Deciphering Trust: Multi-Modal Affective Analysis in Dynamic Decision-Making Games

Darryl Roman[1], Haily Follese[2], Jordan Schotz[2], Shensheng Wang[2], Johnathan Mell[1], Nichole Lighthall[2]

[1]University of Central Florida, Computer Science and Computer Engineering
[2]University of Central Florida, Psychology

Abstract

Investigating how trust is built and maintained is especially important as technological advances make scam and fraud easier and quicker to enact. Fields such as neuroeconomics, psychology, and computer science have devoted considerable attention to the roles that emotional expression possess in determining decision making, with many studies utilizing paradigms including trust games, negotiation games, and dilemma games to model real-world decision-making processes. Current research on player behavior and decision making typically isolates specific aspects, such as acts of betrayal by a trustee or the influence of emotional facial expressions. In contrast, the paper herein describes a study that comprehensively examines both elements while incorporating automatic facial analysis, adding a source of multimodal affective data. This technology, which allows for real-time, objective, and non-intrusive data collection, has been piloted in a dynamic dyadic trust game environment. The following study builds a task framework based on current theories that inform the role of facial expression in decision-making, current models that guide predictive decision making, and the role that automatic facial analysis plays within the aforementioned. We implement that framework to conduct a pilot study investigating human behavioral responses to affective expressions applied to a digital agent.

Introduction

As simulated social exchanges with digital agents become commonplace, trust and believability are key factors in shaping interactions. For human-agent dyadic exchanges, the realism and emotional expressivity of the agent are significant influences on the quality of interactions. Combining aspects of psychological theory with computer analysis capable of interpreting social cues can lead to useful insights into how visual and emotional cues—such as using real human images with naturalistic expressions for digital agents—affect trust and believability, and can therefore improve the quality of the interactions. These insights allow us to design more effective and affective systems.

Incorporating real human images into realistic social exchanges with models that demonstrate emotional expressions as avatars of digital players takes advantage of psychological principles of social cognition, such as the theory of mind and emotional contagion. Humans are skilled in interpreting facial expressions and then using them to infer intentions, emotions, and credibility. These inferences make facial imagery valuable for improving digital interactions. Existing research indicates that realistic emotional content can evoke stronger affective responses (Torre, Goslin, and White 2020), suggesting that bridging the emotional gap between human and digital agents could cultivate a greater sense of trustworthiness and engagement.

This study explores the implications of these principles through the use of emotionally expressive real human images as naturalistic representations of a digital agent. Our work aims to achieve high ecological validity through our specific setup, in the context of dyadic interactions in an investment game. We investigate the resulting behavior and behavioral reactions of human participants. Through experimental evaluation, we are able to control for specific expressions and are therefore able to analyze how these expressions affect:

- Human responses to the agent's reciprocation behavior
- Human reactions to the agent's emotionally expressive displays
- The combined influence of reciprocation and emotional expression

Behavioral data generated through this approach will facilitate the creation of interactive systems that can effectively action plan in emotionally dynamic environments, enhancing their capacity for realistic and adaptive social interactions.

Related Works

Human Behavior in Trust Contexts: Affective Models

Past work in human trust explored the concept through iterative or non-iterative economic trust, dilemma, or negotiation games. According to game theory, participants engaging in these paradigms should act in a 'rational' way, essentially, maximizing gain by betraying a partner. However, participants in various trust games consistently cooperate more than 60% of the time (Hoegen, Stratou, and Gratch 2017). However, this cooperation is conditional and many participants opt for a tit-for-tat strategy, where a participant returns

the same behavior their opponent did— they betray or cooperate in a reactive fashion, returning back a behavior they believe the opponent expresses. Social science researchers, Chang et al., found the ratio as 80% cooperation to 20% betrayal (Chang et al. 2010). Phan et al. found that brain behaviors begin to change close to these same thresholds (Phan et al. 2010).

Thus, interest lies in what drives cooperation and trust between participants in such games and also in what acts to break this trust. There are opposing perspectives on the factors that influence trust and can predict the behavior of players. Some researchers place emphasis on the behavior itself (cooperation, defection, betrayal, use of differing strategies), while others consider the reaction to visual stimuli, such as facial expressions of an opponent, as vital in determining a participant's future actions (Dang and Ignat 2016).

For example, Stratou et al. (2017) focus solely on the power of facial expressions to predict participant behavior (Stratou et al. 2017). Specifically, they utilize facial action units (AUs) to assess participant behavior. They grouped individual AUs into different factors and analyzed data from an iterative prisoner dilemma (IPD) corpus, finding that it is not individual AUs that are most powerful in predicting behavior, but rather the co-occurrence of AUs. Although this model does not incorporate the decision-making behavior of previous participants to make its predictions, it exhibits external validity and shows support for the power of facial expression as a reliable indicator of predict behavior.

