Detecting Deception in Natural Environments Using Incremental Transfer Learning

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

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

Sunbul M. Ahmad
Cardiff University
Cardiff, UK
ahmads31@cardiff.ac.uk

ABSTRACT

Existing work on detecting deception has mainly relied on collecting datasets evolving from contrived user interactions. We argue that naturally occurring deception behaviours can inform more reliable datasets and improve detection rates. Therefore, in this paper, we discuss the findings of two experiments which enabled participants to freely and naturally engage in deceptive and truthful behaviours in a game environment. We collected physiological and oculomotor behaviour (PB, & OB) data including electrodermal activity, blood volume pulse, heart rate, skin temperature, blinking rate, and blinking duration during the deceptive and truthful states. We investigate the changes in both PB and OB across repeated interactions and explore the potential of incremental transfer learning in detecting deception. We found significant differences in electrodermal activity, and skin temperature between deception and non-deception groups in both studies. The incremental transfer learning method with a logistic regression classifier detected deception with 80% accuracy, outperforming previous research. These results highlight the importance of collecting data from multiple sources and promote the use of incremental transfer learning to accurately detect deception in real time.

CCS CONCEPTS

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

KEYWORDS

Deception, Measurement, Dataset, Physiological and oculomotor behaviours, Bluff Game, Human-Robot Game Interaction

ACM Reference Format:
Muneeb Imtiaz Ahmad, Abdullah Alzahrani, and Sunbul M. Ahmad. 2024. Detecting Deception in Natural Environments Using Incremental Transfer Learning. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '24), November 04–08, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3678957.3685702

1 INTRODUCTION

Deception can be defined as an agent *acting or speaking to induce a false belief in a target or victim* [28]. Deception detection has been widely studied across many fields such as psychology, social science, criminology and neuroscience due to its pervasive nature in human interactions and implication in many social contexts [21].

The traditional approach to detecting deception has mainly considered polygraph tests that extract physiological measurements such as heart rate, respiration rate, skin conductance, and skin temperature [19]. Researchers have raised concern about the use of these measures to reliably and non-invasively detect deception. Further, findings suggest that trained individuals can trick the system resulting in bias and error [27]. However, we argue that due to advancements in wearable technology design (see [17, 55]), improved machine learning classification methods for deception detection [18], and combining multiple psychophysiological indicators or using a hybrid approach [6], current devices can reliably and non-invasively collect various psychophysiological indicators in real-time to support deception detection. For instance, together with physiological measurements, studies have shown that blink duration and count can be useful for the detection of deception [22, 40, 43]. Recent work by George et al. [22] has shown that participants' blink duration and count were significantly higher in the deception condition. Similarly, Marchak [40] has found that training machine learning classifiers on blink rate and response time can help identify deceptive and non-deceptive behaviours. Consequently, we consider both oculomotor and physiological behaviours (OB, & PB) for detecting deception.

While current work has considered hybrid approaches, the work on the collection of data to detect deception has mostly considered rather unnatural or artificially created tasks such as truthfully and quickly answering general questions, interview questions of different categories, or analysing video interactions [6, 8, 24, 47, 49]. Consequently, the existing dataset resulting from these interactions lacks collecting data on deception behaviour in a natural way. Similarly, recent findings have highlighted the limitations of relying on laboratory-created lies to study human lie detection and have called for researching natural means to study human deception detection [53]. Alaskar et al. [6] conducted a comprehensive review on machine-intelligence techniques for detecting deception and concluded that available datasets are not diverse as they are collected in simulated environments which are not realistic to train the deceptive and truthful behaviour, have been based on limited participants, and, have used static questions as a task.

Another aspect lies in investigating the changes in deception behaviour during repeated interaction and understanding how OBs and PBs of deception change over time. Further, how investigating changes in deception behaviour over time can enrich datasets. Existing research on collecting datasets has also used one-off interactions, thus, the change in deception behaviour during repeated

ICMI '24, November 04–08, 2024, San Jose, Costa Rica

ACM ISBN 979-8-4007-0462-8/24/11
https://doi.org/10.1145/3678957.3685702

interaction has also not been studied [6]. We see evidence from deception research on repeated interviews that liars are not less consistent than truth tellers [41]. Thus, it becomes an important question to understand how deception behaviour naturally evolves. Furthermore, most existing datasets have been collected under a single context. Recent research has shown that the context in which deception occurs can influence the likelihood and type of deception [59]. Therefore, applying incremental transfer learning to detect deception across multiple contexts is important. However, due to most datasets being collected within a single context and to the best of our knowledge, the role of incremental transfer learning to detect deception has not been explored yet.

Considering these aspects, this paper uniquely employs the GAME (*Game As a Method of eliciting Emotions*) paradigm proposed by Ahmad et al. [5], Shahid [48] to collect data on naturally occurring deception behaviour. We consider collaborative and competitive interaction game contexts and investigate both PBs and OBs during naturally occurring deception behaviour. Lastly, as social robots have begun to take on different social yet professional roles such as an interviewer [4, 30], or a teacher [36], or a therapist [11] or a detective [24], we consider the Human-robot game interaction context and foresee a future where robots detect deception in real-time. The paper investigates the following research questions (RQs):

- RQ1: How do naturally occuring human OBs and PBs differ between deception and truthful states during interactions with a robotic agent?
- RQ2: How do naturally occuring human OBs and PBs evolve as individuals gain experience during repeated interactions?
- RQ3: Do collecting data on OBs and PBs during truthful and deceptive states naturally improve the deception detection accuracy?
- RQ4: Which naturally occuring human OBs and PBs are predictive of deception and truthful behaviours?

To investigate these RQs, we conducted two experiments that tasked participants to play a game involving showing natural instances of both deception and truthful behaviour to and with a NAO robot. We recorded both OBs and PBs, including electrodermal activity (EDA), blood volume pulse (BVP), heart rate (HR), skin temperature (SKT), blinking rate (BR), and blinking duration (BD) to detect humans' deception in real-time. The novel contributions of this paper are as follows:

- Using an incremental transfer learning technique, we have shown that data gathered on (OB) and (PB) in natural interactions can effectively differentiate between deceptive and truthful states with an accuracy rate of 80%. This method surpasses most existing techniques and highlights the potential for reliable deception detection and development of adaptive robotic systems.
- We show that repeated exposure to the same deceptive scenarios can lead to habituation and emotional desensitisation, resulting in fewer physiological changes and consistent behaviours.
- We share the study materials and evolving datasets with the community to advance knowledge on deception which can be found *here*.

