, 7 (Jun. 2022), 465.

Séverin Lemaignan, Charlotte Edmunds, and Tony Belpaeme. 2019. *The PInSoRo Dataset—Anonymized Version*. Zenodo. Retrieved from https://doi.org/10.5281/zenodo.3248670.

Bojana Lobe and David L. Morgan. 2021. Assessing the effectiveness of video-based interviewing: A systematic comparison of video-conferencing based dyadic interviews and focus groups. *International Journal of Social Research Methodology* 24, 3 (2021), 301–312.

Wairagkar Maitreyee and Vaidyanathan Ravi. 2021. *EEG Data for Emotive Response to Robot Facial Expressions*. Version 1.0. Zenodo. Retrieved from https://doi.org/10.5281/zenodo.5784856

Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. 2021. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods* 53, 4 (2021), 1689–1696.

Jeremy A. Marvel, Shelly Bagchi, Megan Zimmerman, and Brian Antonishek. 2020. Towards effective interface designs for collaborative HRI in manufacturing: Metrics and measures. *ACM Transactions on Human-Robot Interaction* 9, 4 (2020), 1–55.

Christoforos Mavrogiannis, Patrícia Alves-Oliveira, Wil Thomason, and Ross A. Knepper. 2022. Social momentum: Design and evaluation of a framework for socially competent robot navigation. *ACM Transactions on Human-Robot Interaction* 11, 2 (2022), 1–37.

Paolo Melillo, Marcello Bracale, and Leandro Pecchia. 2011. Nonlinear heart rate variability features for real-life stress detection. case study: Students under stress due to university examination. *Biomedical Engineering Online* 10, 1 (2011), 96.

Reuth Mirsky, Xuesu Xiao, Justin Hart, and Peter Stone. 2024. Conflict avoidance in social navigation—A survey. *ACM Transactions on Human-Robot Interaction* 13, 1 (2024), 1–36.

Frederik Moesgaard, Lasse Hulgaard, and Mads Bødker. 2022. Incidental encounters with robots. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 377–384.

Nicolas E. Neef, Sarah Zabel, Mathis Lauckner, and Siegmar Otto. 2023. What is appropriate? On the assessment of human-robot proxemics for casual encounters in closed environments. *International Journal of Social Robotics* 15, 6 (2023), 953–967.

Margot M. E. Neggers, Simon Belgers, Raymond H. Cuijpers, Peter A. M. Ruijten, and Wijnand A. IJsselsteijn. 2024. Comfortable crossing strategies for robots. *International Journal of Social Robotics* 16, 7 (2024), 1541–1560.

Neo4j database. [n.d.]. Retrieved from https://neo4j.com/docs/cypher-manual/current/introduction/

Lorelli S. Nowell, Jill M. Norris, Deborah E. White, and Nancy J. Moules. 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods* 16, 1 (2017), 1609406917733847.

Bjorn Petrak, Katharina Weitz, Ilhan Aslan, and Elisabeth Andre. 2019. Let me show you your new home: Studying the effect of proxemic-awareness of robots on users’ first impressions. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–7.

Amanda Prorok, Matthew Malencia, Luca Carlone, Gaurav S. Sukhatme, BrianM Sadler, and Vijay Kumar. 2021. Beyond robustness: A taxonomy of approaches towards resilient multi-robot systems. arXiv:2109.12343. Retrieved from https://arxiv.org/abs/2109.12343

Qinjun Qiu, Kai Ma, Hairong Lv, Liufeng Tao, and Zhong Xie. 2023. Construction and application of a knowledge graph for iron deposits using text mining analytics and a deep learning algorithm. *Mathematical Geosciences* 55, 3 (2023), 423–456.