Hoegen, Stratou, and Gratch (2017) combined both of these components (behavior and facial expression) to create a model capable of predicting an opponent's decision in the Iterative Prisoner Dilemma (IPD) game based on previous behavior and their emotional signaling during previous rounds. Here, the researchers found that the predictive model, which was created using a combination of action and facial expressions from the IPD corpus, outperformed all the following in predicting opponent behavior:

- Naive baseline model: Always assumed cooperation
- Second baseline model: Assumed a tit-for-tat strategy
- 'Action-only' model: Predicted behavior based solely on action (cooperate or defect)
- Emotion-only model: Predicted behavior solely on emotional expression.

Interestingly, they also found that when defection was present, emotion was more tightly coupled to the prediction of the outcome. This implies disruption in the expectation of cooperation may create a "need to plan the next move and that includes estimating the opponent's intentions by all available input (actions and emotions)" (Hoegen, Stratou, and Gratch 2017).

Emotional Expressivity and Congruence

Further research has focused on this finding, exploring how the congruence/incongruence between emotion displayed and behavior shown impacts player behavior. In particular, Hoegen et al. (2018) found that congruent expression mimicry by a virtual agent caused an increase in participant smiling, but this congruent mimicry did not affect behavioral decision making by the participant (Hoegen et al. 2018). In contrast, Torre, Goslin, and White (2020) found that smiling voices of autonomous agents, which presumably transmit emotional affect as smiling can be detected through voice, elicited greater investment overall, regardless of whether the agent acted in an incongruent manner (i.e., performed programmed 'generous' or 'mean' behavior). Other studies, such as Lei and Gratch (2023), call for an even more expansive view of emotional expressivity, highlighting the importance of facial expressions and dynamic movement, head movement, and gestures (Lei and Gratch 2023).

Theories on the Role of Emotion on Trust

Despite this call for a more expansive view of what encompasses emotional expressivity, the role of facial expressions in dictating trust has been largely dominated by several theories in the world of affective computing, psychology, and computer science, such as the EASI model, social appraisal pathway, affective pathway, and reverse appraisal pathway. The EASI model indicates that "emotions are used to make sense of ambiguous situations, and their effect depends on the situation in which the interaction takes place, specifically its cooperative or competitive nature" (Torre, Goslin, and White 2020; Van Kleef, De Dreu, and Manstead 2010). The social appraisal pathway establishes that emotions of others affect situational appraisal, which subsequently affects decision-making (Manstead and Fischer 2001). The affective pathway posits that emotions are 'contagious' and theorizes that the emotions of another person directly influence the emotions of the participant (De Gelder and Hadjikhani 2006). Finally, the reverse appraisal pathway suggests that emotions allow participants to navigate ambiguous situations by inferring the mental states of another agent/being (de Melo, Gratch, and Carnevale 2013).

Through 5 separate experiments, De Melo et al. (2013) find support for the reverse appraisal pathway, finding that participants can successfully infer another person's mental state by observing their emotional expressions. The Facial Action Coding System (FACS) of emotional expressions, which was presented by Ekman in 1971, identifies specific action units as the foundation for appraisal of true emotional expressions (Ekman, Friesen, and Tomkins 1971; Ekman and Friesen 1976). This coding system was further refined to categorize facial expressions into 6 classifications. However, facial emotional expressions are not always genuine and therefore appraisal of emotion may not always lead to accurate inferences of one's mental state, especially if an opponent aims to deceive (Schneider, Hempel, and Lynch 2013).

Shore and colleagues (2023) explored genuine versus regulatory expressions in an IPD utilizing the AFFDEX by Affectiva module of the iMotions software suite, which automatically codes AUs informed by the FACS database (Shore et al. 2023). Surprisingly, they found that cooperation actually increased when a partner perceived their opponent to be regulating their emotions, contrasting previous research

indicating that people can distinguish genuine vs. manipulative or forced expressions and may cooperate less with non-genuine expressions (Duchenne smiles versus forced regulatory smiles) (Mehu, Little, and Dunbar 2007). AFFDEX is comparable with Facial electromyography (fEMG) in properly detecting facial expressions, and can provide rich FEA data more quickly than manual encoding (Kulke, Feyerabend, and Schacht 2020).