2 BACKGROUND

Theoretical Knowledge on Deception - Numerous theories have been proposed to explain why individuals engage in deceptive behaviours [9, 62]. We focus on the four-factor theory of deception and the interpersonal theory of deception, both suggesting that liars exhibit increased general arousal, emotional load, cognitive load, and attempts at control and impression management to appear honest. These conditions can lead to changes in verbal and nonverbal behaviours, such as increased blinking and pupil dilation, heightened voice pitch, speech errors, pausing, and other speech hesitations, ultimately affecting physiological behaviours like SKT, EDA, and HR. Additionally, we consider the truth default [38] and interpersonal theory, which justify lying for reasons such as goal attainment, where honesty is seen as counterproductive. Thus, using games that naturally create situations requiring deception to win is considered.

Methods for Detecting Deception - Broadly, three methods have been applied to detect deception: 1) psychological, 2) professional, and 3) computational [6]. Psychological approaches examine the relationship between nonverbal and verbal behaviours and the act of lying, including physiological changes such as increased pupil size and higher-pitched voices [14]. Studies have also identified unique hand movements and speech-related gestures as indicators of deception, using measures of facial expression smoothness and asymmetry to connect them to deceptive acts [1].

Physiological techniques, including polygraph tests and fMRI, are historically considered limited due to their need for complex setups and skilled operators [50]. However, recent advancements in hardware design and technology may challenge these limitations [17, 55]. Lastly, significant advancements in data mining and machine learning algorithms have led to the rise of computational methods [6]. These methods analyse micro-expressions, voice stress, heart rate, skin activity, and breathing patterns to detect deception. Interdisciplinary research continues to refine these methods, aiming to develop reliable tools for various settings. Despite advancements, existing work is often limited to datasets collected under a single context and does not use increment transfer learning to detect deception [6]. This paper addresses this gap by collecting datasets from two experiments and applying transfer learning to investigate the accuracy of deception detection.

Datasets on Detecting Deception - Various methods have been used to create datasets for detecting deception, encompassing three primary categories: verbal, non-verbal, and hybrid approaches. In verbal methods, researchers have leveraged features such as text sequences [31], linguistic attributes [13], sparse elements [60], and acoustic characteristics [57, 61]. Non-verbal methods involve features such as EEG signals [7], facial expressions [33, 54], and micro eye movements [34]. Hybrid methods, combine different feature categories to create comprehensive datasets. These hybrid datasets encompass a range of features, including visual and vocal characteristics [32], MFCCs (Mel-Frequency Cepstral Coefficients), and statistical measures of speed, pitch, and loudness [23], as well as facial expressions and body motions extracted from videos [15], and integrated data from videos, audios, EEG, and gaze tracking [25, 35]. Some hybrid datasets combine EEG, video, audio, and gaze data, while others integrate gaze and speech features [20, 42].

In summary, many studies resulting in datasets to detect deception demonstrate low accuracy rates, often below 70% [6]. Deception detection models are frequently trained on datasets lacking diversity and realism, collected in simulated environments that may not accurately reflect real-world deceptive behaviours. These datasets typically involve a limited number of subjects, and participants are often asked static questions, restricting response authenticity. Additionally, there are conflicting findings on whether verbal or visual cues are more important in detecting deception. Psychological studies suggest verbal cues are primary, while professional investigators and deep learning models emphasize visual cues. These inconsistencies highlight the challenges in developing effective deception detection methods. Therefore, we create a dataset based on naturally occurring deceptive behaviour and use a hybrid approach combining physiological and oculomotor behaviour from a large number of participants. We apply an incremental transfer learning algorithm to detect deception [12].

Game as a Method to Detect Deception - Researchers have used various games to study deception behaviour, including Cheap-Talk Games signalling game [45], iterated prisoner's dilemma [52], and the Mafia party game [13]. Signalling games involve strategic communication where one party sends a message, and the other decides if it is truthful or deceptive [45]. The iterated prisoner's dilemma allows players to repeatedly choose cooperation or defection, with opportunities for deception to gain higher payoffs [52]. The Mafia party game involves players with secret roles lying to conceal their identity and objectives [13]. The Mafia party game has been widely used to create video-based datasets for detecting deceptive behaviours [13, 29]. Additionally, Bag of Lies is a game-based approach where participants describe images honestly or deceptively [25].

However, researchers have identified several shortcomings in the Mafia party game, including being context-specific [29], imposing a high cognitive load on players that may impact their ability to deceive naturally [29, 50], and offering low classification accuracy in identifying deceptive players in the real world [25]. In this paper, we introduce a simple card game known as the Bluff game [56], inspired by Ahmad and Alzahrani [2], to study the truthfulness of the robot. The game allows players to depict truthful and deceptive behaviours without a high cognitive load, making it ideal for studying deception behaviours. The cognitive load in a game can vary depending on factors such as the number of players, game complexity, and player familiarity [44]. To minimise impact, the game is played by two players with minimal complexity. Cards are managed and distributed to each player, with multiple sessions conducted to ensure familiarity. In summary, this paper employs the Bluff game to create a dataset to detect deception, testing players' abilities to lie naturally and convincingly.

3 METHODOLOGY

We conducted two studies that involved participants to naturally engage in deceptive and truthful behaviours in a fun and entertaining manner while playing the bluff game. The two studies differed in context. In study 1, we enabled participants to play against the robot thus presenting a competitive context while in study 2, participants played the game where the robot acted as an advisor, hence presenting a cooperative context. In both studies, we only focused on the instances where participants were engaged in displaying deceptive and truthful behaviours. Such interactions were not informed or mediated by the the role of robot in both contexts. We investigate the following hypotheses:

- **[H1]**: Human PBs and OBs, including EDA, BVP, HR, SKT, BR, and BD, will show significant differences when participants are engaging in deceptive versus truthful behaviours during interactions with a robotic agent in both competitive (**H1a**) and cooperative (**H1b**) settings.
- **[H2]**: Significant interaction effects between the session number (1, 2, 3, and 4) and the chosen PBs and OBs will be observed, indicating differences in PBs and OBs responses to deceptive and truthful behaviours in both competitive (**H1a**) and cooperative (**H1b**) settings.
- **[H3]**: Classification algorithms will be able to classify instances of deception with potentially high accuracy, demonstrating the feasibility of using PBs and OBs for real-time detection of deception in different settings.

Study 1 investigates H1a, H2a, and H3 while study 2 investigates H1b, H2b and H3 respectively. In essence, H1 is based on the existing findings that both PBs and OBs tend to differ during deceptive and truthful acts of humans [50]. H2 is based on the existing research that familiarity with the situation can significantly influence humans' deceptive behaviour [39]. Lastly, H3 is based on the finding suggesting that data collected through humans naturally occurring deceptive and truthful behaviour can improve the reliability and detection rates [6].

Ethics - We submitted an application to the university's ethics committee to ensure the ethical integrity of our research involving human participants. After review, the application was approved [160322/5031].

