

REVIEW PAPER

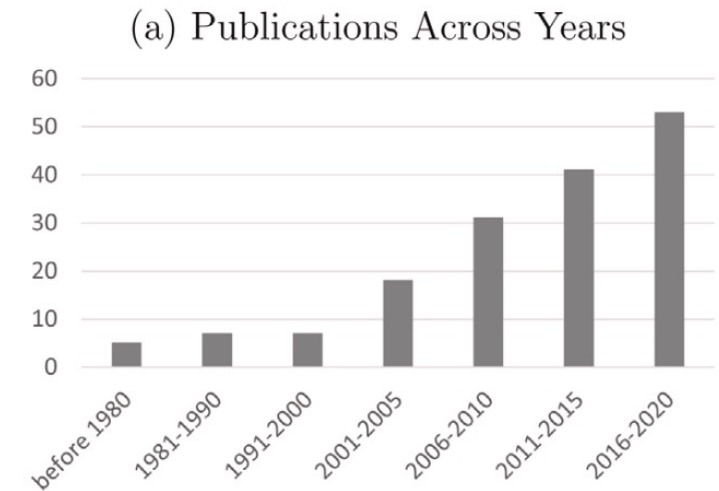
Strategic interactions between humans and artificial intelligence:
Lessons from experiments with computer players

Christoph March (2021)

Presented by Michelle Lee

Review Paper Outline

Aim: to conduct a literature review of the experimental studies that examine the strategic interaction of humans with AI



- **Why is it relevant?**
 - **Increasing interactions with AI in the world**, e.g., humans vs. algo-traders in financial markets
 - Growing use of AI-related decision aids, e.g., robo-advisor for investment advice
 - Greater focus on the transparency and explainability of AI
- **Common findings of human-human interaction:**
 - Humans tend to underestimate other humans' rationality causing them to e.g., learn too little from others' actions
 - Humans tend to have social preferences when interacting with other humans, e.g., showing reciprocity or altruism

Key findings

1. **Humans behave differently when interacting with computer players (CP)**

- They adapt to CP even when information about the CPs is limited -> they can exploit CP with simple algorithms but are exploited by CP with more sophisticated algorithms
- They apply different cognitive processes when CP is present -> they behave more rationally, more selfishly and less emotionally, but CP can also be designed to induce cooperation

2. **Consequently, outcomes of strategic games differ from those with only human players**

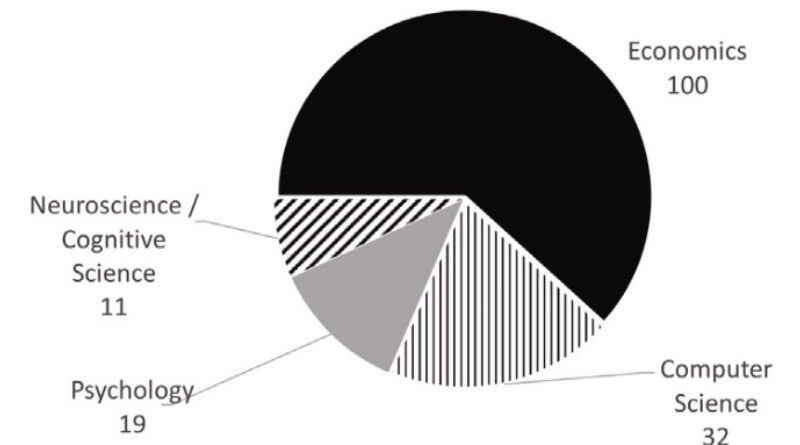
- CPs may enhance efficiency of the experimental task at hand (e.g., improve price convergence in markets, assist cooperation in strategic games, and improve bargaining outcomes)

3. **The impact of CP strongly depends on (the interaction of) the strategy implemented by the CP, the information given to the participants, the strategic setting, and the participants' cognitive ability to understand the task and the strategy implemented by the CP .**