Computer-Controlled Actions: Perceived Agency

Prior work utilized naturalistic datasets to analyze and predict outcomes in several trust paradigms (with the exception being Stratou's mimicry experiment). Incorporating naturalistic data ensures that the learned models reflect actual human behavior. However, these naturalistic data are more difficult to analyze independently and generally reduce overall experimental controls. Thus, another method used in trust games is generating algorithmic responses which allows for more experimental control, but can also reduce the believability of the interactions and impact trust. This dilemma emphasizes investigating the impact of perceived agency on behavioral decision-making.

Importantly, in real-world contexts, scams often involve agents impersonating real people, making the detection of agency crucial for helping individuals decide whether to trust or cooperate in online environments. For de Melo, Gratch, and Carnevale (2013), the "computers as social actors" theory is evaluated, which contends that humans aimlessly treat computers the same as they would humans, if they (the computers) possess sufficient social traits (de Melo, Gratch, and Carnevale 2013). Instead, de Melo and colleagues found that actual agency is difficult for participants to detect, but the mere belief that virtual humans (VH) were being controlled by a human and not a computer algorithm led to increased cooperation in a trust game. However, this finding does not follow in other applications. Lucas and colleagues (2014) and other studies investigating trust in personal information disclosure (PID) find that when participants perceive that a computer program controls an interview, self-disclosure increases, as the perceived lack of agency of a computer likely reduces the apprehension associated with being evaluated (Lucas et al. 2014). These contrasting findings highlight that the effects of perceived agency on trust are highly context-dependent, with different dynamics at play in trust games versus personal disclosure scenarios.

Methodologically ensuring this believability means showing a participant that the digital agent with which they are interacting "possess[es] the ability to suspend the users' disbelief, by providing an illusion of life" and "[...] making the human user accept they are interacting with a living character, whose existence is consistent and coherent in the context of the virtual world it is situated in" (Avradinis, Panayiotopoulos, and Anastassakis 2013).

Methodology

This study uses a task designed to evaluate the impact of both behavior and perceived emotional expressions on the likelihood that human participants increase their investments in a dyadic economic game. The task is based on prior research on investment games, the use of digital avatars, the interpretation of emotional expressions, and behavioral responses to these expressions. Economic games often explore decision making under uncertainty, resource allocation, and strategic interactions between participants (Hula et al. 2021; Becchetti and Degli Antoni 2010; Gneezy, Güth, and Verboven 2000).

The Stimuli

Thus far, many studies which have analyzed participant cooperation and reaction to betrayal have utilized corpuses of natural human behavior data to train their models. These resultant datasets are relevant and valuable for the creation of accurate models for intelligent systems which require an understanding of naturalistic human behavior. However, these data can also confound extraneous variables resulting from the unpredictable nature of human-human behavior (Vinciarelli et al. 2015). In this study, we elicit trust learning and betrayal response behaviors by utilizing algorithmic agentic behavior based on game theory, while recording facial expressions.

To enhance the likelihood that participants believed they were interacting with real people despite maintaining algorithmic agentic control, we utilized stimuli from the Virtual Avatar Facial Stimuli Set (VAFSS). Six photos (anger, disgust, happiness, fear, neutral, and sadness) for each young adult male face were converted to video format and processed using iMotions, with each photo being presented for two seconds. iMotions does not immediately 'plateau' to stable ratings for each photo, so ratings used were based on averages for each photo from 666ms-1666ms (average from 30 samples/frames in that 1000ms range). Models who displayed identifiable emotional expressions for all six emotion categories were selected. In addition, emotional expressions used from the VAFSS data set were also rated with iMotions Affectiva and the models used were those found to display the emotional expression with low detectable intensity (Jordan Schotz 2025). Systematically choosing faces with minimally detectable emotional expressions functioned as a mechanism to increase realism by decreasing emotional exaggeration, which can disrupt perceived immersion in a task or environment. Studies have shown that subtle expressions have equal or greater ecological validity and value as signals of emotional content (Matsumoto and Hwang 2014; Schneider et al. 2022). Finally, images were processed in Adobe Photoshop to increase the graininess of each photo to match the graininess commonly found in webcam recordings, such as the device used to record participants during the Trust Game.

The Task

Our task extends the principles of game theory through a two-player investment game consisting of three blocks of 15 trials each. In each trial, participants, known as the Investor, decide how much to invest (ranging from $0 to $10). The investment is then multiplied by 4. Then, a second player, a digital agent known as the Trustee, can either return 50%