Martina Raue, Bernhard Streicher, and Eva Lermer. 2019. *Perceived Safety*. Springer.
Jakob Reinhardt, Lorenz Prasch, and Klaus Bengler. 2021. Back-off: Evaluation of robot motion strategies to facilitate human-robot spatial interaction. *ACM Transactions on Human-Robot Interaction* 10, 3 (2021), 1–25.
Stephanie Riger and Rannveig Sigurvinsdottir. 2016. Thematic analysis. In *Handbook of Methodological Approaches to Community-Based Research: Qualitative, Quantitative, and Mixed Methods*. L. A. Jason and D. S. Glenwick (Eds.), Oxford University Press, New York, 33–41.
Yara Rizk, Mariette Awad, and Edward W. Tunstel. 2019. Cooperative heterogeneous multi-robot systems: A survey. *ACM Computing Surveys* 52, 2 (2019), 1–31.
Kate Roberts, Anthony Dowell, and Jing-Bao Nie. 2019. Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development. *BMC Medical Research Methodology* 19 (2019), 1–8.
Robot Operating System. 2007. Retrieved from http://wiki.ros.org/costmap_2d
Silvia Rossi, Daniela Conti, Federica Garramone, Gabriella Santangelo, Mariacarla Staffa, Simone Varrasi, and Alessandro Di Nuovo. 2020. The role of personality factors and empathy in the acceptance and performance of a social robot for psychometric evaluations. *Robotics* 9, 2 (2020), 39.
Matteo Rubagotti, Inara Tusseyeva, Sara Baltabayeva, Danna Summers, and Anara Sandygulova. 2022. Perceived safety in physical human–robot interaction—A survey. *Robotics and Autonomous Systems* 151 (2022), 104047.
Matthew Rueben, Jeffrey Klow, Madelyn Duer, Eric Zimmerman, Jennifer Piacentini, Madison Browning, Frank J. Bernieri, Cindy M. Grimm, and William D. Smart. 2021. Mental models of a mobile shoe rack: Exploratory findings from a long-term in-the-wild study. *ACM Transactions on Human-Robot Interaction* 10, 2 (2021), 1–36.
Ismail Sahin. 2006. Detailed review of rogers' diffusion of innovations theory and educational technology-related studies based on rogers' theory. *Turkish Online Journal of Educational Technology* 5, 2 (2006), 14–23.
Lizawati Salahuddin, Jaegeol Cho, Myeong Gi Jeong, and Desok Kim. 2007. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 4656–4659.
Chris Seiffert, Taghi M. Khoshgoftaar, Jason Van Hulse, and Amri Napolitano. 2008. RUSBoost: Improving classification performance when training data is skewed. In *2008 19th International Conference on Pattern Recognition*, 1–4.
Sahaya G. Selvam and Joanna Collicutt. 2012. The ubiquity of the character strengths in African traditional religion: A thematic analysis. In *Well-Being and Cultures: Perspectives from Positive Psychology*. Springer, Dordrecht, Netherlands, 83–102.
Masahiro Shiomi, Francesco Zanlungo, Kotaro Hayashi, and Takayuki Kanda. 2014. Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model. *International Journal of Social Robotics* 6, 3 (2014), 443–455.
Kavan Singh Sikand, Logan Zartman, Sadegh Rabiee, and Joydeep Biswas. 2021. Robofleet: Open source communication and management for fleets of autonomous robots. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 406–412.
Lauryn A. Spearing, Amal Bakchan, Leigh C. Hamlet, Keri K. Stephens, Jessica A. Kaminsky, and Kasey M. Faust. 2022. Comparing qualitative analysis techniques for construction engineering and management research: The case of arctic water infrastructure. *Journal of Construction Engineering and Management* 148, 7 (2022), 04022058.
Mariacarla Staffa and Silvia Rossi. 2022. Enhancing affective robotics via human internal state monitoring. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 884–890.
Jessi Stark, Roberta R. C. Mota, and Ehud Sharlin. 2018. Personal space intrusion in human-robot collaboration. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 245–246.
Chee Sheng Tan, Rosmiwati Mohd-Mokhtar, and Mohd Rizal Arshad. 2021. A comprehensive review of coverage path planning in robotics using classical and heuristic algorithms. *IEEE Access* 9 (2021), 119310–119342.
Ada V. Taylor, Ellie Mamantov, and Henny Admoni. 2022. Observer-aware legibility for social navigation. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1115–1122.
Gareth Terry, Nikki Hayfield, Victoria Clarke, and Virginia Braun. 2017. Thematic analysis. In *The SAGE Handbook of Qualitative Research in Psychology*, 25.
Julie M. Turner-Cobb, Mashal Asif, James E. Turner, Chris Bevan, and Danae Stanton Fraser. 2019. Use of a non-human robot audience to induce stress reactivity in human participants. *Computers in Human Behavior* 99 (2019), 76–85.
Sanne Van Waveren, Rasmus Rudling, Iolanda Leite, Patric Jensfelt, and Christian Pek. 2023. Increasing perceived safety in motion planning for human-drone interaction. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 446–455.
Christian Vassallo, Anne-Hélène Olivier, Philippe Souères, Armel Crétual, Olivier Stasse, and Julien Pettré. 2017. How do walkers avoid a mobile robot crossing their way? *Gait and Posture* 51 (2017), 97–103.
Valeria Villani, Fabio Pini, Francesco Leali, and Cristian Secchi. 2018. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics* 55 (2018), 248–266.