3.1 The Game

We have created a card game called the "Bluff Game" using the Python programming language. The game can be played in two ways: a human player can compete against a robot (**study 1**), or a human and a robot can team up against an adversary (**study 2**). The game involves a deck of 52 cards with four sets of each number from 1 to 10 and the face cards (jack, queen, and king). The game interface has play and decision buttons, which make it easy for players and the game to interact seamlessly. At the start of the game, each player receives 15 cards, and to win the game, the goal is to eliminate all the cards before the opponent. The game is turn-based, and at each turn, a player must choose a set of 2-4 cards to discard. This requires the player to decide whether to deceive or be truthful about the cards in their hand. The opponent then has to decide whether to believe the player is telling the truth or lying. If the opponent believes that the player is truthful, the cards are discarded and remain unseen. The opponent then takes the next turn, and the game continues. If the opponent does not believe that the player is truthful, the discarded cards are revealed. If the player is found to be truthful, the opponent loses the round, and the opponent receives the player cards. If the player is found to be deceptive, the player must take back the cards, and the game continues. The game ends when one player has discarded all their cards. The list of each player's cards is updated dynamically after every turn.

3.2 Study 1

3.2.1 Interaction Scenario. The Nao robot was designed to interact verbally with participants across the game events. We used the Wizard of Oz method (WOz) to manage the game's control without revealing this to the participants, ensuring unbiased responses. The interaction had three phases: welcoming and introducing the game, playing the game, and concluding the game. At the beginning, the robot greeted the participant warmly and offered 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 four times with a 5-minute break between each session. In the second, third, and fourth sessions, the robot thanked the participants and reintroduced the game by saying: "Hello again. Thank you for playing. We are going to play another game. Are you ready?" and "Let us start" respectively. At the start of the game, the Nao robot informed the participant by saying "the game starts now". The robot initiated the first turn and followed the game rules by interacting with the participant during various game events in the following manner:

(1) When the robot selected its set of cards and declared them, e.g., "I selected two Kings"..
(2) When the participant believed the robot's claim, it responded with: "It is your turn".
(3) When the participant did not believe the robot and the robot's card declaration was accurate, the robot stated: "I was telling the truth".
(4) When the participant did not believe 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 believed the participant, it said: "I trust you, and it is my turn".
(6) When the robot did not believe the participant, it said: "I think you are bluffing". If the participant told the truth, the robot said: "Oh, I was wrong, and it is your turn now".
(7) If the robot did not believe 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 or wished the participant luck for the next round. Upon a victory, the robot cheered: "Congratulations! You've won. Thank you, and see you in the next round". In case of defeat, the robot encouraged by saying: "You've just lost the game. Good luck in the following rounds". In the final session, the robot bid farewell as it concluded the experiment.

3.2.2 Participants. The study initially aimed to involve 45 people aged between 18 and 60 years old. However, data collection issues were encountered with two of the participants, and the effective number was adjusted to 43, with a mean age of 29.53 and a standard deviation of 6.71. The group consisted of 16 females, 26 males, and one individual who preferred not to specify their gender. We recruited participants by sending out invitations through university email lists and posting flyers around the campus. Interested individuals signed up via the online platform, *Calendly.*

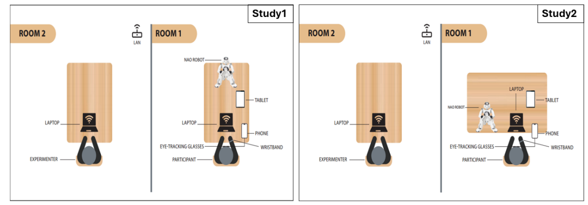

Figure 1: Experimental setup of study 1 and study 2

To determine the participants' level of familiarity with robotics, we categorised them based on their experience as high, medium, low, or none. Participants who had controlled or constructed a robot were categorised as having high experience. Those who had repeated use of robots were categorised as having medium experience, and those with occasional interactions were considered to have low experience. The remaining participants, who had never interacted with robots, were noted as having no experience. The final tally showed that 2 participants had high experience, 2 had medium experience, 24 had low experience, and 15 had no prior interactions with robots.

3.2.3 Experimental Setup and Equipment. The experimental setup was splitted across two rooms, depicted in Figure 1. The first room hosted the interactive game setup, where participants were seated at a table facing the Nao robot, with a laptop facilitating the card game. To capture physiological responses, participants were equipped with Pupil Invisible Eye Tracking Glasses and the Empatica E4 Wristband. A tablet was also provided to enter demographic details. This room was specially arranged to ensure uniform environmental conditions, such as controlled lighting and temperature, to prevent any external influence on physiological data such as BD, BR, and SKT. In the second room, an experimenter oversaw the experiment and manipulated the robot's actions through a laptop, ensuring seamless interaction between the participant and the robot. PBs and OBs data collection was carried out using two sophisticated devices: the Empatica E4 Wristband and Pupil Invisible Eye Tracking Glasses. The E4 Wristband is known for its accuracy in measuring heart rate, electrodermal activity, and body temperature, making it an invaluable tool for this study. Similarly, the Pupil Invisible Eye Tracking Glasses, with their high-resolution cameras and sensors, were pivotal in tracking eye movements and providing insights into the participants' focus and cognitive engagement during interactions with the robot.

3.2.4 Experimental Procedures. The study has the following steps:

(1) Initially, each participant was briefed about the study through an informational sheet and provided with detailed game instructions, followed by the signing of a consent form.
(2) Next, they filled out a demographic questionnaire detailing their prior experiences with robotics.
(3) The participants then equipped themselves with eye-tracking glasses and a physiological data recording wristband. Upon starting the data recording, the experimenter vacated the room.
(4) The actual game play commenced with participants engaging in the card game against the Nao robot, which was manipulated remotely by the experimenter from a separate room.

(5) Following the completion of each game round, the experimenter re-entered the room to pause the data recording and requested the participants to fill out a questionnaire assessing their interaction with the robot. It's important to note that the responses to this questionnaire were not analyzed in this paper due to their irrelevance to the study's primary objectives.
(6) The steps 3, 4, and 5 were repeated in four different game sessions.
(7) To conclude, participants were informed of their £10 Amazon voucher as appreciation for their contribution to the study.

3.3 Study 2

3.3.1 Interaction Scenario. The interaction was consistent during the welcoming and introducing and concluding the game phases. Due to the different role of robot, the role of robot interaction was limited to occurrences where participants need to take advise on whether to accept or reject the opponent move during the game phase. Following the game rules, the robot interacted with the participants during various game events. The game's flow involved the robot interacting with the participants during decisions and other situations in the game as follows:

(1) During the experiment, the robot consistently followed a predefined protocol and strategy when participants asked about the decision-making process in the accept condition. The robot provided feedback as follows: "Given the game has just started, I think we could accept the claim for now; what do you think?", "I think we could accept, what do you think?", "I suggest accept, what do you think?", or "I think it seems reasonable to accept the claim, what do you think?".
(2) In the reject claims condition, the robot said, "I think they might want to discard non-similar cards first, what do you think?", "I think they are bluffing, what do you think?", "I suggest rejecting the claim; what do you think?"
(3) If the participants agreed with the robot's suggestion to accept, the robot said "Okay, let's continue", "Okay, let's proceed", or "Okay, let's see how to conclude".
(4) If the participants agreed with the robot's suggestion of rejecting the claims, the robot said "Okay, let's see".
(5) If the participants disagreed with the robot's suggestion, the robot said "Okay, it is up to you".
(6) If the participants asked the robot to repeat the suggestion, the robot repeated the suggestion for them.
(7) If the robot did not hear the participants, the robot said "Sorry, I did not hear that, could you please repeat it".
(8) If the participants seem to have been occupied with something else, the robot said "You seem occupied with something else, could you please focus on the game".
(9) If the participants asked the robot for anything else during the game, the robot said "I can only advise you when you are deciding to accept or reject".