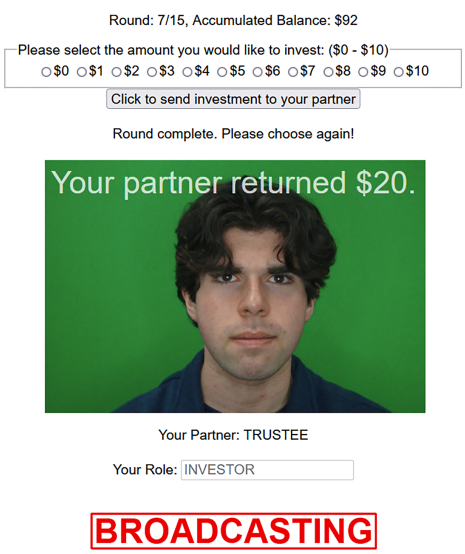

Figure 1: The Task Interface

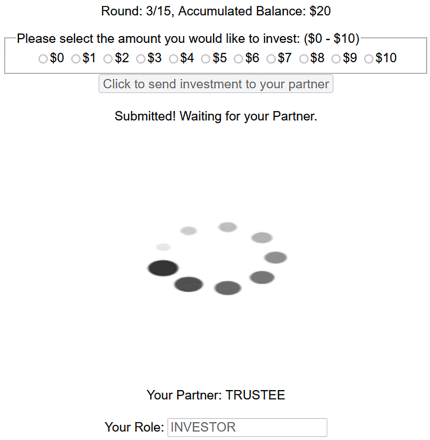

Figure 2: The Jitter Stage Interface

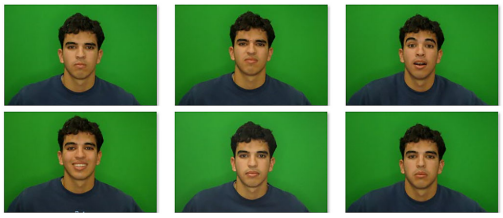

Figure 3: Portrait 6 - Anger, Disgust, Fear, Happy, Neutral, Sadness

of the multiplied amount, effectively doubling the initial investment, or betray the investor by withholding the entire amount. The participants are informed that the second player represents another human participant. This setup requires participants to continuously assess the risks and benefits of their investment decisions, taking into account the behavior and perceived strategy of their counterpart.

The digital agent is presented as a stationary avatar with randomized emotional expressions through a set of portraits, which may or may not align with its behavior (e.g., betrayal paired with a neutral or friendly expression). This design introduces scenarios that test participants' interpretations of emotional cues and their influence on trust, believability, and cooperation. By manipulating the congruence between expressions and behavior, the task provides an opportunity to study the psychological and strategic dynamics underpinning trust, believability, and decision making.

Participants are likely to adopt various investment strategies, ranging from conservative approaches to minimize risk to aggressive strategies aimed at maximizing the potential return on investment. Outcomes and emotional cues presented in each trial can shape subsequent decisions, eliciting psychological responses such as reciprocity or retaliation. In successive trials, participants can adapt their strategies based on feedback, creating a dynamic process of strategic evolution.

To further examine the role of perceived identity, participants are informed at the start of each 15-trial block that they are interacting with a new human counterpart, represented by a different portrait, although the algorithmic behavior of the digital agent remains consistent across all blocks. This feature isolates the effect of perceived emotional expressions and player identity from the underlying behavior of the agent. By systematically varying emotional expressions and observing participant responses, this framework provides a robust method for studying decision-making processes, economic behavior, and the influence of perceived emotional states on human-computer interactions.

The Stages

The first trial of our investment game allows the human participant to select their initial investment. Once selected and for the remaining 14 trials of each game, the dyadic agent's behavior is partitioned into 4 stages per trial:

- Jitter
- Portrait
- Reciprocation
- Investment

The Jitter stage is a randomized wait between 3 and 9 seconds that is utilized to increase the believability of the study by giving the appearance of a human selecting a reciprocation response. During this stage, the portrait is set to a white background with an animated Wait graphic giving the impression that the trustee is making their selection and the investment selection is deactivated.

Once the randomized jitter has passed, the portrait representing the digital agent, the Trustee, is painted on the interface. The portrait is randomly selected from a set of 6 images, each of which represents a single emotional expression, anger, disgust, happiness, fear, neutral, or sadness. This stage continues for 2 seconds.

Round	Chance of Betrayal
1-4	25%
5-11	33%
12, 13	50%
14, 15	100%

*Maximum of 3 Betrayals.

