Crucial Clues: Investigating Psychophysiological Behaviors for Measuring Trust in Human-Robot Interaction

Abdullah Alzahrani
Swansea University
Swansea, UK
2043528@swansea.ac.uk

Muneeb Imtiaz Ahmad
Swansea University
Swansea, UK
m.i.ahmad@swansea.ac.uk

ABSTRACT

Existing work on the measurements of trust during Human-Robot Interaction (HRI) indicates that psychophysiological behaviours (PBs) have the potential to measure trust. However, we see limited work on the use of multiple PBs in combination to calibrate human's trust in robots in real-time during HRI. Therefore, this study aims to estimate human trust in robots by examining the differences in PBs between trust and distrust states. It further investigates the changes in PBs across repeated HRI and also explores the potential of machine learning classifiers in predicting trust levels during HRI. We collected participants' electrodermal activity (EDA), blood volume pulse (BVP), heart rate (HR), skin temperature (SKT), blinking rate (BR), and blinking duration (BD) during repeated HRI. The results showed significant differences in HR and SKT between trust and distrust groups and no significant interaction effect of session and decision for all PBs. Random Forest classifier achieved the best accuracy of 68.6% to classify trust, while SKT, HR, BR, and BD were the important features. These findings highlight the value of PBs in measuring trust in real-time during HRI and encourage further investigation of trust measures with PBs in various HRI settings.

CCS CONCEPTS

• **Human-centered computing → Human Robot interaction** ; User studies; • **Computer systems organization** → Robotics.

KEYWORDS

Trust, Measurement, Psychophysiological behaviours, Human-Robot Interaction, Real-time

ACM Reference Format:
Abdullah Alzahrani and Muneeb Imtiaz Ahmad. 2023. Crucial Clues: Investigating Psychophysiological Behaviors for Measuring Trust in Human-Robot Interaction. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '23), October 9–13, 2023, Paris, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3577190.3614148

1 INTRODUCTION

In an increasingly interconnected world, human-robot interactions (HRI) are becoming more prevalent across various collaborative and competitive settings, such as healthcare, manufacturing, economics, and sports [28, 45, 46, 48]. In these settings, establishing and maintaining trust is essential for successful and efficient HRI, as it directly impacts user acceptance, safety, and overall performance [20]. Consequently, humans need to have a balanced level of trust in robots, which is an optimal level of trust that neither underestimates nor overestimates the robot's capabilities. This helps to prevent disuse and the loss of benefits from using the system [43]. Identifying the balanced level of trust highlights the importance of developing real-time, online trust measurement methods during HRI. However, measuring human trust in robots presents a challenge, as factors affecting trust, such as context, robot characteristics, and individual differences, can vary significantly [8].

In HRI, researchers use two methods to assess human trust: subjective trust and objective trust measurement [25]. The subjective measuring method involves evaluating the responses of experiment participants to questionnaires meant to determine people's trust in the robots [16]. In contrast, objective trust measuring methods analyse how experiment participants interact with robots, as opposed to depending on participants' assumptions about themselves [29]. Objective trust measurement methods are less frequently used in human-robot trust studies compared with subjective trust measurement methods [25]. However, the subjective method may not capture the dynamic nature of trust during real-time interactions [9]. Objectively measuring trust can analyse user behaviours during an interaction with robots in real-time which can be beneficial for understanding and optimizing these interactions. Real-time trust measurement ensures effective and efficient interactions with robots over time, including optimising decision-making [38]. It allows for real-time capture of trust levels, enabling robotic systems to adapt their communication strategies and provide more trustworthy and persuasive information, thereby improving the overall user experience [38].

Considering the challenges of representing human trust mathematically in robots within HRI [17], researchers have explored alternative approaches for assessing humans trust in robots. One such approach, which builds upon the concept of objective trust measurement, is the use of human psychophysiological behaviours (PBs) [41]. PBs, such as electrodermal activity (EDA), blood volume pulse (BVP), heart rate (HR), blinking rate (BR) and blinking duration (BD), offer a promising avenue for understanding and assessing trust in real-time [5]. These behaviours indicate an individual's emotional and cognitive states during interactions with robots. By monitoring these responses, we can gain insights into the dynamic nature of trust and its impact on HRI [4].

The importance of measuring trust during long-term HRI has been recognized, but there is limited work in this area [36]. Trust

ICMI '23, October 9–13, 2023, Paris, France

ACM ISBN 979-8-4007-0055-2/23/10…$15.00
https://doi.org/10.1145/3577190.3614148

calibration is essential in long-term HRI, as it involves adjusting trust levels based on the robot's performance, human experience, and evolving expectations [14]. In this context, research in HRI indicates that trust levels change according to experiences gained over time [13, 21]. To the best of our knowledge, the analysis of PBs to assess and monitor trust in repeated interactions remains unexamined. Addressing this gap by incorporating PBs could provide valuable insights into trust dynamics and contribute to the development of adaptive robots capable of fostering positive long-term interactions.

In light of these considerations, we investigate the following research questions:

- RQ1: How do specific human PBs differ between trusting and distrusting states during interactions with a robotic agent?
- RQ2: How do PBs evolve as individuals gain experience during repeated HRI?
- RQ3: Which classification algorithms demonstrate the highest accuracy and performance in predicting trust levels based on PBs during HRI?
- RQ4: Which psychophysiological features are predictive of trust or distrust behaviours during HRI?

To investigate these RQs, we conducted an experiment that enabled participants to play a game involving instances to either trust or distrust the NAO robot. We recorded PBs, including EDA, BVP, HR, SKT, BR, and BD to measure humans' trust in robots in real-time. The novel contributions of this paper are as follows:

- We present a comprehensive analysis of the relationship between human PBs and trust levels in HRI, providing insights into the most indicative behaviours for real-time trust measurement.
- We show that the use of multiple psychophysiological as a combination can potentially classify two different levels of trust during long-term HRI, highlighting the potential for trust calibration and the design of adaptive robotic systems.
- We share the study materials and evolving dataset with the community to advance knowledge on trust in HRI which can be found here.

The remainder of this paper is organized as follows: Section 2 provides background and discusses relevant literature. Section 3 thoroughly describe the study. Sections 4 & 5 present the results and their discussion. Finally, section 6 concludes the paper.