3.3.2 Participants. The recruitment process for participants in Study 2 was similar to that of Study 1, with the aim of maintaining consistency in the sample population by targeting a similar demographic profile. Due to data collection issues, only 41 participants were included in the study, instead of the intended 45. The participants, with an average age of 30.45 years and a standard deviation of 4.14, were a balanced mix of genders. This diversity of demographics provides varied perspectives on robotic interaction, which enriches the research. To classify the participants based on their experience with robotics, we analysed their frequency of interaction with robots, ranging from daily usage to no interaction at all. We categorised them into four groups: high, medium, low, or none. 3 participants were classified as having high experience due to their daily engagement with robots, 6 as having medium experience reflecting weekly interaction, 20 as having low experience, and 12 as having no experience with robots.

3.3.3 Setup and Materials. The setup and materials for Study 2 closely mirror those outlined in Study 1, with adjustments made to accommodate the robot's position to be next to the participant (see Figure 1). Key components include the interactive card game environment, physiological sensors for capturing participants' responses, and the data collection system. The detailed descriptions of these elements are provided in the Setup and Materials section of Study 1.

3.3.4 Experiment Procedure. The experiment procedure for Study 2 follows the structure established in Study 1. This includes participant orientation, sensor attachment and calibration, introduction to the game's mechanics and the robot's function, gameplay sessions, and post-experiment questionnaire. For a detailed explanation of these procedures, please refer to the experiment Procedure section of Study 1.

3.4 Measurements for Deception Detection

3.4.1 Physiological Measures. We collected participants' PBs and OBs responses in real-time during deceptive or truthful decisions using eye tracking technology for recording BR and BD, and a wristband for monitoring EDA, BVP, HR, and SKT. These responses were chosen for their relevance to deception detection and non-intrusive nature of collection.

3.4.2 Behavioural Measures. The game interactions captured the decision-making process of individuals choosing to display a deceptive or truthful behaviour. The outcomes of each participant's decision were documented and coded as binary values, with 0 representing truthfulness and 1 representing deception. Additionally, the researchers recorded the timestamps for the beginning and end of each decision phase to better understand the physiological responses associated with specific moments of decision-making related to deception or honesty.

3.5 Data Preparation for Deception Analysis

3.5.1 Behavioral Data Processing. Behavioural data from the game sessions were processed to extract meaningful analytics:

(1) **Outcomes processing:** The choices of whether to deceive or to be honest were counted and sorted into different categories for statistical analysis.
(2) **Timing analysis:** We carefully matched the decision-making periods with physiological data collection, ensuring precise analysis of participants' responses during critical moments.

3.5.2 Physiological Data Preprocessing. To begin with, we performed the following steps prior to analysing the physiological data:

(1) **Noise reduction:** A low-pass filter was applied to remove noise and artifacts, ensuring signal integrity for accurate analysis.
(2) **Data segmentation:** The physiological data stream was segmented according to the game rounds, aligning with the decision-making phases for each participant.
(3) **Feature extraction:** For each decision-making interval, we calculated average values for EDA, BVP, HR, SKT, BR, and BD. This step transformed raw data into a structured format conducive to detecting deception-related physiological changes.
(4) **Dataset compilation:** The final dataset was compiled by averaging physiological measures across sessions and decisions, creating a comprehensive profile for each participant's deceptive and honest behaviours.

By following these steps, we successfully generated two datasets suitable for analysing deception and honest behaviours using PBs and OBs. The datasets alongside codes can be accessed here. In the given *link*, the files named as "Dataset1_sessions" and "Dataset2_sessions" represent the datasets for session 1, 2, 3 and 4 respectively, while, the files named as "Dataset1_all" and "Dataset2_all" represent all session data.

4 RESULTS

In this section, we present the findings of both studies. We investigate whether there are significant differences in OBs and PBs during deceptive or truthful states in both studies. In addition, we investigated whether familiarity with the task enabled a significant change in OBs and PBs during deceptive or truthful states during repeated interactions. Lastly, we investigated how accurately we could classify deception in the two datasets and also applied incremental transfer learning to find whether we could detect deception with high accuracy.

To test **H1a, & H1b** and **H2a, & H2b**, we conducted a repeated-measures ANOVA with deception and truthful states as a between-subject variable and the interactive session (session 1, session 2, session 3, and session 4) as a within-subject variable on the physiological measures (EDA, BVP, HR, SKT, BR, and BD) as dependent variables (DVs). We found that there was a significant effect of deception in study 1 on EDA (($F(1,69) = 4.270$, $p = .04$, $\eta_p^2 = .058$)) and SKT ($F(1,69) = 20.124$, $p < .001$, $\eta_p^2 = .226$) scores and in study 2 on EDA ($F(1,76) = 8.730$, $p = .004$, $\eta_p^2 = .103$), HR ($F(1,76) = 4.141$, $p = .045$, $\eta_p^2 = .052$), and SKT ($F(1,76) = 23.570$, $p < .001$, $\eta_p^2 = .237$). However, we did not see a significant effect of deception on BVP ($F(1,69) = .591$, $p = .445$, $\eta_p^2 = .008$), HR ($F(1,69) = 2.925$, $p = .092$, $\eta_p^2 = .041$), BR ($F(1,69) = .00$, $p > .983$, $\eta_p^2 = .00$) and BD ($F(1,69) = .066$, $p > .798$, $\eta_p^2 = .001$) respectively in study 1 and BVP ($F(1,76) = .834$, $p = .364$, $\eta_p^2 = .011$), BR ($F(1,76) = .428$, $p = .515$, $\eta_p^2 = .006$), and BD ($F(1,76) = .796$, $p = .375$, $\eta_p^2 = .010$) in study 2. The mean and standard deviation for all the DVs in both studies can be seen in Table 4.