Table 1: Probability of Betrayal

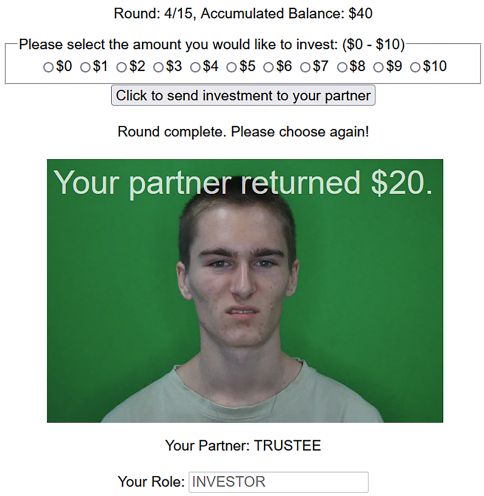

Figure 4: Notification of Reciprocation

Once 2 seconds has elapsed, a translucent notification is painted over the portrait to indicate what the Trustee has returned from the 4x investment amount that the investor selected. The algorithm for non-reciprocation is based on a probability selection to match the 80%/20% ratio by allowing a maximum of 3 betrayals per block of 15 trials.

And finally, the investment selector is reactivated and the human participant may now choose their investment and begin the loop again.

Results

Participants

51 participants were recruited through the local in-person participant pool composed of students attending the University of Central Florida (UCF). Participants were awarded course credit for participation and a $5 Amazon Gift Card. Despite all participants receiving the same amount of money on their Amazon Gift Card, they were informed during the consent process that the amount they would receive would depend on their performance during the task and that they could receive up to $5 (incentive-compatible payment mechanism). The fixed rate of payment was later disclosed during the debriefing process.

Participants in our prior 3 studies using this econonomic game returned a mean believability rating of 4.49 on a 7-point Likert scale, in which 1 is the highest possible score. This study found a mean believability rating of 3.4, indicating that participants found this version of the task highly believable compared to prior studies.

Two participants were excluded due to technical issues that were later discovered to interfere with the video recordings. Three participants were excluded due to an inability to see without behaving in such a manner that interfered with recording facial expressions (clearly without squinting or moving too close to the screen). One participant was excluded because they reported during the post-task believability check that they did not believe their partners to be live human agents. The remaining sample included 45 participants to be used in the following analyses. This sample was mainly female (63.04% assigned female at birth, 36.96% assigned male at birth), approximately 19 years of age (M = 18.79, SD = 2.01), and primarily Caucasian (69.57% Caucasian, 15.22% Multiracial, 8.7% Asian, and 6.5% Black).

Although all stimulus images used in the trust game depicted young adult males, participants' gender identity (female, male, other) did not show a relationship with investment behavior ($F(2, 34) = 1.67$, $p = .191$, $R^2 = .139$). Furthermore, participant investment behaviors did not significantly differ according to participants' self-reported race ($F(3, 34) = 0.40$, $p = .806$, $R^2 = .047$). These results indicate that despite the use of homogeneous visual stimuli (all young adult males), neither participants' racial identity nor their gender identity influenced overall average investment behaviors. We explore two main hypostheses with regards to this data set.

Hypothesis 1: Investments will be greater on trials following a partner's reciprocation

A linear mixed-effects model (LME) was conducted to test whether participants adjusted their investment decisions based on their partner's reciprocation on the previous trial, as well as whether partner emotional expressions modulated responses to reciprocity or betrayal in the trust game. Our model included reciprocation as a binary fixed-effect predictor (reciprocated vs. not reciprocated) and partner facial expression (anger, disgust, happiness, fear, neutral, or sadness) as fixed effects, while accounting for random intercepts for participants to control for individual differences.

The results revealed a significant fixed effect of reciprocation, $b = 3.55$, $SE = 0.35$, $t(1576.88) = 10.03$, $p < .001$, which demonstrates that participants invested approximately 3.55 dollars more following reciprocation compared to trials following no reciprocation. This confirms that the participants were responsive to their partner's reciprocation behavior and adjusted their subsequent investments accordingly.

Additionally, we conducted a separate analysis on participants' trial-wise investment behavior using a repeated-measures ANOVA of 2 (reciprocation: reciprocated versus did not reciprocate) × 6 (partner emotional expression: anger, disgust, happiness, fear, neutral, sadness). Consistent with the previous model, we again found a significant main effect of reciprocation on investment decisions, $F(1,8) = 25.59$, $p < .001$, further supporting that reciprocation increased subsequent investments. However, unlike earlier analyses that revealed significant interactions between emotional expressions and demographic variables, this ANOVA did not find a main effect for emotional expression alone, $F(5,40) = 0.94$, $p = .47$, and no significant interaction be-

tween emotional expression and reciprocation. Furthermore, sex at birth and gender identity did not significantly moderate how participants responded to reciprocation (all interaction p-values $> .05$).

Hypothesis 2: Investments will be greater on trials after viewing a positively-valenced facial expression of the partner

In further assessing the LME conducted to test Hypothesis 1, the fixed effect of partner facial expression was not significant for the positively-valenced happiness condition, b = 0.55, SE = 0.44, t(1576.63) = 1.26, p = .21, suggesting that facial expression was only a single factor in investment decisions.