2 BACKGROUND & RELATED WORK

2.1 Trust conceptualization

Trust is a multifaceted and complicated concept, and despite extensive research efforts, there is still no comprehensive and universally accepted definition [1, 20]. The Merriam-Webster Dictionary lists trust as "*assured reliance on the character, ability, strength, or truth of someone or something*". Rotter [42] defined trust as a "*generalized expectancy held by an individual that the word, promise, oral or written statement of another individual or group can be relied on*". The term *generalized expectancy* represents the combined effect of an individual's experiences and interactions with another entity, such as a person or technology, which ultimately influences their trust level. This generalized expectancy can be inferred from physiological measures. Relaxation may indicate trust, while heightened alertness may indicate distrust or uncertainty towards a robot. Ajenaghughrure et al. [5] defined trust as a subconscious compound cognitive process. This involves mental deliberation, reasoning, and mental processing, which include memory, learning, and accumulated knowledge. Physiological measures can offer insight into real-time trust evaluations, even when they do not align with conscious trust affirmations. In this paper, we adopt and build on the above definitions. These definitions are highly relevant to our research for several reasons. First, they highlight the dynamic and evolving nature of trust, which is crucial when examining and assessing human trust in robots in real-time using PBs. Second, they underscore the importance of trust in competitive situations where a robot's truthfulness significantly influences human decision-making.

2.2 Measurement of trust

In addition to subjective [35, 44, 50] and objective methods [23, 25, 29] for measuring trust during HRI, researchers have explored the use of PBs to assess trust in collaborative contexts [3, 19, 22, 24]. Past research has identified two methods in which PBs were used to access an individuals trust in robots. 1) Empirical evaluations, and 2) Machine learning methods.

Empirical Evaluations and Psychophysiological Behaviors - Psychophysiology is the scientific field that examines the relationship between human physiological responses and psychological states, such as emotions and trust, which gives rise to the term psycho + physiology [47]. This discipline entails using physiological sensors to capture and record human physiological changes (psychophysiological signals) during psychological experiences, like emotions and trust [15, 27]. These physiological sensors constantly track and document alterations in four distinct human organs: (1) the brain, through measuring neurological activity via an electroencephalogram (EEG); (2) the heart, through measuring HR, or BVP; (3) skin, through measuring EDA or skin temperature (SKT); and (4) eyes, through measuring BR or BD [2].

PBs have been studied across various disciplines, including psychology, neuroscience, medicine, games and human-robot interaction [4, 10, 27]. In HRI, these PBs have been proposed as alternative methods for assessing trust in HRI [11]. These behaviours reflect an individual's emotional and cognitive states and can provide valuable insights into their trust-related responses during interactions [5]. Several studies have explored the use of PBs in human-robot trust research such as [6, 26, 32, 33].

Khawaji et al. [26] investigated the use of galvanic skin response (GSR) in measuring trust and cognitive load in a text-based chat environment. The study evaluated the GSR signals at four gradients and overlapping trust and cognitive load conditions. Participants engaged in a text-chat conversation while playing an investment game. Lu and Sarter [33] explored eye movement as a measure of trust in automation. Participants engaged in a target identification task. Eye fixation data were collected to assess trust in the automation based on participants' visual attention and system reliability. The results suggest that eye tracking may be a valuable tool to trust calibration based on priming and system reliability. The results showed that GSR signals were significantly affected by trust conditions and were higher in the high level of trust. This finding provided evidence that GSR can be used as a reliable tool for measuring trust in HRI. Gupta et al. [19] assessed human trust in a virtual assistant using physiological sensing in virtual reality during a cooperative information retrieval task. They employed heart rate variability, skin conductance, and facial electromyography to evaluate trust under four conditions: low cognitive load with low accurate assistance, low cognitive load with high accurate assistance, high cognitive load with low accurate assistance, and high cognitive load with high accurate assistance. The results showed that HRV was a reliable indicator of trust levels towards the virtual assistant, with participants displaying higher HRV when their level of trust was high. However, no significant differences were observed in EEG or GSR measures between different levels of trust. This suggests that further investigation into PBs for trust assessment is needed.

Machine learning to estimate Trust - Ajenaghughrure et al. [3] developed a predictive model for assessing user trust in a conversational user interface using PBs. In their study, participants engaged in an information search game, where they answered questions with the help of Google Assistant. The authors used heart rate variability and skin conductance to assess trust and distrust behaviours. Their findings showed increased trust when the system provided accurate assistance, as indicated by changes in physiological signals in their trust levels. They achieved a mean accuracy of 77.8%, demonstrating the model's effectiveness in evaluating trust through physiological signal analysis. Khalid et al. [24] examined subjective measures such as ability, benevolence, and integrity alongside PBs, including facial expressions, voice, and heart rate, to estimate trust levels in natural dialogues of real-world scenarios involving human-robot-human interactions. Heart rate variability and skin conductance were used as measures of trust. By employing a neuro-fuzzy neural network and integrating both objective and subjective indicators, their results showed a 67% accuracy in trust estimation, demonstrating that PBs can asses human trust during HRI. Hu et al. [22] developed a trust sensor model that maps PB measurements to human trust levels in real-time during collaboration with a machine to perform a simulated car driving task to reach a target location while avoiding obstacles. They used electroencephalography and GSR to capture PBs during trust and distrust. The study employed multiple classification methods, including binary classification techniques such as support vector machine and logistic regression. The results showed that EEG and GSR features were correlated with trust and were most significant when a human's trust level in an automated system was low. This demonstrates that PB measurements can be effectively used to sense trust in real-time. Lochner et al. [32] explored using PBs to calibrate human trust in automation. They collected GSR from participants during a semi-automated UAV operation task and measured trust and cognitive load based on their experience with the system. The authors used a decision tree algorithm to classify trust and achieved an accuracy of 80%. This demonstrates that PBs can effectively measure trust and cognitive load during human-automation interaction.

In summary, the described empirical studies have employed various PBs, such as GSR, eye fixation, and EEG, to assess trust in collaboration contexts in a single interaction. Besides, existing work on the use of machine learning to estimate trust has used limited PBs in combination and has also created a dataset based on one-off interaction. In addition, most of these studies had simulated environments rather than real HRI. Moreover, existing research has mainly focused on collaborative settings, with the only exception we identified being [40], which investigates trust in competitive contexts. In competitive contexts, factors influencing human trust, such as the robot's truthfulness in providing advice or information and the capability of robots to outperform humans, may not be present in collaborative settings. Furthermore, past research did not consider psychophysiological during repeated HRI. This paper aims to bridge these gaps by assessing human trust in robots in real-time using BPs within a competitive task during repeated HRI.

