Trust and Deception: The Role of Robot Voice Gender in Strategic Gameplay

Tongwen Zheng, Xiaopeng Zhang, Wanyun Song

1 Research Question

How does the gender of a robot's voice (male vs.female) impact a human's trust in the robot, confidence in their decisions, and willingness to engage in deception during a simplified Texas Hold'em poker game?

2 Game Design

- Simplified Texas Hold'em: All community cards are revealed at once, and there is only one bidding phase.
- Bidding rules: One player bids first. The other may follow, raise, or fold. If raised, the first must follow or fold. Then, cards are revealed.
- Robot deception: 50% of rounds to bluff using facial expressions or speech inconsistent with its hand.
- Human deception: Players may also use deceptive cues during gameplay.

3 Constructs You Intend to Investigate

- Trust
- Deception Detection
- Deception Engagement
- Risk-Taking
- Confidence in Decisions

4 Hypotheses

- **H1 (Trust Bias):** Participants will follow bids more and report higher trust for female-voiced robots.
- **H2 (Deception Susceptibility):** Participants will be more deceived by female-voiced robots.

- **H3 (Deception Engagement):** Participants will bluff more frequently against male-voiced robots.
- **H4 (Risk Assessment):** Male-voiced robots will elicit riskier strategies from humans (higher bids).
- **H5 (Confidence in Decisions):** Participants will report higher confidence against female-voiced robots.

5 Experimental Conditions

- **Voice Gender Condition:** The robot will use one of two voice genders:
  - male voice,
  - female voice.

  Each voice will be designed to clearly represent its gender category while maintaining similar speech patterns, vocabulary, and intonation patterns.
- **Robot Deception:** The robot will engage in deceptive behavior:
  - In 50% of rounds, displaying facial expressions that contradict its actual hand quality (showing happy expressions with poor cards, or sad expressions with good cards);
  - In the other 50% of rounds, the robot will display honest behaviors that match its hand quality.
- **Hand Quality:** Participants will be dealt predetermined hands of varying quality
  - higher than robot's
  - the same as robot's
  - lower than robot's

  to control for the effect of actual card value on bidding behavior.
- **Game Rounds:** Participants play 2 sets (6 rounds each), one set with a male-voiced robot, another with a female-voiced robot.
- **Chips:** Each player starts with 18 chips per set.

6 Experimental Design

Design: We intend to conduct within-subjects experiment. To mitigate the confounding effect, we assign participants to random orders of gender voice, and we give participants short break between switching gender.

7 Measures

7.1 Measurement Strategy:

- **Voice Gender Perception:** Post-game Likert scale on perceived voice gender.
- **Robot Deception (Truth/Bluff):** Tracked by the system per round (50% deceptive).
- **Hand Quality:** Pre-assigned each round.
- **Trust(H1):** Likert-scale ratings after each round; bid-following rate.
- **Deception Susceptibility (H2):** % of bids increased in response to robot bluffs.
- **Risk-Taking (H3):** Average bid size by voice condition.
- **Deception Engagement (H4):** Number of participant-initiated deceptive actions.
- **Confidence (H5):** Self-reported decision confidence after each round.
- **Win/Loss:** Number of chips in hands.
- **Strategy Differences:** Open-ended self-report post-game.

7.2 Manipulation check:

- **Voice Gender Perception Test:** Participants will rate each voice on gender identification scales to confirm they correctly perceive the intended gender of each voice.
- **Deception Awareness:** Post-experiment assessment of whether participants detected the robot's systematic deception patterns during gameplay.

7.3 Subjective measures:

- **Robot Perception Questionnaire:** We will employ the Anthropomorphism and Likeability subscales from Bartneck et al. (2009) to assess participants' perceptions of the robot's human-likeness and appeal.
- **Trust in Robot Scale:** We will use the validated Trust in Automation Scale by Jian et al. (2000), adapted for human-robot interaction. This 12-item scale includes statements such as "The robot is deceptive" and "I can trust the robot" rated on a 7-point Likert scale.
- **Game Strategy Questionnaire:** Participants will complete a custom questionnaire about their strategic approach when playing against the robot, including items about risk assessment, deception choice and deception detection.
- **Robotic Social Attributes Scale (RoSAS):** The warmth and competence subscales from Carpinella et al. (2017) will be used to measure perceived social capabilities of the robot.

7.4 Objective measures:

- **Bid Following Rate:** Proportion of times participants matched or raised the robot's bid, used to indicate behavioral trust in the robot.
- **Deception Vulnerability Index:** Proportion of conditions where participants changed their bidding strategy in response to the robot's misleading expressions, reflecting susceptibility to the robot's bluffing behavior.
- **Risk Behavior (Average Bid):** Average number of chips wagered per round.
- **Win/Loss Record:** Total number of chips at hands.
- **Bluff Detection Accuracy:** Percentage of robot deception rounds where participants correctly identified the bluff and adjusted their strategy accordingly (e.g., folded or reduced bid).
- **Human-Initiated Deception Frequency:** Count of rounds where participants intentionally used verbal or nonverbal cues (e.g., misleading expressions or statements) inconsistent with their hand quality, to measure engagement in deceptive behavior.

7.5 Interview questions:

- "What factors influenced your decision to trust or not trust the robot's expression?"
- "How did your trust in the robot's expressions change throughout the experiment?"
- "Did you feel more competitive or confident with one robot voice?"
- "Did your strategies differ by robot?"

8 Procedure

- **Introduction:** Participants are briefed on simplified Texas Hold'em rules and robot behavior. They engage in a short play simulation to become familiar with the gameplay dynamics. Robot competitors are introduced and presented as seemingly autonomous.
- **Set 1 (6 rounds):** Participants play six rounds against either a male-voiced or female-voiced robot (voice condition is randomized across participants).
  - Each round consists of one bidding phase only.
  - All three community cards are revealed simultaneously.
  - The robot exhibits deceptive behavior (misleading expressions or speech) in 50% of the rounds, randomized across trials.
- **Break:** Participants take a short rest to minimize fatigue and reset expectations before continuing.
- **Set 2 (6 rounds):** Participants complete six additional rounds, now with the opposite robot voice condition (i.e., female if male was first, or vice versa).

- **Post-Game:** Participants complete the following:
  - Standardized questionnaires assessing trust, confidence, perception of deception, and gender stereotypes.
  - A semi-structured qualitative interview designed to explore subjective impressions, strategies, and perceived differences between robot conditions.

9 Sample Size

The target is 34 participants (based on repeated measures ANOVA: f = 0.25, a = 0.05, power = 0.80). We plan to recruit 40 participants (accounting for possible attrition due to early chip loss or withdrawal).

10 Timeline

Time	**Task**
Week 4	Develop study protocol, questionnaires, consent forms, etc. and **submit our IRB application**
Week 5	Program the robot and go back-and-forth with the IRB if they request any changes to your IRB application
Week 6	Pilot the study
Weeks 7–8	Run human subjects
Week 9	Conduct data analysis
Week 10	Presentation and final report

Table 1: Project Timeline