Ziming Wang, Ziyi Hu, Björn Rohles, Sara Ljungblad, Vincent Koenig, and Morten Fjeld. 2023. The effects of natural sounds and proxemic distances on the perception of a noisy domestic flying robot. *ACM Transactions on Human-Robot Interaction* 12, 4 (2023), 1–32.

Karl E. Weick, Kathleen M. Sutcliffe, and David Obstfeld. 2005. Organizing and the process of sensemaking. *Organization Science* 16, 4 (2005), 409–421.

Christian J. A. M. Willemse, Alexander Toet, and Jan B. F. Van Erp. 2017. Affective and behavioral responses to robotinitiated social touch: Toward understanding the opportunities and limitations of physical contact in human–robot interaction. *Frontiers in ICT* 4 (2017), 12.

Shuen-De Wu, Chiu-Wen Wu, Shiou-Gwo Lin, Chun-Chieh Wang, and Kung-Yen Lee. 2013. Time series analysis using composite multiscale entropy. *Entropy* 15, 3 (2013), 1069–1084. Retrieved from https://www.mdpi.com/1099-4300/15/3/1069

Atef Ben Youssef, Giovanna Varni, Slim Essid, and Chloé Clavel. 2019. On-the-fly detection of user engagement decrease in spontaneous human-robot interaction. *International Journal of Social Robotics* 11, 5 (2019), 815–828.

Kaiyu Zheng, Anirudha Paul, and Stefanie Tellex. 2023. A system for generalized 3D multi-object search. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1638–1644.

Tian Zhou, Jackie S. Cha, Glebys Gonzalez, Juan P. Wachs, Chandru P. Sundaram, and Denny Yu. 2020. Multimoda physiological signals for workload prediction in robot-assisted surgery. *ACM Transactions on Human-Robot Interaction* 9, 2 (2020), 1–26.

Appendices

A Robot Deployment Information

This appendix describes information relevant to the autonomous deployment of the Go1 and Spot robots. The Go1 robot is equipped with an Intel NUC computer that is connected to an RPLidar A3 2D LiDAR sensor that casts a 360-degree view of the robot's surroundings. Similarly, the Spot is equipped with an MSI GS66 Stealth laptop connected to a Velodyne Puck 3D LiDAR sensor. We use the raw 2D LaserScan output from the sensor. Both robots utilize a pre-built map of the environment, represented as a set of vectors. The robots utilize Episodic Non-Markov Localization [Biswas and Veloso, 2017] in order to maintain an accurate pose estimate throughout each experiment trial. Open source code for this method can be found at https://github.com/ut-amrl/enml. They utilize a graph-based navigation tool in order to follow a series of pre-defined waypoints pre-generated to facilitate reproducibility between different sessions. Open source code for the navigation tool can be found at https://github.com/ut-amrl/graph_navigation. Both robots maintain communication with a base station operated by the researcher. Communications are achieved over a local network using Robofleet [Sikand et al., 2021]. This enables the researcher to send the new waypoint series to the robot at the beginning of each individual trial.