A significant interaction effect of session and deception was observed for BVP measures in study 2 ($F(3,74) = 2.664$, $p = .054$, $\eta^2 = .097$). However, we did not observe a significant interaction effect of session and deception (session * deception) on EDA ($F(3,67) = 1.073$, $p = .366$, $\eta^2 = .046$), BVP ($F(3,67) = .606$, $p = .613$, $\eta^2 = .026$), HR ($F(3,67) = .727$, $p = .539$, $\eta^2 = .032$), SKT ($F(3,67) = .447$, $p = .720$, $\eta^2 = .020$), BR [$F(3,67) = 1.036$, $p > .382$, $\eta^2 = .044$], and BD ($F(3,67) = .392$, $p > .759$, $\eta^2 = .017$) in study 1 and EDA ($F(3,74) = .374$, $p = .772$, $\eta^2 = .015$), HR ($F(3,74) = .241$, $p = .867$, $\eta^2 = .010$), SKT ($F(3,74) = .061$, $p = .980$, $\eta^2 = .002$), BR ($F(3,74) = .170$, $p = .916$, $\eta^2 = .007$), and BD ($F(3,74) = .570$, $p = .637$, $\eta^2 = .023$) in study 2. Lastly, we observed that only SKT varies across sessions in both study 1 and 2 ($F(3,67) = 18.957$, $p >< .001$, $\eta^2 = .020$), ($F(3,74) = 22.834$, $p < .001$, $\eta^2 = .481$), respectively.

We conducted post-hoc Bonferroni tests to assess whether SKT differed significantly between sessions 1, 2, 3, and 4 in two studies. In Study 1, the analysis confirmed that SKT was significantly higher in session 1 compared to session 2 ($p < 0.001$), session 3 ($p < 0.001$), and session 4 ($p < 0.001$). No significant differences were found when comparing session 2 to sessions 3 and 4, nor between session 3 and session 4, suggesting that SKT levels were more stable across these later sessions. In Study 2, similar trends were observed, with SKT significantly higher in session 1 compared to session 2 ($p < 0.001$), session 3 ($p < 0.001$), and session 4 ($p < 0.001$). These findings highlight the variability in SKT responses across different sessions, underscoring the impact of session-specific factors on skin temperature measurement.

To test **H3**, which aimed to investigate whether PBs and OBs can be utilised to classify truthful and deception behaviour, we followed the structured approach proposed by Ahmad et al. [3]. Seven classifiers were implemented: Random Forest (RF), Logistic Regression (LR), Support Vector Machines (SVM), Decision Tree (DT), AdaBoost (AB), Neural Network (NN), and Naive Bayes (NB). The performance of these classifiers was evaluated using 5-fold cross-validation. The findings revealed that RF, LR, and SVM achieved the highest accuracies at 75%, 71%, and 71%, respectively, while the other classifiers also performed well (refer to Table 1).

To provide a more detailed analysis of the accuracy findings, we have presented the results in the form of a classification report in Table 1. This report shows the F1 score for each class, which evaluates the performance of each classifier. The results indicate that for the RF, LR, and SVM classifier, both deception and non-deception were predicted correctly, with a 71% both, 70% and 71% and 71%, 74% accuracy rate on the test data, suggesting that LR and SVM have relatively higher accuracy compared to the other classifiers.

4.1 Feature importance for Deception and Non-deception

We examined each PB in the datasets and evaluated their ability to determine whether the subject was being truthful or deceptive. We calculated the F1 score for each class separately to gauge the effectiveness of each feature in accurately classifying the subjects. The RF, LR, and SVM classifiers exhibited the best performance in predicting deception or non-deception. Thus, we only present the feature importance for these classifiers. In Study 1, the feature importance for the deception and truthful states were: EDA (0.39, 0.62), BVP (0.58, 0.39), HR (0.58, 0.58), SKT (0.67, 0.66), BR (0.52,

Classifier	Accuracy (%)										F1-Scores			
	Study 1					Study 2					Study 1		Study 2	
	S1	S2	S3	S4	All	S1	S2	S3	S4	All	D	T	D	T
RF	60%	60%	67%	64%	69%	65%	77%	63%	63%	75%	0.70	0.66	0.76	0.74
LR	67%	70%	62%	67%	71%	71%	69%	66%	64%	69%	0.71	0.71	0.75	0.72
SVM	65%	66%	59%	69%	71%	64%	69%	70%	62%	69%	0.70	0.71	0.71	0.66
DT	55%	51%	55%	59%	60%	47%	60%	53%	67%	67%	0.61	0.59	0.68	0.65
AB	60%	51%	56%	60%	61%	54%	59%	57%	63%	69%	0.63	0.58	0.69	0.66
NN	72%	59%	56%	62%	70%	64%	77%	67%	67%	68%	0.70	0.69	0.71	0.67
NB	66%	63%	66%	65%	67%	60%	64%	71%	58%	65%	0.69	0.64	0.71	0.50

Table 1: Performance Metrics of Classifiers across Different Sessions (S) and Studies for Deception (D) and truthful (T) behaviours.

Classifier	Source (Study 1)	Target (Study 2)
RF	79%	77%
LR	77%	80%
SVM	76%	77%
DT	73%	58%
AB	76%	75%
NN	82%	77%
NB	71%	67%

Table 2: Performance of various classifiers using incremental transfer learning on two datasets.

0.49), and BD (0.43, 0.57) for LR respectively. For RF in Study 2, the values were: EDA (0.56, 0.52), BVP (0.53, 0.53), HR (0.53, 0.52), SKT (0.56, 0.56), BR (0.65, 0.66), and BD (0.55, 0.55). SVM in Study 1 showed: EDA (0.28, 0.63), BVP (0.69, 0.18), HR (0.50, 0.56), SKT (0.66, 0.65), BR (0.66, 0.06), and BD (0.36, 0.51). The features HR and SKT were consistent and effective in predicting deception and truthful classes.

4.2 Incremental Transfer Learning Results

We conducted two studies that produced two datasets. To handle this, we adopted an incremental transfer learning approach utilising seven classifiers as proposed by Chui et al. [12]. Our process involved selecting two datasets, one as the source (Dataset 1) and the other as the target (Dataset 2). We divided each dataset into equally sized subsets and trained an initial model (Model 1.1) on the first subset of Dataset 1. Then, we transferred the knowledge from Model 1.1 to train Model 2.1 on the first subset of Dataset 2. We continued updating the models with subsequent subsets until the last subsets were used. We utilised seven models, including RF, LR, SVM, DT, AB, NN and NB, with parameters illustrated in Table 3. Our method achieved significant accuracies: RF achieved 79% (source) and 77% (target), LR scored 77% and 80%, SVM showed 76% on both, DT reported 73% and 58%, AB recorded 76% and 75%, NN reached 82% and 77%, and NB showed 71% and 67% (see table 2 for more information).

5 DISCUSSION

This study investigated whether PBs & OBs can be collectively used to detect deception. In this section, we discuss whether the hypotheses were accepted or rejected in the light of the findings.