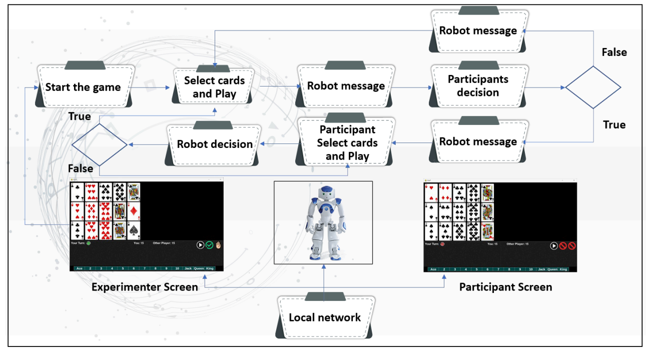

Figure 1: System description

3 STUDY DESIGN

This study aimed to investigate whether PBs can be collectively used to sense humans' trust in robots. To collect data, we involved participants playing with the Nao robot across four game sessions. We tested the following hypotheses:

H1 : Human PBs, including EDA, BVP, HR, SKT, BR and BD, will show significant differences between trusting and distrusting states during interactions with a robotic agent.[5].
H2 : Significant interaction effects between session (1, 2, 3, and 4) and the chosen PBs will be observed during HRI.
H3 : The classification algorithms will be able to classify levels of trust with potentially higher accuracy, demonstrating the potential of using PBs to sense trust in real-time.

H1 is based on the understanding that trust and distrust states can be reflected in an individual's PB responses, which are associated with emotional and cognitive factors [5]. **H2** acknowledges the potential influence of repeated interactions on PBs, as trust development is a dynamic process, and trust levels may change over time [36]. **H3** is supported by previous work in various domains, demonstrating the effectiveness of machine learning classifiers in analyzing and predicting human behaviour based on physiological measures [3, 6].

Ethics - Given the involvement of human participants in the study, we submitted an application to the university's ethics committee to guarantee the ethical integrity of our research. After review, the application was approved [160322/5031].

3.1 System description

The system, as illustrated in Figure 1, consists of two main components: 1) an interactive card game specifically designed to generate trust or distrust situations between participants and the robot, and 2) a semi-autonomous robot equipped to engage in playing the card game alongside human participants. By collecting data on PBs as humans showed trust and distrust behaviours, our primary aim was to explore participants' PB responses in real-time within the context of HRI.

The Bluff Game - We designed a Python-based interactive card game called *Bluff Game*, where participants competed against a robot. The game was two players game (participant vs robot) and comprised 52 cards, including four sets of aces, numbered cards 1 through 10, jacks, queens, and kings. Additionally, there were play and decision buttons (trust and distrust) for the participants to interact with during the game. At the beginning of the game, each player received 15 cards. The objective was to be the first to discard all the cards in hand. Participants took turns, and during each turn, a player chose a set of 2-4 cards to dispose of. Their opponent then decided whether to trust or distrust the player based on the truthfulness of the stated set of cards being discarded. For instance, if a player claimed to have a pair of kings, the opponent had to choose to trust or distrust the claim. If the opponent trusted the player, the cards were discarded, and the opponent took their turn. In this case, the opponent could not view the player's cards. However, if the opponent distrusted the player, the player had to show their cards. If the claim was accurate, the opponent received the cards, and they were added to the opponent's hand. If the claim was false, the cards were returned to the player, and the opponent took their turn. The game progressed in this manner until one player had successfully discarded all their cards. The game dynamically updated each player's card list after every turn. We designed this game to collect psychophysiological data in trust or distrust situations.

Interaction Scenarios - The Nao robot was programmed to interact verbally with participants throughout various game events. We utilized the Wizard of Oz method (WOz) to control the game, without informing the participants to prevent bias. The interaction consisted of three phases: welcoming and introducing the game, playing the game, and concluding the game.

Initially, the robot greeted the participant and provided a brief introduction: "Hello. I am a Nao robot. Today, we will be playing a card game against each other. Are you ready?" Participants played the game on four separate occasions with a time gap of 5 minutes between each session. During the second, third, and fourth sessions, the robot thanked the participants and reintroduced the game by stating: "Hello again. Thank you for playing. We are going to play another game. Are you ready?" and "Let us start" respectively.

As the game started, the Nao robot informed the participant that "the game starts now". The robot took the first turn. Following the game rules, the robot interacted with the participant during various game events as follows:

(1) When the robot selected its set of cards and declared them, e.g., "I selected three queens".
(2) When the participant trusted the robot, it responded with: "It is your turn".
(3) When the participant distrusted the robot, and the robot's card declaration was accurate, the robot stated: "I was telling the truth".
(4) When the participant distrusted the robot, and the robot's card declaration was incorrect, the robot stated: "You got me, and it is your turn".
(5) When the robot trusted the participant, it said: "I trust you, and it is my turn".
(6) When the robot distrusted the participant, it said: "I think you are bluffing". If the participant was truthful, the robot said: "Oh, I was wrong, and it is your turn now".
(7) If the robot distrusted the participant and the participant was incorrect, the robot stated: "Yes, I got you, and it is my turn now".

After each game, the robot congratulated the participant or wished them good luck for the next round. In case of a win, the robot said: "Congratulations! You win, thank you and see you in the next round". If the participant lost, the robot encouraged them by saying: "You just lost the game, good luck in the following rounds". In the final session, the robot added goodbye to its message, announcing the end of the experiment.

3.2 Participants

The study initially recruited 45 participants between 18 and 60 years old. However, we faced issues in data collection for 2 participants, resulting in a final sample size of 43 participants (M = 29.53 years, SD = 6.71). The gender distribution included 16 females, 26 males, and 1 participant who chose not to disclose their gender. We sent messages asking for participants using university mailing lists and placed flyers across the campus to recruit the participants. Participants registered for the study through the (*Calendly*) online scheduling platform.

To evaluate the participants' prior experience with robots, we divided them into four categories: high, medium, low, and no experience. Those who had previously controlled or built a robot were considered highly experienced, while participants who had used robots multiple times were deemed to have medium experience. Low experience referred to individuals who had only interacted with robots occasionally. The distribution of participants across these categories was as follows: 2 participants with high experience, 2 with medium experience, 24 with low experience, and 15 with no experience interacting with robots.

3.3 Setup and Materials

The study was conducted in two separate rooms, as illustrated in Figure 2. In Room 1, a laptop was placed on a table for participants to play the game, with the Nao robot positioned opposite them. Participants wore Pupil Invisible Eye Tracking Glasses and the to record PBs while seated in front of the robot. They also used a tablet to complete demographic information. This room was designed to maintain consistent environmental conditions during the study, including steady room temperature and consistent lighting. This was done to minimize potential influences on physiological measures, specifically BD, BR, and SKT. In Room 2, the experimenter monitored the interaction and remotely controlled the robot from a laptop.