B *ndEDRfreq* and HRV_{CMNSEn} Plots over Trials for Individual Participants

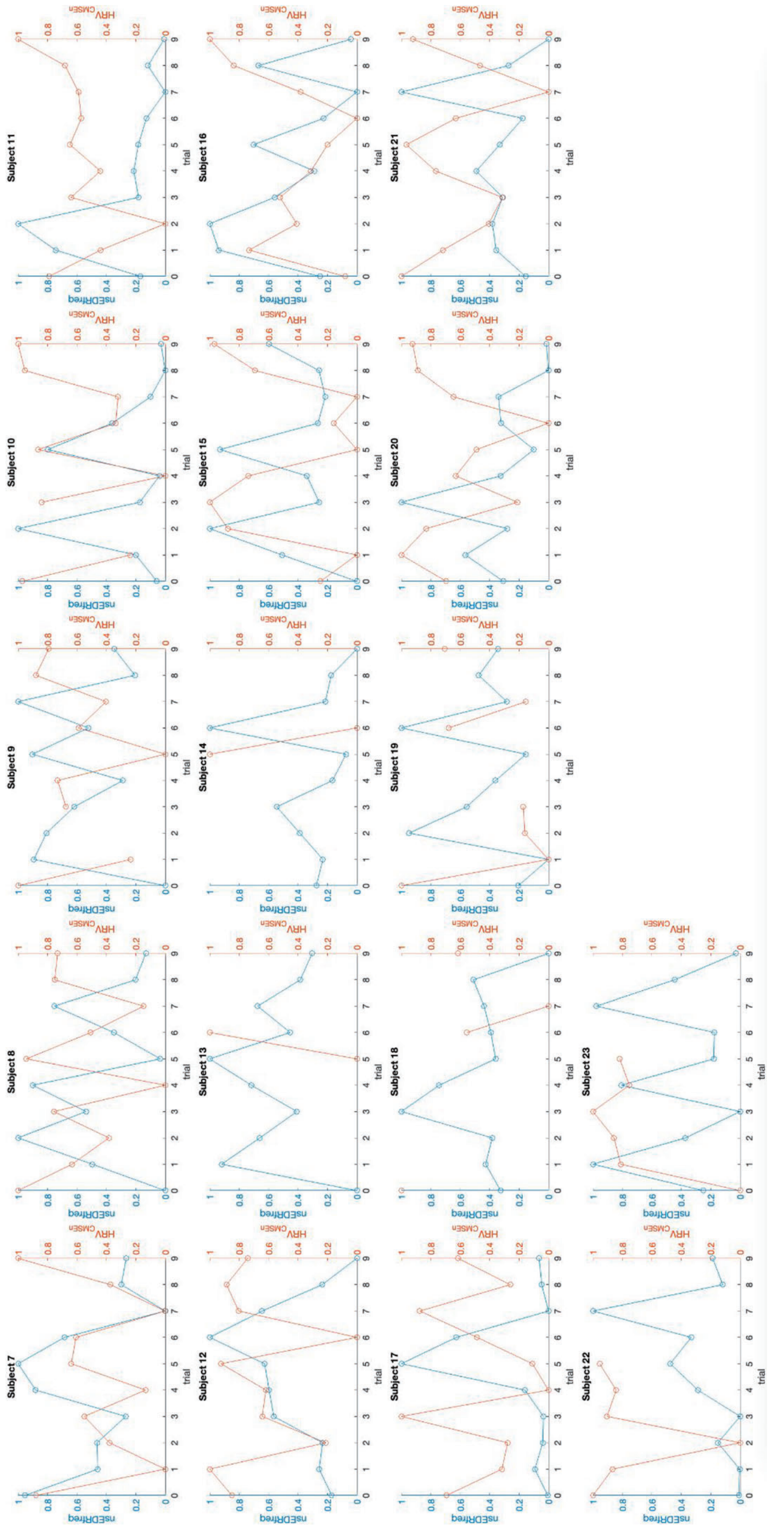

Fig. B1. Normalized features HRV_{CMSEn} and $nsEDRfreq$ over trials, i.e., experiment time. Plots are given for the N = 17 participants with analyzed physiological data.