Model	Parameters
RF	n_estimators=150, max_depth=10, criterion='entropy'
LR	penalty='l2', tol=0.0001, C=1.0, fit_intercept=True, solver='lbfgs', max_iter=100
SVM	probability=True
AB	base_estimator=DecisionTreeClassifier(max_depth=1), n_estimators=50
DT	max_depth=3
NN	hidden_layer_sizes=(100,), max_iter=500
NB	GaussianNB()

Table 3: Models Parameters

Feature	Non-deception				Deception			
	Study 1		Study 2		Study 1		Study 2	
	M	SD	M	SD	M	SD	M	SD
EDA	0.99	2.59	0.37	0.44	0.38	0.46	0.99	0.84
BVP	0.02	0.18	0.03	0.30	0.03	0.32	0.01	0.17
HR	105.1	18.1	101.5	14.7	101.4	18.01	105.9	20.64
SKT	28.1	1.4	26.87	1.5	26.9	1.5	28.02	1.53
BR	2.5	2.04	2.4	1.63	2.5	1.9	2.29	2.03
BD	310.4	125.52	317.02	120.5	319.08	132.6	302.1	144.94

Table 4: Mean (M) and Standard Deviation (SD) for the physiological features under truthful and deceptive states across two sets of data.

H1 suggested a significant difference in human PBs & OBs responses, such as EDA, BVP, HR, SKT, BR, and BD, between deceptive and truthful states during HRI. We found that both EDA and SKT differed significantly during deceptive and truthful states in both experiments. EDA and SKT are physiological measures linked with galvanic skin responses that can detect deception [51]. EDA measures skin conductivity, which increases during stress or arousal states related to deception. SKT reflects changes in blood flow to the skin, which can vary due to the complex interplay between the sympathetic and parasympathetic nervous systems. Studies have shown that lying often induces nervousness or stress, as well as cognitive load, both of which are related to increased (sympathetic nervous system activity [51].

We observed a significant difference in HR between deceptive and truthful states in study 2, which represented a cooperative context. However, no such effects were seen in Study 1, which presented a competitive context. Previous research has shown that the variability in HR response to deception can be influenced by several situational and individual factors, which can explain why HR may differ in one situation and not in another [39]. The emotional response to lying can vary depending on the stakes involved, the potential consequences of being caught, and the individual's moral compass. Although cognitive load and stress were consistent in both studies, we believe that participants may have felt more pressure in the presence of the robot and attempted to maintain a moral compass [58]. This finding can be due to interpersonal dynamics, as the relationship between the deceiver and the observer can influence HR. For instance, lying to a stranger may not elicit the same physiological response as lying to a loved one [39]. Furthermore, studies have shown inconsistencies in the correlation between HR and deception. Some studies have found an increased HR in guilty individuals, while others have indicated that lying could decrease HR [26]. Lastly, a recent study suggests that the presence of robots can have a similar impact on HR as working with other humans, potentially due to the development of trust and the integration of robots as team members [16].

On the other hand, the other PBs and OBs (BVP, BR, and BD) did not show significant differences between truthful and deceptive states. We understand that BVP and blinking rates can be influenced by stress and cognitive load. The regulation of BVP is complex and can be maintained across different emotional states, and blinking rates are subject to voluntary control and are influenced by various contextual factors. Therefore, these measures could not differ significantly in truthful and deceptive states. In addition, the context of the interaction can influence blinking rates. For example, if an individual is in a relaxed and informal setting, they may blink less frequently, regardless of whether they are being truthful or deceptive [10]. Moreover, changes in BVP and other PBs correlate with anxiety, but such conditions may not have been present during the game-based context. The absence of pressure elements, such as time constraints and individual differences in stress response and cognitive load, may have contributed to the variability in physiological responses.

In summary, the hypothesis **H1a** and **H1b** were *partially accepted* as we did not find significant differences for all the PBs.

H2 hypothesised an interaction effect (session and decision to be truthful and deceptive) 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 except BVP in study 2. We understand that deception can cause consistent physiological behaviours during repeated interactions due to several psychological and physiological factors [46]. Individuals can adapt to the act of deception over time, leading to a decrease in physiological responses. With repeated exposure to the same deceptive scenarios, individuals can become habituated to the stress associated with lying leading to a reduction in physiological responses. Frequent deception can make individuals more skilled at lying, resulting in less pronounced physiological changes and more consistent behaviours. Repeatedly engaging in deceptive behaviour can lead to emotional desensitisation, where the emotional impact of lying diminishes over time, resulting in fewer physiological changes and consistent behaviours. Increased efficiency at lying reduces physiological indicators of deception. Individuals have different baseline physiological responses, and familiarity with the situation reduces the physiological response to deception. If the context of the repeated interactions remains consistent, it makes it harder to discern differences between truthful and deceptive behaviours.

H3 suggested that classification algorithms will classify instances of deception with potentially high accuracy. The results of the two studies were promising, with LR, SVM, NN, and RF classifiers detecting deception with accuracy rates of over 70% and 75%. In both studies, HR, and SKT features were crucial for detecting deception in the best-performing classifiers. The importance of these features is linked to their association with emotional arousal, cognitive effort, and rapid physiological changes that typically occur in response to deception in game contexts [37].

Our study utilised a new incremental transfer learning algorithm and achieved an accuracy rate of 80%, surpassing the current deception detection rates based on PBs and OBs [6]. This indicates that our hypothesis (H3) was accepted. The high accuracy was possible due to the negative transfer avoidance algorithm included in incremental transfer learning, which reduces the risk of transferring irrelevant information and facilitates the transfer of knowledge [12]. Additionally, the use of multiple PBs and OBs is crucial since PBs are often dependent on the task or environment [3]. To summarise, our dataset consisted of natural and repeated interactions, including both deceptive and truthful states, resulting in a large and diverse set of data that helped us achieve good results.

6 CONCLUSION & FUTURE WORK

This paper highlights the limitations of current datasets used for deception detection, as they lack diversity and realism and have a limited number of subjects in simulated environments that do not accurately reflect real-world deceptive behaviours. To overcome this, we used the GaME (game as a method to elicit emotions naturally) paradigm and created a dataset based on organically depicted deceptive and truthful interactions. The dataset was based on repeated interactions, which is a further significant improvement from existing work that only offers a dataset on one-off interactions. We conducted two experiments involving 83 participants to investigate whether different physiological and oculomotor behaviours (PBs & OBs) collected naturally during deceptive and truthful states differ significantly. Additionally, we explored whether combining PBs and OBs can accurately predict humans in deceptive and truthful states during HRI. Our findings confirmed that PBs such as EDA and SKT differed in deceptive and truthful states. It indicated that multiple PBs collectively detect deception in real time during HRI. For the first time, we used the novel incremental transfer learning to detect deception and achieved an 80% accuracy, surpassing most of the existing work. We encourage the research community to use the GaME paradigm in different contexts to improve the rate of deception detection. We promote incremental transfer learning techniques to yield optimal results in the target (new) models.

