Where you look in the past biases your social inference

Sangkyu Son (ss.sangkyu.son@gmail.com)
Department of Biomedical Engineering, Sungkyunkwan University
Suwon, Republick of Korea

Seng Bum Michael Yoo (sbyoo@g.skku.edu)
Department of Biomedical Engineering, Sungkyunkwan University
Suwon, Republick of Korea

Abstract

Humans are adept at inferring others' intentions by watching their actions. What do they observe while making an inference, and how would observation influence the inference? In this study, participants played an interactive pursuit game with a computerized opponent whose hidden intentions varied. Participants formed negatively biased inferences after the opponent behaved in an unexpectedly unfriendly way. However, it did not occur after unexpectedly helpful behavior. Such asymmetrical inference persisted even after the opponent's unfriendly behavior ceased, as long as it had been building up over time—a phenomenon of history and path dependence known as hysteresis. The inferential biases were associated with gaze selection: participants looked more at the unexpectedly unfriendly opponent, even though it increased variability in their control. These findings extend action-based inference models by incorporating what we choose to observe and accounting for asymmetrical hysteresis in social understanding.

Keywords: Social inference, observation, hysteresis, energy landscape

Results

Forty-four participants played an interactive game where they inferred the friendliness of a computerized opponent and decided whether to speed it up or slow it down while chasing prey (Figure 1A). Unbeknownst to them, the opponent's friendliness could change after the decision, becoming less friendly (betrayal) or more friendly (unexpected help).

After betrayal, participants consistently inferred the opponent as unfriendly compared to the veridical friendliness, unlike unbiased inference after unexpected help (Figure 1C). We modeled inference as a ball rolling on a landscape: highly variable inferences sit in high-energy regions, while low-variable ones settle into low-energy valleys (Figure 1D). We found that betrayal shifted the inference landscape toward unfriendly interpretations, leading to different final inferences from the identical opponent behavior (Figure 1E). This suggests that prior betrayal experiences can leave lasting, history-dependent effects on social inference.

We next examined observational focus. During betrayal, participants shifted gaze to the opponent, unlike in unexpected help (Figure 2A). This gaze shift impaired the control of the player's avatar (Figure 2B) and predicted stronger unfriendly bias in the next trial (Figure 2C). These results suggest that observation, despite its cost of losing control, contributes directly to inferential bias.

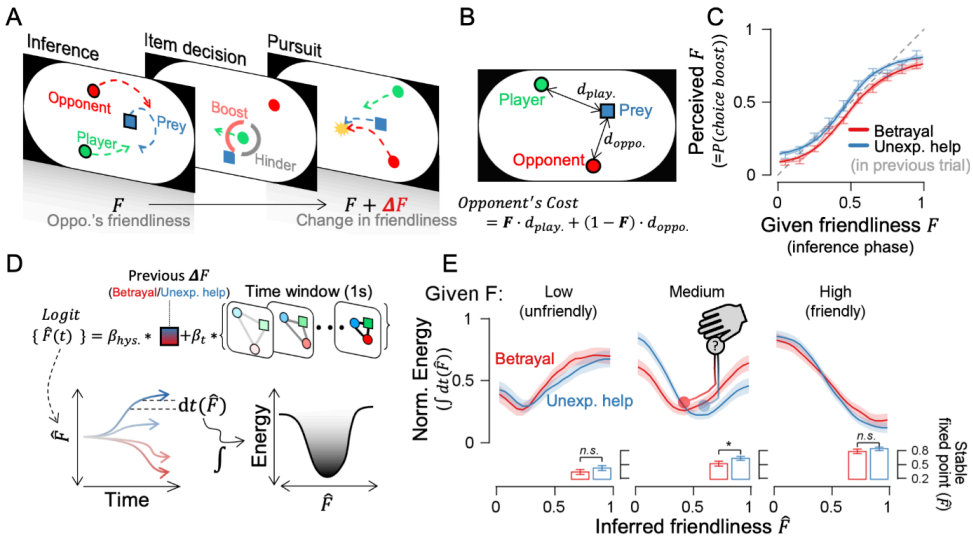

Figure 1. Asymmetrical biases in inference. (A) Participants watched a video of their avatar and a computer opponent pursuing prey. Opponent's hidden intention (F) ranged from competitive to cooperative. Based on the movie, participants either boosted or hindered the opponent's speed. Then, they pursued the prey by themselves. A decrease in F after the decision indicated betrayal; an increase, unexpected help. **(B)** At F = 1, it herded prey toward the player; at F = 0, it intercepted prey competitively. **(C)** Psychometric curves show reduced boost choices after betrayal, but no change after unexpected help. **(D)** A logistic regression estimated the evolution of inference over time, using the characters' trajectories and the recent experience of ΔF. The energy landscape was reconstructed from the changes in inferred friendliness. **(E)** Reconstructed energy landscapes from data. The subfigures show the stable fixed point, the minimum energy point where the inference ultimately settles.