Interestingly, there was a significant interaction between reciprocation and partner facial expression (disgust), b = -1.08, SE = 0.49, t(1576.60) = -2.20, p = .028. This signifies that participants were less likely to increase their investment after a reciprocated exchange when their partner expressed disgust, implying that disgust dampens the reinforcement effects of betrayal and reciprocity.

The random intercept variance for the participants was 2.75 (SD = 1.66), and the residual variance was 5.68 (SD = 2.38). Model comparison indicated that including random intercepts for participants significantly improved model fit compared to a fixed-effects-only model, $\chi^2(1) = 5.12$, p = .024. Marginal $R^2 = 0.24$, and Conditional $R^2 = 0.52$, indicating that the fixed effects explained 24% of the variance, while the full model (including random effects) explained 52%. Residual diagnostics confirmed that the model assumptions were met.

Memory Data

There was a moderate positive correlation (r = 0.53) between the participants' trustworthiness ratings of the partners and their estimates of the partners' reciprocation rates, which were each collected immediately following the completion of the trust game. This result indicates that participants who rated their partners as more trustworthy also tended to remember or estimate higher reciprocation rates from those partners. This correlation analysis included all partners (both those participants actually interacted with and those they rated without interaction), but excluded any participants who indicated disbelief that their partners were real (rated 7 on a 7-point Likert scale).

Change Across the Block

To examine whether participants' investments changed systematically over time within each partner interaction (block), we tested whether the trial number predicted investment decisions. Specifically, we ran an LME with the investment amount as the dependent variable and the trial number as a fixed effect predictor. This approach allowed us to assess whether participants tended to invest more (or less) as they gained experience with a given partner across the 15 trials in each block. In addition to the trial number, we include the participant as a random effect to account for repeated measures and individual differences in overall investment behavior. This means that the model allowed each participant to have their own baseline level of investment while estimating an overall trend across trials. Additional predictors were not included in this model.

The analysis revealed a significant positive effect of trial number on investment, b = 0.10, $p < .001$, indicating that participants' investments increased by approximately 0.10 dollars for each additional trial. This result indicates that as participants continued to interact with a partner throughout the course of a block, they gradually increased their investments, which could reflect growing trust, increased cooperation, or learning about the partner's behavior. This pattern of responding is appropriate considering that trustee reciprocation rates were experimentally set at 80%.

Discussion

This study explores how trust is built and maintained in human-agent interactions using a dyadic investment game. Our methodology differs from previous design paradigms in that prior research primarily analyzes collections of existing data or a reduced set of basic emotional expressions, such as happiness or sadness. In contrast, our study leverages iMotions to record real-time emotional facial analysis and time match to behavioral decisions, while simultaneously examining betrayal, agency, and a broad spectrum of expressions - which are highly relevant factors in understanding scams involving digital agents, particularly those that exploit older adults through emotionally manipulative, yet deceptively human-like cues.

By controlling for agent expressions and behavior, we can more effectively isolate human behavioral responses to these conditions and provide results that inform the design of digital agents for the benefit of human-computer interactions. Our findings reveal a significant and expected correlation between trustworthiness ratings and participants' investment decisions, in which prior work demonstrated perceived trustworthiness as a critical factor in economic decision making. Furthermore, our results indicate that the behavioral trust responses of human participants to reciprocation and betrayals were modulated by seeing an image of their respective partner expressing disgust. Specifically, this finding indicates that participants' observing disgust from their partner following a reciprocation reduced their investment on the next trial, but observing disgust from their partner following a betrayal (failure to reciprocate) led to relatively greater investments on the next trial. This outcome is consistent with previous research that stipulates that emotional expressions play a critical role in social decision-making. Our study differed from prior work insofar as it did not indicate a similar effect for emotional expressions other than disgust. Expressions of disgust are often displayed as signals of moral or social disapproval. As such, these expressions appeared to reduce evaluations of trust when coupled with a reciprocation. In contrast, when partners failed to reciprocate and expressed disgust, this increased participants' investments, consistent with a response to initiate relationship repair. These findings further emphasize the importance of emotional cues in shaping human perceptions of digital agents and elucidate some stimulus to avoid.

Furthermore, these results highlight the dynamic and context-dependent nature of trust and the need for paradigms

that can model these context-dependent effects. Although participants exhibited a general tendency to adjust their investment behavior in favor of perceived trustworthiness, the pronounced effect of disgust expressions implies that negative emotions may have an out-sized influence on trust decisions. This highlights the need for digital agents to manage the emotional expressions that are presented and how to maintain human trust.