C Full Mann–Whitney–Wilcoxon Test Statistics

Here, we report a table with the full results of the Mann–Whitney–Wilcoxon tests including the adjusted p-values after multiple comparisons correction wherever applicable, and the U statistics (Tables C1 and C2).

Table C1. Mann–Whitney–Wilcoxon Test Results for the Comparisons of HRV Features Conducted in Figure 5

HRV	p-value	U Statistic
Baseline vs. robot	$9.252x10^{-6}$	2,316
Baseline vs. separate	0.005171	1,077
Separate vs. together	0.001612	1,987
Baseline vs. together	$2.521x10^{-6}$	1,238
Search vs. navigation	0.004798	1,938
Baseline vs. search	0.001317	1,116
Baseline vs. navigation	0.000017	1,200
Baseline vs. isolated	0.0001483	1,172
Isolated vs. social	1	1,284
Baseline vs. social	0.0002006	1,144

Table C2. Mann–Whitney–Wilcoxon Test Results for the Comparisons of EDA Features Conducted in Figure 5

EDA	p-value	U Statistic
Baseline vs. robot	$2.318x10^{-6}$	1,099
Baseline vs. separate	0.0001899	592
Separate vs. together	1	2,203
Baseline vs. together	$1.246x10^{-5}$	507
Search vs. navigation	0.04054	1,744
Baseline vs. search	0.001993	676
Baseline vs. navigation	$5.984x10^{-7}$	423
Baseline vs. isolated	$2.143x10^{-5}$	486
Isolated vs. social	1	2,203
Baseline vs. social	0.0001078	613

D Additional Research Materials

Pre-Experiment Questionnaire
Demographics: –What is your age? –What is your role at the university? –What is your gender?
Baseline Stress: For each question choose from the following alternatives: 0—never, 1—almost never, 2—sometimes, 3—fairly often, 4—very often –In the last month, how often have you felt nervous and stressed? –In the last month, how often have you found that you could not cope with all the things that you had to do? –In the last month, how often have you felt that you were on top of things? –In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?
Personality: Here are a number of personality traits. Please rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other. 1—Disagree Strongly, 2—Disagree Moderately, 3—Disagree a little, 4—Neither agree nor disagree, 5—Agree a little, 6—Agree moderately, 7—Agree strongly. I see myself as: –Extraverted, Enthusiastic –Critical, Quarrelsome –Dependable, Self-disciplined –Anxious, Easily upset –Open to new experiences, Complex –Reserved, Quiet –Sympathetic, Warm –Disorganized, Careless –Calm, Emotionally stable –Conventional, Uncreative

Between-Trial Questionnaire
Please rate your agreement with the below statements regarding the preceding robot encounter on the 7-point scale: –During the last trial I felt uncomfortable. –Sharing space with mobile robots is confusing for me. –During the last trial I felt comfortable. –Sharing space with mobile robots is uneasy for me. –Sharing space with mobile robots is uncomfortable for me. –During the last trial I felt anxious. –Sharing space with mobile robots is difficult for me. –Sharing space with mobile robots is annoying for me. –During the last trial I felt relaxed. –Sharing space with mobile robots is disappointing for me.

Semi-Structured Interview Protocol

Hi everyone, my name is Emily, and I will be guiding your conversation today. I'm a PhD student at the Moody College of Communication, and I am working alongside Ryan in order to understand more about your experience today.

Throughout the next 45 minutes I will be asking you all a series of questions, and you can feel free to respond directly to me, to your fellow participants, or not at all. Note that we will be recording this conversation, but all of your information will be anonymized—including changing your name to a pseudonym.