Studies reviewed

- 162 experimental studies using CP in economics, computer science, neuroscience and psychology.
- Exclusion criteria:
 - Studies that use deception, e.g., told participants that their partner is a human but actually a CP or participants not aware of the presence of CP
 - Studies with no strategic interdependence between the human and CP (e.g., interaction with chatbots, use of algorithmic advice)
 - Studies not grounded in economic theory (e.g., online shopping)
 - Studies that don't explicitly refer to the CP

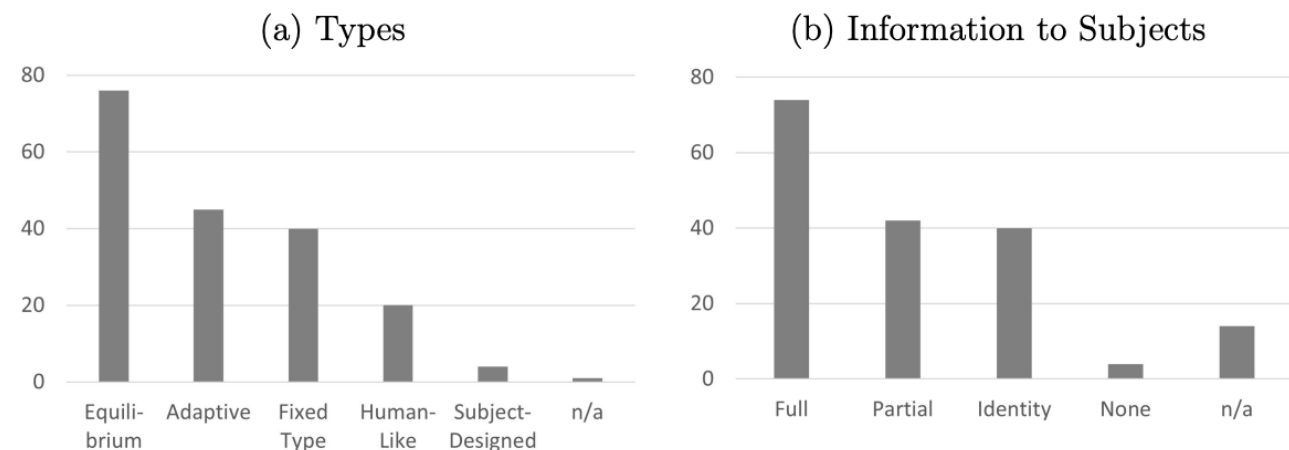
(b) Research Fields



Using CP in experiments

1. Type of CP: equilibrium, fixed, adaptive or human-like
 - Most studies used simple/equilibrium strategies
2. Information to participants: trade-off between experimental control and external validity, also depends on participants' ability to understand the CP
 - Most studies reveal the entire strategy of the CP to participants
3. Treatment comparison: only around half of the studies compare human-AI interaction with human-human interaction