Limitations

There were a set of notable limitations in our study. First, reliance solely on iMotions for facial expression analysis (FEA) presents a risk to convergent validity, as the lack of alternative software for comparison may introduce inconsistencies in expression interpretation. In future studies we will run varied sources of FEA and cross-compare.

Additionally, internal validity was affected by laboratory conditions. Notably including glare from participants' glasses that distorted the FEA readings, and the use of two distinct webcams, recording at the same resolution but potentially resulting in small variations in data quality.

Finally, the use of static images from a preexisting dataset introduced a potential compromise to ecological validity by reducing the perceived authenticity of the scenario. To address this concern, a cover story was implemented to enhance believability, and data from participants who reported disbelief in the authenticity of their interaction partner were excluded from the analysis.

Future Work

This research and the methodology developed herein is our first step in the further analysis of human participants and reactions to specific emotional expressions. Our intention is two-fold.

First, we also recorded the emotional expressions of human participants as they engage with the investment game and the digital agents. These data will enable us to directly analyze the expressive behavior of the human in response to the controlled expressions of the agent, which will supply a corpus of data on human expressions and predictive behaviors that could a model then be trained on for live affective interpretation towards immediate behavioral adaptations.

Second, we will leverage this framework to examine the behavior of older adult participants and their ability to effectively assess emotional expressions displayed by partners in context, as well as those that are in conflict with the context surrounding those expressions. Do preexisting trust or baseline trust override context-dependent expressions when presented with facial expressions of persons with whom they already had a prior trust relationship? These questions are of particular interest due to the heightened vulnerability of older adults to various scams which can be perpetuated by bad actors, especially those powered by emotionally-cognizant agentic tools (Ebner, Pehlivanoglu, and Shoenfelt 2023).

Finally, can we use models based on these trust behaviors to examine the evolving cognitive process and effectively predict cognitive decline? Specifically, can these models be used to help reduce vulnerability in older adults by training this cohort how to assess both situational context and partners' expression for a more accurate understanding of partner behavior, in particular, on reducing susceptibility to scams and fraud committed by trusted family members or by a malicious agent acting utilizing voice or visual mimicry?

Ethical Impact Statement

This research adheres to the highest ethical standards in the design, implementation, and dissemination of its findings. All experimental protocols were reviewed and approved by the UCF Institutional Review Board before participant recruitment and user studies. Participants granted their fully informed consent, acknowledging their understanding of the procedures, including the recording of their data during the study. They were informed of their rights, the purpose of the research, and the handling of their data following the study.

While all participants were advised that they were playing against human partners, they were fully debriefed following their completion and informed that the partner was consistently a digital agent. This debriefing ensured transparency and maintained ethical standards in participant treatment.

Additionally, all individuals whose portraits were included in the study granted informed consent, acknowledging that their images could be included in future research. Their understanding and agreement were recorded to ensure ethical compliance.

To ensure privacy and confidentiality, recorded data were assigned coded identifiers, making it non-identifiable. Data will not be released or shared. The anonymity of participants' was strictly maintained, and all data were securely stored. These measures comply with the ethical guidelines set by the research institution.

The research team acknowledges the potential limitations of the participant sample. As participants were recruited from the UCF student body, the sample set may bias towards a younger demographic with higher education backgrounds, which may limit the generalizability of the findings.

The social impacts of this research were carefully considered. The findings contribute to the advancement of affective computing and human-computer interaction. Potential misuse or unintended consequences were evaluated and safeguards were implemented to mitigate risks.

Our research team remains steadfastly committed to transparent communication of findings and to promoting ethical debate. This statement reflects our dedication to conducting ethical research that benefits society while minimizing harm and protecting the dignity and privacy of all participants.

Acknowledgments

We express our gratitude to Noah Ari for his valuable contributions in editing this manuscript. We also extend our sincere thanks to Sophia Sakakibara Capello for her assistance with the literature review. Their support and expertise have greatly enhanced the quality of this work.

References

Avradinis, N.; Panayiotopoulos, T.; and Anastassakis, G. 2013. Behavior believability in virtual worlds: agents acting when they need to. *SpringerPlus*, 2(1): 246.

Becchetti, L.; and Degli Antoni, G. 2010. The sources of happiness: Evidence from the investment game. *Journal of Economic Psychology*, 31(4): 498–509.

Chang, L. J.; Doll, B. B.; van't Wout, M.; Frank, M. J.; and Sanfey, A. G. 2010. Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive psychology*, 61(2): 87–105.

Dang, Q.-V.; and Ignat, C.-L. 2016. Computational trust model for repeated trust games. In *2016 IEEE Trustcom/BigDataSE/ISPA*, 34–41. IEEE.