The humanoid Nao robot, developed by Aldebaran Robotics, was used in this study. Nao is 58 cm tall and equipped with an inertial

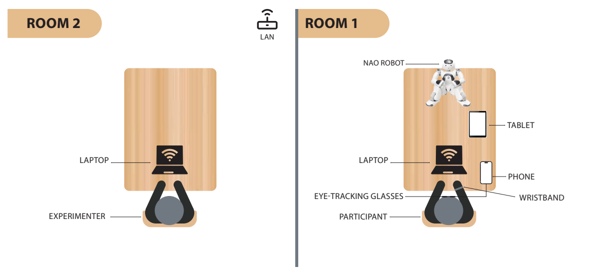

Figure 2: Experimental setup: The experimenter remotely controls the robot from one room (left) while the participant plays the game against the robot in a separate room (right).

sensor, two cameras, eyes with full-colour RGB LEDs, and various other sensors. It is designed for research and educational purposes, providing a versatile platform for human-robot interaction studies.

We captured physiological data by using the Empatica E4 Wristband and Pupil Invisible Eye Tracking Glasses. The E4 Wristband, developed by Empatica, is a wearable device that measures physiological signals such as heart rate, skin conductance, and temperature. The Pupil Invisible Eye Tracking Glasses, developed by Pupil Labs, are lightweight glasses equipped with high-resolution cameras and sensors that record eye movement and gaze behaviour, offering valuable insights into attention and cognitive processes during human-robot interactions.

3.4 Procedure

The study was carried out in the following stages:

(1) Participants were provided with an experiment information sheet, game instruction sheet, and consent form, which they were required to sign.
(2) Participants completed a demographics questionnaire, which included information about their experience with robots.
(3) Participants put on glasses and a wristband. The experimenter initiated data collection from these devices and then left the room.
(4) Participants performed a calibration by playing the game against the Nao robot while the experimenter remotely controlled the robot from another room.
(5) After each game, the experimenter returned to the room and stopped collecting physiological data, and asked the participant to complete the questionnaire to rate the robot during the game. However, the questionnaire is not included in the analysis of this paper, as it falls outside the scope of the described contributions.
(6) Steps 3, 4, and 5 were repeated in the other three sessions.
(7) Finally, participants were thanked for their participation and informed that they would receive a £10 Amazon voucher for their involvement in the study.

3.5 Measurements

3.5.1 Psychophysiological Measures. In this study, we collected the following real-time PBs during decision periods, from when the robot played cards until the player made a decision:

(1) The eye tracking recorded participants' eye BR and BD.
(2) The wristband measured participants' EDA, BVP, HR, and SKT.

Our choice of physiological signals prioritizes participant comfort and non-intrusiveness. Wearable devices capture chosen signals and have strong empirical evidence for trust measurement presented in the literature.

3.5.2 Behavioural Measures. We collected data on participants' in-game decisions, including their choices to trust or distrust the robot and each decision's start and end time. This information enabled us to assess the participants' PB responses during their decision.

3.6 Data Preparation

3.6.1 Behavioural Data Processing. The behavioural data collected during the game were processed to obtain relevant metrics for analysis:

(1) **Decision outcomes:** Trust and distrust decisions made by participants were logged and coded as binary variables (0 for distrust, 1 for trust) for subsequent statistical analyses.
(2) **Decision period:** Each participant's decision start and end time was logged to extract psychophysiological data in the given interval. The gameplay log was maintained to extract the decision's start and end times. The start decision time is when the robot plays cards, and the end time is when the player presses one of the decision buttons.

3.6.2 Psychophysiological Data Preprocessing. Before analyzing the psychophysiological data, we performed the following:

(1) **Noise and Artifact Removal:** The psychophysiological data underwent an essential preprocessing step where we applied a low-pass filter to remove noise and artifacts. This ensured that the subsequent analyses were based on clean and accurate signals.
(2) **Segmentation:** The psychophysiological data were recorded with timestamps to mark the start and end of the time during the session and to lead us to align them with the exact decision period that was logged in the game. The psychophysiological data were then segmented into epochs corresponding to the four rounds of the game for each participant.
(3) **Feature extraction:** Based on each participant's decision start and the end time logged during the game, the physiological samples were aggregated by computing the average value of EDA, BVP, HR, and SKT. In addition, the number of blinks and the average blink duration during the decision period were extracted. The average values were computed because the physiological data in the raw data was logged in each millisecond. As the raw physiological data was recorded in milliseconds and the decision period was in seconds, averaging the values per second (as a 1-second minimum decision period) enabled a more meaningful data comparison. This allowed us to understand better and analyse the changes in PBs during decision-making.
(4) **Dataset Generation:** To generate the dataset for the analysis and classification task in the study, we followed these steps:
(a) First, we computed the value for each PBs (EDA, BVP, HR, SKT, BR, and BD) during trust and distrust stated in all the sessions (1, 2, 3 and 4).
(b) Next, in each session (1, 2, 3 and 4), for each participant, we averaged the value for each PBs (EDA, BVP, HR, SKT, BR, and BD) in the trust and distrust states. For instance, if in session 1,

		Session 1				Session 2				Session 3				Session 4			
		Trust		Distrust		Trust		Distrust		Trust		Distrust		Trust		Distrust	
Feature (Unit)	N	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
EDA (μS)	43	0.69	1.13	0.87	1.21	0.93	2.46	0.92	2.34	0.88	2.22	0.76	2.09	0.91	2.33	0.81	2.28
BVP (μV)	43	0.31	1.32	0.37	1.57	-0.01	0.24	-0.004	0.15	-0.09	0.65	0.14	0.68	0.33	1.28	0.03	0.25
HR (bpm)	43	107.77	20.76	94.58	31.08	103.97	17.66	98.17	34.61	104.23	14.62	90.11	39.97	102.65	24.84	89.09	40.61
SKT ($^{\circ}C$)	43	27.80	1.22	25.39	7.21	28.19	1.30	26.21	7.35	28.48	1.27	24.66	10.11	28.53	1.35	24.57	10.08
BR (count)	43	1.56	1.62	1.11	1.25	1.53	3.62	1.79	5.37	1.38	3.63	2.00	1.82	1.75	3.78	1.74	3.48
BD (s)	43	177.20	149.85	181.85	164.68	197.11	114.75	151.66	145.32	191.83	99.69	183.95	180.80	191.14	102.08	204.90	176.27

Table 1: Mean (M) and Standard Deviation (SD) for the psychophysiological features of trust and distrust states during each session.

we had 4 trust, and 3 distrust states for a participant. In this case, we averaged the 4 and 3 values recorded in trust and distrust states. It resulted in 1 value for trust and distrust for a participant. We did this because the number of trust and distrust decisions were different among participants across all four game sessions.