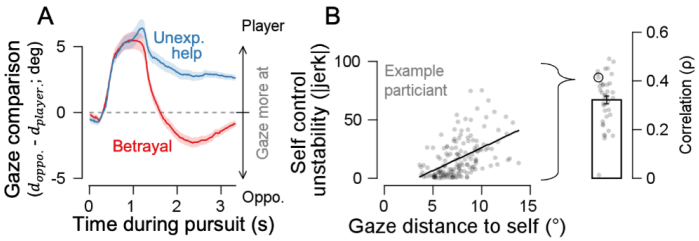

Figure 2. Observation changes inference. **(A)** Gaze during pursuit was measured as the difference in distance between the gaze to the opponent and the player. **(B)** Looking away from one's avatar reduced control stability, reflected in increased jerk. **(C)** Correlation between gaze at the opponent and bias in perceived F-value on the next trial.

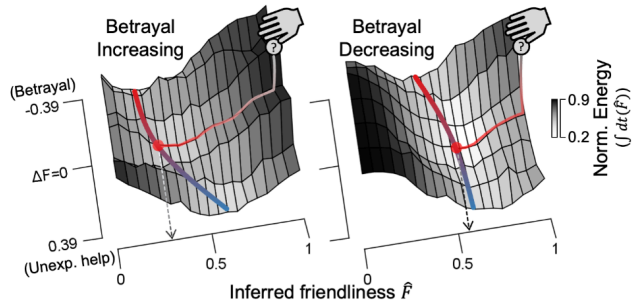

Figure 3. Reconstructed energy landscapes from data over various levels of betrayal intensity in the previous trial. Minima are indicated by lines along the basin. Negative inference biases were stronger during phases of increasing betrayal, even when the previous trial involved no betrayal, as shown by identical ball drops settling in different positions.

Finally, we examined whether inference is not only history-dependent but also the path of history. When the energy landscape (from Figure 1E, middle panel) was separated by trials ranging from the strongest betrayal to the strongest unexpected help, the structure of the inference landscape differed depending on whether betrayal had been increasing or decreasing across prior trials (Figure 3). Specifically, the lowest points (valley) of the energy landscape—representing the final inference—shifted to more negative interpretations only during betrayal increasing phases. This indicates that inference is path-dependent—that is, it reflects not just what was experienced, but how it has been unfolded over time. Such dynamics resemble hysteresis, a phenomenon observed in physical systems, and here extended to social inference.

Discussion

As illustrated in Figure 4, we reconceptualize social inference as a bidirectional process between self and others, where observational selection determines which aspects of the environment are prioritized. Traditional models assume agents have full access to others' actions, implicitly treating observation as passive and complete (Baker, Saxe, and Tenenbaum 2009; Jara-Ettinger 2019). In contrast, we show that limited observability is strategically resolved by prioritizing previously relevant information. Our findings reveal that betrayal experiences shift observation from self-relevant cues toward others' actions, aiming to reduce uncertainty about others' intentions. This selective observation, in turn, shapes the evidence available for future inference. The current framework unifies previously unaccounted aspects, including information value shaping observational learning (Grabenhorst et al. 2019) and social gaze (Dal Monte et al. 2022; McMahon and Isik 2023). This was possible by adopting a dynamical systems perspective, which captures the history-dependent and continuous nature of social inference.

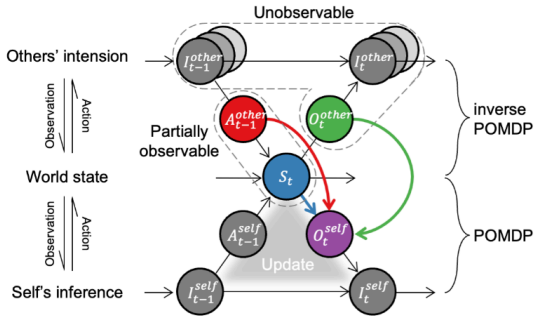

Figure 4. Social Inference is bidirectional between the self and others. Self-world interaction is modeled as a POMDP, a partially observable Markov decision process, while inferring others is treated as an inverse POMDP. The agent knows its own observations and actions but not the full world state, especially others' intentions or perceptions. When others strongly affect the world, the agent focuses more on their actions (red) and observations (green) than the external state (blue), modifying inference in the next time step.

Acknowledgement

This research was supported by RS-2023- 00211018 (SS and SBMY). The authors declare no competing financial interests.

References

Baker, Chris L., Rebecca Saxe, and Joshua B. Tenenbaum. 2009. “Action Understanding as Inverse Planning.” Cognition 113 (3): 329–49.

Dal Monte, Olga, Siqi Fan, Nicholas A. Fagan, Cheng-Chi J. Chu, Michael B. Zhou, Philip T. Putnam, Amrita R. Nair, and Steve W. C. Chang. 2022. “Widespread Implementations of Interactive Social Gaze Neurons in the Primate Prefrontal-Amygdala Networks.” Neuron 110 (13): 2183–97.e7.

Grabenhorst, Fabian, Raymundo Báez-Mendoza, Wilfried Genest, Gustavo Deco, and Wolfram Schultz. 2019. “Primate Amygdala Neurons Simulate Decision Processes of Social Partners.” Cell 177 (4): 986–98.e15.

Jara-Ettinger, Julian. 2019. “Theory of Mind as Inverse Reinforcement Learning.” Current Opinion in Behavioral Sciences. Elsevier Ltd. https://doi.org/10.1016/j.cobeha.2019.04.010.

McMahon, Emalie, and Leyla Isik. 2023. “Seeing Social Interactions.” Trends in Cognitive Sciences, October. https://doi.org/10.1016/j.tics.2023.09.001.