De Gelder, B.; and Hadjikhani, N. 2006. Non-conscious recognition of emotional body language. *Neuroreport*, 17(6): 583–586.

de Melo, C. M.; Gratch, J.; and Carnevale, P. J. 2013. The effect of agency on the impact of emotion expressions on people's decision making. In *2013 Humaine association conference on affective computing and intelligent interaction*, 546–551. IEEE.

Ebner, N. C.; Pehlivanoglu, D.; and Shoenfelt, A. 2023. Financial fraud and deception in aging. *Advances in geriatric medicine and research*, 5(3): e230007.

Ekman, P.; and Friesen, W. V. 1976. Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1): 56–75.

Ekman, P.; Friesen, W. V.; and Tomkins, S. S. 1971. Facial affect scoring technique: A first validity study. *Semiotica*.

Gneezy, U.; Güth, W.; and Verboven, F. 2000. Presents or investments? An experimental analysis. *Journal of Economic Psychology*, 21(5): 481–493.

Hoegen, R.; Stratou, G.; and Gratch, J. 2017. Incorporating emotion perception into opponent modeling for social dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 801–809.

Hoegen, R.; Van Der Schalk, J.; Lucas, G.; and Gratch, J. 2018. The impact of agent facial mimicry on social behavior in a prisoner's dilemma. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 275–280.

Hula, A.; Moutoussis, M.; Will, G.-J.; Kokorikou, D.; Reiter, A. M.; Ziegler, G.; Bullmore, E.; Jones, P. B.; Goodyer, I.; Fonagy, P.; et al. 2021. Multi-round trust game quantifies inter-individual differences in social exchange from adolescence to adulthood. *Computational Psychiatry*, 5(1): 102.

Jordan Schotz, N. L., Haily Follese. 2025. VAFSS (Virtual Avatar Facial Stimuli Set)—a database of face stimuli and corresponding computerized representations: Development and validation.

Kulke, L.; Feyerabend, D.; and Schacht, A. 2020. A comparison of the Affectiva iMotions Facial Expression Analysis Software with EMG for identifying facial expressions of emotion. *Frontiers in psychology*, 11: 329.

Lei, S.; and Gratch, J. 2023. Sources of facial expression synchrony. In *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–8. IEEE.

Lucas, G. M.; Gratch, J.; King, A.; and Morency, L.-P. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37: 94–100.

Manstead, A.; and Fischer, A. H. 2001. Social appraisal. *Appraisal processes in emotion: Theory, methods, research*, 221–232.

Matsumoto, D.; and Hwang, H. C. 2014. Judgments of subtle facial expressions of emotion. *Emotion*, 14(2): 349.

Mehu, M.; Little, A. C.; and Dunbar, R. I. 2007. Duchenne smiles and the perception of generosity and sociability in faces. *Journal of Evolutionary Psychology*, 5(1): 183–196.

Phan, K. L.; Sripada, C. S.; Angstadt, M.; and McCabe, K. 2010. Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences*, 107(29): 13099–13104.

Schneider, J. N.; Matyjek, M.; Weigand, A.; Dziobek, I.; and Brick, T. R. 2022. Subjective and objective difficulty of emotional facial expression perception from dynamic stimuli. *Plos one*, 17(6): e0269156.

Schneider, K. G.; Hempel, R. J.; and Lynch, T. R. 2013. That "poker face" just might lose you the game! The impact of expressive suppression and mimicry on sensitivity to facial expressions of emotion. *Emotion*, 13(5): 852.

Shore, D.; Robertson, O.; Lafit, G.; and Parkinson, B. 2023. Facial regulation during dyadic interaction: Interpersonal effects on cooperation. *Affective Science*, 4(3): 506–516.

Stratou, G.; Van Der Schalk, J.; Hoegen, R.; and Gratch, J. 2017. Refactoring facial expressions: An automatic analysis of natural occurring facial expressions in iterative social dilemma. In *2017 Seventh international conference on affective computing and intelligent interaction (ACII)*, 427–433. IEEE.

Torre, I.; Goslin, J.; and White, L. 2020. If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior*, 105: 106215.

Van Kleef, G. A.; De Dreu, C. K.; and Manstead, A. S. 2010. An interpersonal approach to emotion in social decision making: The emotions as social information model. In *Advances in experimental social psychology*, volume 42, 45–96. Elsevier.

Vinciarelli, A.; Esposito, A.; André, E.; Bonin, F.; Chetouani, M.; Cohn, J. F.; Cristani, M.; Fuhrmann, F.; Gilmartin, E.; Hammal, Z.; et al. 2015. Open challenges in modelling, analysis and synthesis of human behaviour in human–human and human–machine interactions. *Cognitive Computation*, 7(4): 397–413.