(c) Later, all this resulted in a dataset containing 43 average values for each PB measure (EDA, BVP, HR, SKT, BR, and BD) corresponding to trust and distrust decisions in each session.

(d) Lastly, to form the dataset for all sessions, we merged the data of all the sessions into one.

By following these steps, we successfully generated a dataset suitable for analyzing trust and distrust in HRI using PBs. The dataset alongside codes can be accessed here. In the given link, the file named as "Data_sessions" represents the dataset for session 1, 2, 3 and 4 respectively, while, the file named as "Data_all" represents the all session data.

4 RESULTS

We present the results of the analyses regarding the differences in PBs between trust and distrust groups, the effects of sessions on these behaviours, and the accurate classification of trust levels in real-time during HRI using machine learning classifiers.

To test **H1** and **H2**, a repeated-measures ANOVA was conducted to determine whether there is an effect of the decision (trust vs distrust) and the interactive session (session 1, session 2, session 3, and session 4) on the physiological measures (EDA, BVP, HR, SKT, BR, and BD).

We found that there was a significant effect of decission on HR ($F(1, 84) = 11.652$, $p < .001$, $\eta_p^2 = .122$.) and SKT ($F(1, 84) = 13.473$, $p < .001$, $\eta_p^2 = .138$) scores. However, we did not see a significant effect of decision on BVP ($F(1, 84) = .001$,$p = .970$, $\eta_p^2 < .001$), EDA ($F(1, 84) = .001$, $p = .977$, $\eta_p^2 < .001$), BR ($F(1, 84) = .050$, $p = .823$, $\eta_p^2 = .001$) and BD ($F(1, 84) = .218$, $p = .642$, $\eta^2 = .003$) respectively.

Furthermore, We did not observe a significant interaction effect of session and decision (session * decision) on EDA [$F(3, 82) = .353$, $p = .787$, $\eta^2 = .013$], BVP [$F(3, 82) = 1.8$, $p = .154$, $\eta^2 = .062$], HR [$F(3, 82) = .376$, $p = .77$, $\eta^2 = .014$], SKT [$F(3, 82) = .517$, $p = .672$, $\eta^2 = .019$], BR [$F(3, 82) = .993$, $p = .906$, $\eta^2 = .007$], and BD [$F(3, 82) = .983$, $p = .405$, $\eta^2 = .035$] respectively.

We conducted a post-hoc Bonferroni test to assess whether HR, SKT, and other measures differed significantly between the trust and distrust classes within each session (sessions 1, 2, 3, and 4). The analysis confirmed that HR significantly differed between trust and distrust states in session 1 ($p < 0.03$), session 3 ($p < 0.01$), and session 4 ($p = 0.01$). Moreover, a slightly significant difference was found in session 2 ($p = 0.086$). In all of the sessions, we observed a significantly higher mean value of HR in the trust state as compared to the distrust state. Additionally, the analysis further showed that SKT significantly differed between trust and distrust decisions in session 1 ($p < 0.03$), session 3 ($p < 0.01$), and session 4 ($p < 0.01$). Furthermore, slightly significant difference was found in Session 2 ($p = 0.086$). In all of these sessions, we observed a significantly higher mean value of SKT in the trust state as compared to the distrust state. Intriguingly, a post-hoc test confirmed a slightly significant difference for BVP in session 3 ($p = 0.090$). The mean and Standard deviation for the PB features during trust and distrust across all sessions can be seen in Table 1 respectively.

To test **H3**, which was to investigate whether PBs can be used to classify trust, we used the structured approach proposed by Ahmad et al. [2]. We implemented five classifiers: Random Forest (RF), Logistic Regression (LR), Support Vector Machines (SVM), Decision Tree (DT), and AdaBoost (AB). The performance of these classifiers was evaluated using 5-fold cross-validation. We found that RF and DT achieved the best accuracies at 68.6%, and 62.2% respectively. The remaining classifiers performed above chance level (see Table 2).

We understand that the fundamental concept of RF is that it functions as an ensemble. RF builds models—trees—that produce class predictions. Based on the class that received the most votes, the model is forecasted. The low correlation between the trees is the secret to improved performance. We recognise that the DT's high predictive accuracy had an impact on the RF's performance since the ensemble of trees that the RF generated may have improved the classifier's predictive capabilities.

To further delve into the accuracy findings, in Table 3, we show the classification report for all the classifiers to highlight the F1 score for each class. This indicates that for RF, trust and distrust were predicted correctly 70% and 65% on the test data, showing a relatively higher accuracy of RF as compared to other classifiers.

4.1 Feature importance for Trust and Distrust

Using one feature at a time, we investigated which PBs in our dataset were predictive of either class (trust or distrust). We then computed the F1-score for each class. The goal was to determine how well each feature performed on its own in reliably classifying each class in the dataset. Due to the RF classifier's superior performance in

Classifier	Accuracy (%)				
	Session 1	Session 2	Session 3	Session 4	All session
SVM	58.9 +/- 0.11	53.8 +/- 0.09	55.5 +/- 0.04	50.5 +/- 0.11	**53.5 +/- 0.02**
RF	68.2 +/- 0.10	46.8 +/- 0.08	69.1 +/- 0.085	60.2 +/- 0.11	**68.6 +/- 0.04**
LR	55.5 +/- 0.07	49.4 +/- 0.08	49.9 +/- 0.07	46.4 +/- 0.04	**50.5 +/- 0.05**
DT	63.4 +/- 0.09	64.2 +/- 0.07	64.9 +/- 0.05	59.3 +/- 0.10	**62.2 +/- 0.02**
AB	63.4 +/- 0.05	57.8 +/- 0.11	67.6 +/- 0.06	54.0 +/- 0.08	**53.6 +/- 0.05**

Table 2: Classifier Accuracies for Psychophysiological Behaviors in Trust Classification.

Decision	Classifier	F1-score
Trust	AB	0.534
	RF	**0.708**
	DT	0.692
	SVM	0.644
	LR	0.572
Distrust	AB	0.536
	RF	**0.658**
	DT	0.492
	SVM	0.304
	LR	0.402

Table 3: F1-scores for the five classifiers to predict human's trust and distrust levels. Bold RF is the classifier that achieves the highest accuracy.