While this study shows promise, it is important to note its limitations. Findings are specific to game-based robot interactions and may not apply to other contexts or human interactions. The limited demographic characteristics, mainly consisting of students, may restrict the generalisability of the results. Future research will involve testing in various environments and including participants from diverse backgrounds to enhance the findings. We also plan to explore how combining facial and speech features along with PBs and OBs in different contexts can improve detecting deception rates. We aim to use such detection mechanisms to develop adaptive robotic systems that can have wider applications.

REFERENCES

[1] Rachel Adelson. 2004. Detecting deception. *APA Monitor on Psychology* 35, 7 (2004), 70–73.

[2] Muneeb Ahmad and Abdullah Alzahrani. 2023. Crucial Clues: Investigating Psychophysiological Behaviors for Measuring Trust in Human-Robot Interaction. In *Proceedings of the 25th International Conference on Multimodal Interaction*. 135–143.

[3] 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.

[4] Muneeb Imtiaz Ahmad, Omar Mubin, and Hiren Patel. 2018. Exploring the potential of NAO robot as an interviewer. In *Proceedings of the 6th International Conference on Human-Agent Interaction*. 324–326.

[5] Muneeb I Ahmad, Hassan Tariq, Mehreen Saeed, Suleman Shahid, and Emiel Krahmer. 2011. Guess who? An interactive and entertaining game-like platform for investigating human emotions. In *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments: 14th International Conference, HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part III 14*. Springer, 543–551.

[6] Haya Alaskar, Zohra Sbaï, Wasiq Khan, Abir Hussain, and Arwa Alrawais. 2023. Intelligent techniques for deception detection: a survey and critical study. *Soft Computing* 27, 7 (2023), 3581–3600.

[7] Faryal Amber, Adeel Yousaf, Muhammad Imran, and Khurram Khurshid. 2019. P300 based deception detection using convolutional neural network. In *2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE)*. IEEE, 201–204.

[8] Alexander Mois Aroyo, Jonas Gonzalez-Billandon, Alessia Tonelli, Alessandra Sciutti, Monica Gori, Giulio Sandini, and Francesco Rea. 2018. Can a humanoid robot spot a liar?. In *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. IEEE, 1045–1052.

[9] David B Buller and Judee K Burgoon. 1996. Interpersonal deception theory. *Communication theory* 6, 3 (1996), 203–242.

[10] Judee K Burgoon and Norah E Dunbar. 2016. Accuracy of distinguishing truth from lie. *The social psychology of perceiving others accurately* (2016), 71–97.

[11] Luis Vicente Calderita, Luis J Manso, Pablo Bustos, Cristina Suárez-Mejías, Fernando Fernández, and Antonio Bandera. 2014. THERAPIST: towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. *JMIR rehabilitation and assistive technologies* 1, 1 (2014), e3151.

[12] Kwok Tai Chui, Varsha Arya, Shahab S Band, Mobeen Alhalabi, Ryan Wen Liu, and Hao Ran Chi. 2023. Facilitating innovation and knowledge transfer between homogeneous and heterogeneous datasets: Generic incremental transfer learning approach and multidisciplinary studies. *Journal of Innovation & Knowledge* 8, 2 (2023), 100313.

[13] Sergey Demyanov, James Bailey, Kotagiri Ramamohanarao, and Christopher Leckie. 2015. Detection of Deception in the Mafia Party Game. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Seattle, Washington, USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 335–342. https://doi.org/10.1145/2818346.2820745

[14] Bella M DePaulo and Wendy L Morris. 2004. Discerning lies from truths: Behavioral cues to deception and the indirect pathway of intuition. *BM DePaulo, WL Morris, ad. by PA Granhag. The detection of deception in forensic contexts//New York: Cambridge* (2004), 15–40.

[15] Mingyu Ding, An Zhao, Zhiwu Lu, Tao Xiang, and Ji-Rong Wen. 2019. Face-focused cross-stream network for deception detection in videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7802–7811.

[16] Gwendolyn Edgar, Ayca Aygun, Matthew McWilliams, and Matthias Scheutz. [n. d.]. Towards Genuine Robot Teammates: Improving Human-Robot Team Performance Beyond Shared Mental Models with Proactivity. ([n. d.]).

[17] Empatica Inc. Year of access. *Empatica E4 Research Device*. https://www.empatica.com/en-eu/research/e4/

[18] Lynette Gabonthone. 2019. Critical Evaluation of Machine Learning Research Aimed at Improving Lie Detection Accuracy. *Selected Computing Research Papers* (2019), 1.

[19] Andrea Gaggioli et al. 2018. Beyond the truth machine: emerging technologies for lie detection. *Cyberpsychology, Behavior and Social Networking* 21, 2 (2018), 144–144.

[20] Ascensión Gallardo-Antolín and Juan M Montero. 2021. Detecting deception from gaze and speech using a multimodal attention LSTM-based framework. *Applied Sciences* 11, 14 (2021), 6393.

[21] Matthias Gamer and Wolfgang Ambach. 2014. Deception research today. *Frontiers in psychology* 5 (2014), 88426.

[22] Sabu George, MM Manohara Pai, Radhika M Pai, and Samir Kumar Praharaj. 2017. Eye blink count and eye blink duration analysis for deception detection. In *2017 International conference on advances in computing, communications and informatics (ICACCI)*. IEEE, 223–229.

[23] Mandar Gogate, Ahsan Adeel, and Amir Hussain. 2017. Deep learning driven multimodal fusion for automated deception detection. In *2017 IEEE symposium series on computational intelligence (SSCI)*. IEEE, 1–6.

[24] Jonas Gonzalez-Billandon, Alexander M Aroyo, Alessia Tonelli, Dario Pasquali, Alessandra Sciutti, Monica Gori, Giulio Sandini, and Francesco Rea. 2019. Can a robot catch you lying? a machine learning system to detect lies during interactions. *Frontiers in Robotics and AI* 6 (2019), 64.

[25] Viresh Gupta, Mohit Agarwal, Manik Arora, Tanmoy Chakraborty, Richa Singh, and Mayank Vatsa. 2019. Bag-of-lies: A multimodal dataset for deception detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 0–0.

[26] Danielle Hipsky, Elizabeth Parimanath, Joseph Mueller, Mike McGinnis, and Stephanie Wright. 2014. Analyzing Fluctuations in Heart Rate, Respiration, and Perspiration as Physiological Indicators of Deception. (2014).

[27] Charles R Honts and John C Kircher. 1994. Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology* 79, 2 (1994), 252.

[28] Ray Hyman. 1989. The psychology of deception. *Annual review of psychology* 40, 1 (1989), 133–154.

[29] Samee Ibraheem, Gaoyue Zhou, and John DeNero. 2022. Putting the con in context: Identifying deceptive actors in the game of mafia. *arXiv preprint arXiv:2207.02253* (2022).

[30] Koji Inoue, Kohei Hara, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara. 2020. Job interviewer android with elaborate follow-up question generation. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 324–332.