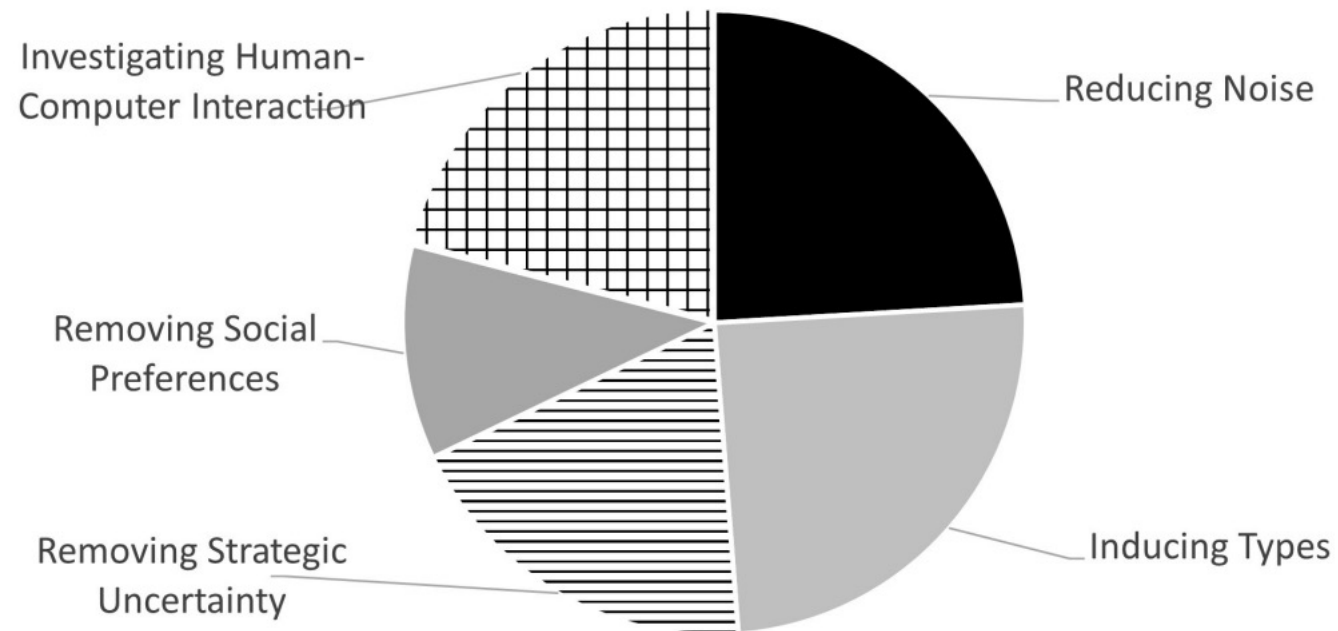
Figure 2: Methodical Aspects of Computer Players



Use of CP w.r.t. research objectives (Pt. 1)

-> mainly been used to increase experimental control

Figure 1: A Classification of Research Objectives



Use of CP w.r.t. research objectives (Pt. 2)

-> mainly been used to increase experimental control

1. Reducing noise in the experimental environment

- E.g., use robot-buyers in a market setting to observe how human sellers set prices or advertising level -> avoid confounding factors from the buyers' side and simplifies the experimental environment for participants
- According to the author, this also *reduces the number of participant payments*, simplifies instructions and increases the statistical power of tests

2. Inducing specific behavioural types

- E.g., let CP adopt a simple strategy, and examine how participants respond to that simple strategy -> do participants understand the CP's strategy and exploit it?

3. Removing strategic uncertainty

- E.g., use CP to implement a fixed and known strategy in order to induce a particular equilibrium outcome -> the effect of removing strategic uncertainty is heavily context-dependent
- Interestingly, one study showed that participants prefer uncertainty generated by computers to uncertainty generated by humans (perhaps it's perceived to be less volatility?)

Use of CP w.r.t. research objectives (Pt. 3)

-> mainly been used to increase experimental control

4. Removing social preferences as a key force of behaviour

- E.g., participants face a CP in a bargaining game -> participants are less likely to reward/punish their opponent and behave more rationally and selfishly
- May be difficult to untangle the effects of strategic uncertainty and social preference
- Social preference may persist when the CP is acting on behalf of another human player

5. Investigating human-AI interactions

- E.g., whether and how market converge to the equilibrium when participants interact with algo-traders in the market -> increases market efficiency
- E.g., how changing the identity of the player to CP affects participants' behaviour in the game -> participants become more rational and less emotional when CP is present
- E.g., whether humans can cooperate with CP -> this depends on the information about the CP that is available to the participants
- E.g., let participants design a CP to play on their behalf -> CP performs better than their designers, possibly because it reduces biases? (future research opportunity?)

Future Research

- Most studies that focus on strategic interaction of humans with AI use CP with simple/equilibrium strategies
-> not realistic
- There's a focus on explainable AI: "it should be possible to demand a suitable explanation of the AI system's decision-making process" (European Commission, 2019)
 - This depends on the technical abilities of the AI, the amount of information presented, the way the information is presented, and the information processing capabilities of the users
- AI may require personal information about user and thus affects the outcome of strategic game
 - Participants' willing to share personal information with the CP will affect the outcome of strategic game
- Ultimately, if humans do, in fact behave differently (e.g., more rationally) in the presence of AI, then it would be important to incorporate such behavioural change in economic models in the future