If at any point during the study you feel uncomfortable and would like to leave or step outside the room, feel free to quietly raise your hand, and we will pause the recording. Furthermore, you are welcome to end your participation at any time. Now, I am going to lay out a few ground rules:

(1) When I am asking you questions, understand that there are not right or wrong answers. We are primarily interested in your thoughts, feelings, about your experience.
(2) You do not always have to agree or echo the sentiment of your fellow participants. While sometimes you might agree, it is also valuable that we are able to have disagreements.
(3) Feel free to express both positive and negative feelings. As I mentioned before, we are most interested in your feelings about the experience.
(4) Please turn off or silence your cell phones. Unintended interruptions can cause your fellow participants to lose their train of thought. Please be respectful of everyone's time.

Begin Recording

(1) Let us start with some basic questions. Tell me a little bit about yourself.
  (a) Are you a student, faculty, or staff member?
  (b) What made you interested in participation?
  (c) What is your prior experience with robots?
(2) Tell me about your experience with the robot(s) today.
(3) How did you feel when the robot first came out from around the corner?
  (a) **Probe:** Did this feeling change throughout the course of the study?
  (b) **Conversation:** Do you agree with [fellow participant]?/what about you?
  (c) What kinds of sounds would you compare the noise to?
(4) How did you feel about wearing the sensors?
  (a) **Probe:** Was there any point that they felt uncomfortable?
(5) Were there any moments during today's experiment that stood out to you?
  (a) **Probe:** Why did that stand out to you?
  (b) **Conversation:** Do you agree? Did that moment stand out to you as well, or not so much?
(6) Did anything happen that made you feel nervous/stressed? Or not so much?
  (a) **Probe:** What about that made you feel stressed? How did you know?
  (b) **Conversation:** Was that stressful/made you nervous too [fellow participant]?
  (c) Why or why not did you find/not find that stressful?
(7) At any point during the experiment, did you feel unsafe or did that not happen to you?
  (a) **Probe:** Help me understand why you felt that way (**rotate safe and unsafe based on the conversation**).
(8) Was there anything that made you feel particularly comfortable or happy?
(9) During the experiment, the got closer to you. How did you feel about that?
  (a) **Probe:** How might you feel if the robot was closer to you for a prolonged period of time?
  (b) **Probe:** How might you feel if the robot if the robot stared at you?

Now let us talk about the differences between sitting alone with others.

(10) Were there any differences in your feelings when you were sitting alone vs. when you were sitting with your fellow participants?

Semi-Structured Interview Protocol

(a) **Probe:** Did you feel more comfortable with the robot alone or with others? Why is that?
(b) **Conversation:** Are robots more scary or dangerous when someone is alone?

(11) What did you notice about how your fellow participants reacted to the robot?
(a) **Probe:** What specific actions or facial expressions did you notice?

Now let us talk about your thoughts and feelings about robots in general.

(12) Let us pretend you had to describe to someone who had no knowledge of robots, your experience today. How would you describe that?

(13) Where have you been exposed to robots before today?
(a) **Probe:** The media? Campus? Research Opportunities/Positions?

(14) What types of tasks could you see robots performing in a household?
(a) What would YOU like for them to do?

(15) Was today's experience similar or different to your prior experience with robots?
(a) **Probe:** (If they mention being closer to a robot than ever before:) How do you feel having a close encounter with a robot?

(16) What is your overall opinion on robots?

(17) How do you think this experience will affect your thinking about robots in the future?

(18) What would you tell others about robots, now that you have had this experience?

(19) How do you think you will react the next time you see a robot "in the wild?"

(If they raise ideas about stress, go here. If not, go to the last question)

Now, I'm going to ask you a few questions about yourself.

(20) How would you describe your personality to someone who has never met you before?

(21) When you think about being stressed, would you say that it happens often or not so often?

(22) How do you know when you are stressed?

(23) What types of things typically cause you to feel stress?

(24) What is something that might help alleviate your stress?

(25) We are nearing the end of the interview. Is there anything you would like to ask or talk to your fellow participant about?
(a) What about the research team?

(26) Is there anything you would like to share with me or your fellow participant as we're wrapping up this conversation?

I will now stop the recording.

If you have any questions or concerns about anything we have done or discussed today, please feel free to reach out to the e-mail address located on your consent form.

Received 24 May 2024; revised 3 January 2026; accepted 5 March 2026