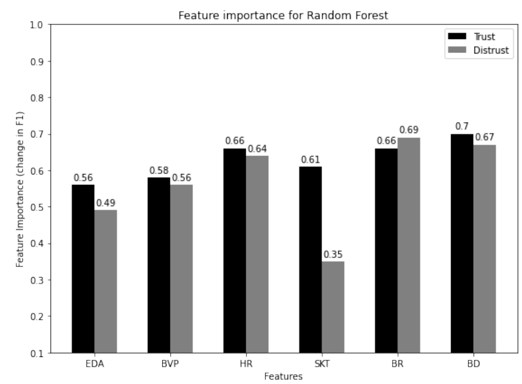

Figure 3: Feature importance for the RF classifier based on the F1-scores for each trust class. The x-axis shows all the PBs while the y-axis shows the accuracies achieved by each PB as one feature to predict the class of trust.

predicting trust or distrust, we only provide the feature importance for trust and distrust for this classifier. In Figure 3, we show the best performing features for the RF classifier. HR, BR, BD, and SKT were the best-performing features for trust and distrust classes. We understand this finding through the lens of the mean and SD values shown in Table 4. We observed mean differences between the trust and distrust behaviours for all four measures (HR, BR, BD, and SKT). It also prompted us to conduct a correlation analysis. We found that all the four measures were significantly ($p < 0.05$) and positively correlated. Consequently, this highlights the reasons for the feature importance findings.

Similarly, as seen in Table 4, both EDA and BVP mean values did not differ for both trust and distrust case resulting in EDA and BVP as the least important features for the RF to predict the trust classes. Further correlation analysis also confirmed that both variables were significantly ($p < 0.05$) and positively correlated.

Feature	N	Trust		Distrust	
		M	SD	M	SD
EDA	43	0.86	2.09	0.85	2.02
BVP	43	0.13	0.99	0.14	0.87
HR	43	104.60	17.66	92.99	36.64
SKT	43	28.25	1.31	25.21	8.75
BR	43	1.55	2.89	1.46	3.37
BD	43	189.32	117.49	180.59	166.94

Table 4: Mean (M) and Standard Deviation (SD) for the psychophysiological features of trust (999 cases) and distrust (480 cases) during all sessions.

5 DISCUSSION

This study investigated whether PBs can be collectively used to sense human trust in robots in real-time during HRI. In this section, we discuss whether the hypotheses were accepted or rejected in the light of the findings.

H1 hypothesizes a significant difference in human PB responses, such as EDA, BVP, HR, SKT, BR, and BD, between trust and distrust states during HRI. The findings of the study showed that HR and SKT were significantly different between trust and distrust groups across all the sessions. This finding is consistent with previous research that identified HR and SKT as important features for assessing human trust in robots across diverse HRI settings [5, 19, 24, 37]. HR and SKT are considered valuable indices of sympathetic arousal changes that can be measured during emotional arousal and cognitive effort [7]. The notable difference in HR and SKT between trust and distrust groups indicates that participants in the trust group may have experienced increased emotional arousal and cognitive effort due to the risks associated with the number of cards remaining in the game. Trusting others in such contexts might require heightened vulnerability and emotion, leading to elevated arousal and cognitive processing [3].

On the other hand, the other PB responses (EDA, BVP, BR, and BD) did not show significant differences between trust and distrust groups. Changes in EDA, BVP, and other PBs correlate with anxiety; however, such conditions may not have been observed during the gameplay. Ganglbauer et al. [18] suggested that assessing trust through psychophysiological signals becomes challenging during natural user interactions. We speculate that participants were relaxed, and no pressure elements, such as time constraints, were part of the gameplay. Furthermore, participants' prior robot interaction experiences, with most having low or none. This range may have contributed to the variability in physiological responses and trust rating, possibly impacting our findings. Future studies should further explore this variable's role in trust during HR. Another factor we acknowledge is the potential impact of mental load, which may have affected participants' physiological responses, particularly in PBs like EDA/GSR that have been significant in previous studies[26]. We conjecture that there was no significant change for BR and BD because participants interacted naturally and had low focus and attention levels, as they are important factors affecting eye blinking [39]. The insignificant differences could also be attributed to individual differences, as existing research suggests that individuals may exhibit distinct physiological responses to the same emotional state [12]. Furthermore, the game data epoch time window could have influenced these results. This window was based on participants' response time to make decisions. In this study, the average decision time was 4 seconds. Employing a longer epoch time window in contexts that are not time-sensitive might improve performance for these other PBs [4].

Although HR and SKT were significant features in our study, the factors affecting them may vary across individuals and settings [11]. Thus, future research should consider HR and SKT as trust indicators and investigate them across diverse individuals and scenarios to ensure the validity of these measures for sensing trust. We believe this will make these more generalisable measures. In summary, the hypothesis **H1** was *partially accepted* as we did not find significant differences for all the PBs.

H2 hypothesized an interaction effect (session and decision to trust or distrust) on PBs. Our results *did not confirm* this hypothesis, as we did not find a significant interaction effect of session and decision (session * decision) on all PB features. We understand that this could be due to the consistent behaviour of the game across the four sessions. We assumed that participants' experience with the interaction partner (robot) will impact the PBs across the four sessions. However, the findings did not support the assumptions. We encourage the community to investigate an individual experience with the robot in the context of trust and PBs during repeated HRI. It may lead to intriguing insights to further enhance our knowledge of sensing trust in real-time using PBs. Furthermore, understanding these factors can contribute to the development of effective trust measures for long-term HRI. Besides, we will consider mitigating this in the context of our experimental setup in the future.

H3 suggested that human trust levels in HRI can be accurately classified using PB data. The results showed that RF classifiers provided the best accuracy in trust level classification, with SKT, HR, BR and BD features crucial for predicting human trust in robots during HRI. We link this feature importance finding to its connection with emotional arousal, cognitive effort, and rapid physiological changes typically occurring in response to trust-related decisions in game contexts [30].

The findings build on the existing literature demonstrating that PB features can predict trust in human-machine interactions [22]. Comparing the findings on classification accuracy with existing results, we see that Khalid et al. [24] used Heart rate variability and skin conductance to sense trust and achieved an accuracy of 67% using a neuro-fuzzy neural network. The findings showed an improved accuracy of 68.6% using an RF classifier to sense trust by using a range of PBs. Other works that have achieved a higher accuracy as compared to our findings for instance (Ajenaghughrure et al. [3], & Lochner et al. [32]) relied on two PBs. We argue that it is critical to use multiple PBs as physiological behaviours tend to be task dependent or sensitive to environments [2]. In addition, we note that past work applied a decision tree with a standard 70 and 30 split for train and test data which may not be a suitable and rigorous strategy to train on a dataset [32, 49]. Moreover, the findings were based on a small sample size (10 participants) [3]. Lastly, most of the settings used in past work had tasks in simulation, consequently highlighting the value of the described work in this paper.