[31] Nishant Jain, Abhay Kumar, Shekhar Singh, Chirag Singh, and Suraj Tripathi. 2019. Deceptive reviews detection using deep learning techniques. In *Natural Language Processing and Information Systems: 24th International Conference on Applications of Natural Language to Information Systems, NLDB 2019, Salford, UK, June 26–28, 2019, Proceedings 24*. Springer, 79–91.

[32] Mimansa Jaiswal, Sairam Tabibu, and Rajiv Bajpai. 2016. The truth and nothing but the truth: Multimodal analysis for deception detection. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 938–943.

[33] Hamid Karimi. 2018. Interpretable multimodal deception detection in videos. In *Proceedings of the 20th ACM international conference on multimodal interaction*. 511–515.

[34] Wasiq Khan, Keeley Crockett, James O'Shea, Abir Hussain, and Bilal M Khan. 2021. Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection. *Expert Systems with Applications* 169 (2021), 114341.

[35] Gangeshwar Krishnamurthy, Navonil Majumder, Soujanya Poria, and Erik Cambria. 2018. A deep learning approach for multimodal deception detection. In *International Conference on Computational Linguistics and Intelligent Text Processing*. Springer, 87–96.

[36] David W Kupferman. 2022. I, robot teacher. , 1513–1522 pages.

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

[38] Timothy R Levine. 2014. Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology* 33, 4 (2014), 378–392.

[39] Timothy R Levine. 2019. An overview of detecting deceptive communication. *The palgrave handbook of deceptive communication* (2019), 289–301.

[40] Frank M Marchak. 2013. Detecting false intent using eye blink measures. *Frontiers in psychology* 4 (2013), 36551.

[41] Jaume Masip, Iris Blandón-Gitlin, Carmen Martínez, Carmen Herrero, and Izaskun Ibabe. 2016. Strategic interviewing to detect deception: Cues to deception across repeated interviews. *Frontiers in psychology* 7 (2016), 1702.

[42] Karnati Mohan and Ayan Seal. 2021. Deception detection on "Bag-of-Lies": integration of multi-modal data using machine learning algorithms. In *Proceedings of International Conference on Machine Intelligence and Data Science Applications: MIDAS 2020*. Springer, 445–456.

[43] Merylin Monaro, Pasquale Capuozzo, Federica Ragucci, Antonio Maffei, Antonietta Curci, Cristina Scarpazza, Alessandro Angrilli, and Giuseppe Sartori. 2020. Using blink rate to detect deception: A study to validate an automatic blink detector and a new dataset of videos from liars and truth-tellers. In *Human-Computer Interaction. Human Values and Quality of Life: Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part III 22*. Springer, 494–509.

[44] Fred Paas and John Sweller. 2014. Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning* 27 (2014), 27–42.

[45] Jeffrey Pawlick and Quanyan Zhu. 2015. Deception by design: evidence-based signaling games for network defense. *arXiv preprint arXiv:1503.05458* (2015).

[46] Richard Reardon, Annette L Folwell, Jode Keehr, and Trevor Kauer. 2019. Effects of deception on the deceiver: An interdisciplinary view. *The Palgrave handbook of deceptive communication* (2019), 107–125.

[47] M Umut Şen, Veronica Perez-Rosas, Berrin Yanikoglu, Mohamed Abouelenien, Mihai Burzo, and Rada Mihalcea. 2020. Multimodal deception detection using real-life trial data. *IEEE Transactions on Affective Computing* 13, 1 (2020), 306–319.

[48] Suleman Shahid. 2012. Fun & Face: Exploring non-verbal expressions of emotion during playful interactions. (2012).

[49] Jeremy Speth, Nathan Vance, Adam Czajka, Kevin W Bowyer, Diane Wright, and Patrick Flynn. 2021. Deception detection and remote physiological monitoring: A dataset and baseline experimental results. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 1–8.

[50] R Weylin Sternglanz, Wendy L Morris, Marley Morrow, and Joshua Braverman. 2019. A review of meta-analyses about deception detection. *The Palgrave handbook of deceptive communication* (2019), 303–326.

[51] Sabine Ströfer, Matthijs L Noordzij, Elze G Ufkes, and Ellen Giebels. 2015. Deceptive intentions: Can cues to deception be measured before a lie is even stated? *PloS one* 10, 5 (2015), e0125237.

[52] Michele K Surbey and Jeffrey J McNally. 1997. Self-deception as a mediator of cooperation and defection in varying social contexts described in the iterated prisoner's dilemma. *Evolution and Human Behavior* 18, 6 (1997), 417–435.

[53] Leanne Ten Brinke, Julia J Lee, and Dana R Carney. 2019. Different physiological reactions when observing lies versus truths: Initial evidence and an intervention to enhance accuracy. *Journal of personality and social psychology* 117, 3 (2019), 560.

[54] Su-Jing Wang, Bing-Jun Li, Yong-Jin Liu, Wen-Jing Yan, Xinyu Ou, Xiaohua Huang, Feng Xu, and Xiaolan Fu. 2018. Micro-expression recognition with small sample size by transferring long-term convolutional neural network. *Neurocomputing* 312 (2018), 251–262.

[55] Whoop. [n. d.]. *WHOOP.* https://www.whoop.com/us/en/ Accessed on January 29, 2024.

[56] Wikipedia contributors. 2024. Cheat (game) — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Cheat_(game) [Online; accessed 3-May-2024].

[57] Yue Xie, Ruiyu Liang, Huawei Tao, Yue Zhu, and Li Zhao. 2018. Convolutional bidirectional long short-term memory for deception detection with acoustic features. *IEEE Access* 6 (2018), 76527–76534.

[58] Aakash Yadav, Patralika Ghosh, Malik Rawashdeh, Diane Lee, Thomas Bolf, and Ranjana K Mehta. 2023. All Human Versus Human-Robot Teaming: Measuring Neurophysiological Synchrony, Team Performance, and Trust during Search and Rescue. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 67. SAGE Publications Sage CA: Los Angeles, CA, 432–433.

[59] Ke Zhang, Lara A Frumkin, Alex Stedmon, and Glyn Lawson. 2013. Deception in context: Coding nonverbal cues, situational variables and risk of detection. *Journal of Police and Criminal Psychology* 28 (2013), 150–161.

[60] Yan Zhou, Heming Zhao, Xinyu Pan, and Li Shang. 2015. Deception detecting from speech signal using relevance vector machine and non-linear dynamics features. *Neurocomputing* 151 (2015), 1042–1052.

[61] Yan Zhou, Heming Zhao, and Li Shang. 2017. Lying speech characteristic extraction based on SSAE deep learning model. In *Intelligent Computing Methodologies: 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part III 13*. Springer, 672–681.

[62] Miron Zuckerman, Bella M DePaulo, and Robert Rosenthal. 1981. Verbal and nonverbal communication of deception. In *Advances in experimental social psychology*. Vol. 14. Elsevier, 1–59.