We employed widely-used machine learning classifiers, such as RF, LR, SVM, DT, and AB, due to their adaptability, robustness, diverse learning mechanisms, and ease of implementation and optimization [34]. The study achieved relatively higher accuracy in predicting trust levels, which can be attributed to the use of multiple PBs that changed behaviour differently in combination across the trust and distrust situations. We speculated to have achieved a higher accuracy but, we understand that our task may have offered an overall low vulnerability [31] to have shown difference among all the PBs. In summary, features with non-significant differences did show a trajectory and indicated that real-time trust calibration in HRI should consider multiple features, which aligns with the literature [5].

6 CONCLUSION & FUTURE WORK

In this paper, we investigated whether different psychophysiological behaviours (PBs) differ significantly when humans trust or distrust a robot during a repeated Human-Robot Interaction (HRI). In addition, we investigate whether PBs can accurately predict human's trust or distrust in the robotic partner during HRI. In a game-based repeated HRI, we examined the differences in PBs between trust and distrust states, the effects of multiple sessions on these behaviours. We also explored the potential of machine learning classifiers in predicting trust during HRI. The findings confirmed that PBs such as HR and SKT differ in trust and distrust behaviours. It indicated that the use of multiple PBs collectively can enable sensing of human trust in robots in real-time during HRI. It further suggested we need to collect data on PBs across various settings to establish the validity of these measures in sensing trust. Doing so will enable the use of PBs to sense trust in real-time. Such measures of trust can be employed to develop robots that can adapt to user trust in real-time. Besides, we share our dataset with the research community to promote further advancements in knowledge.

In future work, we aim to explore additional PB features from multiple relevant body organs in a different HRI setting such as long-term collaborative HRI. Our study and future intention hold the potential to facilitate the development of adaptive robotic systems based on real-time trust measurements, ultimately leading to more effective HRI.

REFERENCES

[1] Hussein A Abbass, Jason Scholz, and Darryn J Reid. 2018. *Foundations of Trusted Autonomy (Studies in Systems, Decision and Control).* Springer Nature, Switzerland. 137 pages.

[2] Muneeb Imtiaz Ahmad, Ingo Keller, David A Robb, and Katrin S Lohan. 2020. A framework to estimate cognitive load using physiological data. *Personal and Ubiquitous Computing* (2020), 1–15.

[3] Ighoyota Ben Ajenaghughrure, Sonia C Sousa, Ilkka Johannes Kosunen, and David Lamas. 2019. Predictive model to assess user trust: a psycho-physiological approach. In *Proceedings of the 10th Indian conference on human-computer interaction.* 1–10.

[4] Ighoyota Ben Ajenaghughrure, Sónia Cláudia Da Costa Sousa, and David Lamas. 2021. Psychophysiological Modelling of Trust in Technology: Comparative Analysis of Psychophysiological Signals.. In *VISIGRAPP (2: HUCAPP).* 161–173.

[5] Ighoyota Ben Ajenaghughrure, Sonia Da Costa Sousa, and David Lamas. 2020. Measuring trust with psychophysiological signals: a systematic mapping study of approaches used. *Multimodal Technologies and Interaction* 4, 3 (2020), 63.

[6] Kumar Akash, Wan-Lin Hu, Neera Jain, and Tahira Reid. 2018. A classification model for sensing human trust in machines using EEG and GSR. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 4 (2018), 1–20.

[7] Veronika Alexander, Collin Blinder, and Paul J Zak. 2018. Why trust an algorithm? Performance, cognition, and neurophysiology. *Computers in Human Behavior* 89 (2018), 279–288.

[8] Abdullah Alzahrani, Simon Robinson, and Muneeb Ahmad. 2022. Exploring Factors Affecting User Trust Across Different Human-Robot Interaction Settings and Cultures. In *Proceedings of the 10th International Conference on Human-Agent Interaction.* 123–131.

[9] Anthony L Baker, Elizabeth K Phillips, Daniel Ullman, and Joseph R Keebler. 2018. Toward an understanding of trust repair in human-robot interaction: Current research and future directions. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 4 (2018), 1–30.

[10] Marsha E Bates and Jennifer F Buckman. 2013. Integrating body and brain systems in addiction neuroscience. *Biological research on addiction: Comprehensive addictive behaviors and disorders* 2 (2013), 187–196.

[11] Cindy L Bethel, Kristen Salomon, Robin R Murphy, and Jennifer L Burke. 2007. Survey of psychophysiology measurements applied to human-robot interaction. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication.* IEEE, 732–737.

[12] Lorenzo Cominelli, Francesco Feri, Roberto Garofalo, Caterina Giannetti, Miguel A Meléndez-Jiménez, Alberto Greco, Mimma Nardelli, Enzo Pasquale Scilingo, and Oliver Kirchkamp. 2021. Promises and trust in human–robot interaction. *Scientific reports* 11, 1 (2021), 1–14.

[13] Ana Cristina Costa, C Ashley Fulmer, and Neil R Anderson. 2018. Trust in work teams: An integrative review, multilevel model, and future directions. *Journal of Organizational Behavior* 39, 2 (2018), 169–184.

[14] Munjal Desai. 2012. *Modeling trust to improve human-robot interaction.* Ph. D. Dissertation. University of Massachusetts Lowell.

[15] Ahmet Cengizhan Dirican and Mehmet Göktürk. 2011. Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Computer Science* 3 (2011), 1361–1367.

[16] Rebecca Flook, Anas Shrinah, Luc Wijnen, Kerstin Eder, Chris Melhuish, and Séverin Lemaignan. 2019. On the impact of different types of errors on trust in human-robot interaction: Are laboratory-based HRI experiments trustworthy? *Interaction Studies* 20, 3 (2019), 455–486.

[17] Amos Freedy, Ewart DeVisser, Gershon Weltman, and Nicole Coeyman. 2007. Measurement of trust in human-robot collaboration. In *2007 International symposium on collaborative technologies and systems.* Ieee, 106–114.

[18] Eva Ganglbauer, Johann Schrammel, Stephanie Deutsch, and Manfred Tscheligi. 2009. Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. In *Workshop on user experience evaluation methods in product development.* Citeseer.

[19] Kunal Gupta, Ryo Hajika, Yun Suen Pai, Andreas Duenser, Martin Lochner, and Mark Billinghurst. 2020. Measuring human trust in a virtual assistant using physiological sensing in virtual reality. In *2020 IEEE Conference on virtual reality and 3D user interfaces (VR).* IEEE, 756–765.

[20] Peter A Hancock, Theresa T Kessler, Alexandra D Kaplan, John C Brill, and James L Szalma. 2021. Evolving trust in robots: specification through sequential and comparative meta-analyses. *Human factors* 63, 7 (2021), 1196–1229.

[21] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.

[22] Wan-Lin Hu, Kumar Akash, Neera Jain, and Tahira Reid. 2016. Real-time sensing of trust in human-machine interactions. *IFAC-PapersOnLine* 49, 32 (2016), 48–53.

[23] Poornima Kaniarasu, Aaron Steinfeld, Munjal Desai, and Holly Yanco. 2013. Robot confidence and trust alignment. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 155–156.

[24] Halimahtun M Khalid, Liew Wei Shiung, Parham Nooralishahi, Zeeshan Rasool, Martin G Helander, Loo Chu Kiong, and Chin Ai-vyrn. 2016. Exploring psycho-physiological correlates to trust: Implications for human-robot-human interaction. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 697–701.

[25] Zahra Rezaei Khavas. 2021. A review on trust in human-robot interaction. *arXiv preprint arXiv:2105.10045* (2021).

[26] Ahmad Khawaji, Jianlong Zhou, Fang Chen, and Nadine Marcus. 2015. Using galvanic skin response (GSR) to measure trust and cognitive load in the text-chat environment. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems.* 1989–1994.

[27] J Matias Kivikangas, Guillaume Chanel, Ben Cowley, Inger Ekman, Mikko Salminen, Simo Järvelä, and Niklas Ravaja. 2011. A review of the use of psychophysiological methods in game research. *journal of gaming & virtual worlds* 3, 3 (2011), 181–199.

[28] Alap Kshirsagar, Bnaya Dreyfuss, Guy Ishai, Ori Heffetz, and Guy Hoffman. 2019. Monetary-incentive competition between humans and robots: Experimental results. In *2019 14th acm/ieee international conference on human-robot interaction (hri).* IEEE, 95–103.

[29] Theresa Law and Matthias Scheutz. 2021. Trust: Recent concepts and evaluations in human-robot interaction. *Trust in human-robot interaction* (2021), 27–57.

[30] Jennifer S Lerner, Ye Li, Piercarlo Valdesolo, and Karim S Kassam. 2015. Emotion and decision making. *Annual review of psychology* 66 (2015), 799–823.

[31] Michael Lewis, Katia Sycara, and Phillip Walker. 2018. The role of trust in human-robot interaction. *Foundations of trusted autonomy* (2018), 135–159.

[32] Martin Lochner, Andreas Duenser, and Shouvojit Sarker. 2019. Trust and Cognitive Load in semi-automated UAV operation. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction.* 437–441.

[33] Yidu Lu and Nadine Sarter. 2019. Eye tracking: a process-oriented method for inferring trust in automation as a function of priming and system reliability. *IEEE Transactions on Human-Machine Systems* 49, 6 (2019), 560–568.

[34] Batta Mahesh. 2020. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]* 9 (2020), 381–386.

[35] Bertram F Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in human-robot interaction.* Elsevier, 3–25.

[36] Linda Miller, Johannes Kraus, Franziska Babel, and Martin Baumann. 2021. More than a feeling—interrelation of trust layers in human-robot interaction and the role of user dispositions and state anxiety. *Frontiers in Psychology* 12 (2021), 592711.

[37] Saeid Nahavandi. 2019. Trust in autonomous systems-iTrust lab: Future directions for analysis of trust with autonomous systems. *IEEE Systems, Man, and Cybernetics Magazine* 5, 3 (2019), 52–59.

[38] Michael Novitzky, Paul Robinette, Michael R Benjamin, Danielle K Gleason, Caileigh Fitzgerald, and Henrik Schmidt. 2018. Preliminary interactions of human-robot trust, cognitive load, and robot intelligence levels in a competitive game. In *Companion of the 2018 ACM/IEEE international conference on human-robot interaction.* 203–204.

[39] Jihoon Oh, So-Yeong Jeong, and Jaeseung Jeong. 2012. The timing and temporal patterns of eye blinking are dynamically modulated by attention. *Human movement science* 31, 6 (2012), 1353–1365.

[40] Yosef S Razin and Karen M Feigh. 2021. Committing to interdependence: Implications from game theory for human–robot trust. *Paladyn, Journal of Behavioral Robotics* 12, 1 (2021), 481–502.

[41] René Riedl and Andrija Javor. 2012. The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics* 5, 2 (2012), 63.

[42] Julian B Rotter. 1971. Generalized expectancies for interpersonal trust. *American psychologist* 26, 5 (1971), 443.

[43] Tracy Sanders, Alexandra Kaplan, Ryan Koch, Michael Schwartz, and Peter A Hancock. 2019. The relationship between trust and use choice in human-robot interaction. *Human factors* 61, 4 (2019), 614–626.

[44] Kristin E Schaefer. 2016. Measuring trust in human robot interactions: Development of the "trust perception scale-HRI". In *Robust intelligence and trust in autonomous systems.* Springer, 191–218.

[45] Josh Siegel and Daniel Morris. 2020. Robotics, Automation, and the Future of Sports. *21st Century Sports: How Technologies Will Change Sports in the Digital Age* (2020), 53–72.

[46] Ana Correia Simões, Ana Pinto, Joana Santos, Sofia Pinheiro, and David Romero. 2022. Designing human-robot collaboration (HRC) workspaces in industrial settings: A systematic literature review. *Journal of Manufacturing Systems* 62 (2022), 28–43.

[47] Robert Morris Stern, William J Ray, and Karen S Quigley. 2001. *Psychophysiological recording.* Oxford University Press, USA.

[48] Tetsuya Tanioka, Rozzano C Locsin, Feni Betriana, Yoshihiro Kai, Kyoko Osaka, Elizabeth Baua, and Savina Schoenhofer. 2021. Intentional observational clinical research design: Innovative design for complex clinical research using advanced technology. *International Journal of Environmental Research and Public Health* 18,

21 (2021), 11184.
[49] Borislava Vrigazova. 2021. The proportion for splitting data into training and test set for the bootstrap in classification problems. *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy* 12, 1 (2021), 228–242.
[50] Rosemarie E Yagoda and Douglas J Gillan. 2012. You want me to trust a ROBOT? The development of a human–robot interaction trust scale. *International Journal of Social Robotics* 4 (2012), 235–248.