Cognitive Psychology 151 (2024) 101654

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Repeated rock, paper, scissors play reveals limits in adaptive sequential behavior

Erik Brockbank [a,*], Edward Vul [b]

[a] Stanford University, United States of America
[b] University of California San Diego, United States of America

ARTICLE INFO

Keywords:
Adaptive reasoning
Adversarial reasoning
Opponent modeling
Rock–paper–scissors

ABSTRACT

How do people adapt to others in adversarial settings? Prior work has shown that people often violate rational models of adversarial decision-making in repeated interactions. In particular, in *mixed strategy equilibrium (MSE)* games, where optimal action selection entails choosing moves randomly, people often do not play randomly, but instead try to *outwit* their opponents. However, little is known about the adaptive reasoning that underlies these deviations from random behavior. Here, we examine strategic decision-making across repeated rounds of rock, paper, scissors, a well-known MSE game. In experiment 1, participants were paired with bot opponents that exhibited distinct stable move patterns, allowing us to identify the bounds of the complexity of opponent behavior that people can detect and adapt to. In experiment 2, bot opponents instead exploited stable patterns in the human participants' moves, providing a symmetrical bound on the complexity of patterns people can revise in their own behavior. Across both experiments, people exhibited a robust and flexible attention to *transition patterns* from one move to the next, exploiting these patterns in opponents and modifying them strategically in their own moves. However, their adaptive reasoning showed strong limitations with respect to more sophisticated patterns. Together, results provide a precise and consistent account of the surprisingly limited scope of people's adaptive decision-making in this setting.

1. Introduction

People's ability to adapt to others in adversarial interactions lies at the heart of sports and games and is a hallmark of popular stories; at a larger scale, it is crucial to negotiations and international relations. In these settings, examples of people's creativity, flexibility, and strategic sophistication abound. For instance, tennis star Andre Agassi famously beat world-class opponent Boris Becker by recognizing that every time he served the ball, Becker unknowingly stuck his tongue out in the direction he was about to serve.[1] On the other hand, adversarial reasoning often poses a number of challenges such as remembering previous decisions (Rapoport & Budescu, 1997), recursive reasoning about others (Stahl & Wilson, 1995), or planning many steps ahead (van Opheusden, Kuperwajs, Galbiati, Bnaya, Li, & Ma, 2023). These challenges have allowed artificial intelligence systems to beat human competitors in a range of adversarial settings, even those once thought to be far beyond the reach of strategic algorithms (Silver et al., 2016). *How do people construct predictive models of an opponent in adversarial settings and what are the limitations of this ability?* In the current work, we investigate a fundamental component of this problem: the ability to recognize exploitable patterns

* Correspondence to: 450 Jane Stanford Way, Building 420, Stanford, CA 94305, United States of America.
E-mail address: ebrockbank@stanford.edu (E. Brockbank).

[1] https://www.facebook.com/watch/?v=1249137535168463 – January 18, 2017.

https://doi.org/10.1016/j.cogpsych.2024.101654
Received 20 February 2023; Received in revised form 30 March 2024; Accepted 8 April 2024
Available online 23 April 2024

in an opponent's actions over time. We ask what sort of structured behavior patterns people can exploit in others and which of these same patterns people can change in their own behavior to avoid exploitation.

The problem of predicting those around us extends beyond adversarial interactions; it is a core part of everyday social behavior. To make sense of others' actions, people rely on an intuitive *theory of other minds* in which behavior is understood as arising from an actor's underlying goals, desires, and beliefs about the situation (Dennett, 1989; Goodman et al., 2006; Gopnik & Wellman, 1992; Premack & Woodruff, 1978). Recent work has argued that a core mechanism for theory of mind reasoning is a principle of *efficient action* (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017; Baker, Saxe, & Tenenbaum, 2009); we predict that others will act in ways that maximize rewards while minimizing costs and we infer their corresponding goals and beliefs on the assumption that they choose utility-maximizing actions (Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016; Jara-Ettinger, Schulz, & Tenenbaum, 2020). While this work has primarily involved predicting and reasoning about behavior in *cooperative* settings, the assumption that others are approximately rational planners can also facilitate determining who is a friend and who is a foe in settings where helping and hindering are both possible (Kleiman-Weiner, Ho, Austerweil, Littman, & Tenenbaum, 2016; Serrino, Kleiman-Weiner, Parkes, & Tenenbaum, 2019; Ullman et al., 2009).

However, adversarial contexts in which one person seeks to gain at the other's expense pose a challenge for traditional theory of mind reasoning. In its simplest form, this challenge is often explored using *zero-sum* economic games (Schelling, 1960). In these settings, an actor's utility-maximizing choice is tied to what they believe their opponent will do; but the opponent's strategic choice depends in turn on the predictions they make about *their own opponent*. In cooperative tasks this reasoning may allow people to quickly align on their best option, but in adversarial interactions, this process of guessing the opponent's beliefs can be infinitely recursive (Binmore, 1987; Stahl & Wilson, 1995). A standard solution to this problem has been to propose that people rely on limited levels of recursive reasoning about an opponent's beliefs when trying to predict them (for example, what does my opponent think that *I think* is the best move?). Models in which individual decision-makers are presumed to vary in their depth of recursion capture behavior in a range of adversarial games and suggest that people rarely reason past one or two levels of recursion (Camerer, Ho, & Chong, 2003, 2004; Costa-Gomes, Crawford, & Broseta, 2001; Stahl & Wilson, 1995).

These results further suggest that when trying to anticipate an opponent's behavior, there may be *persistent* features of their adversarial reasoning such as their recursion depth that support future prediction (Guennouni & Speekenbrink, 2022; Ho, Camerer, & Weigelt, 1998). Broadly, *repeated* games, in which players face off against a stable opponent over multiple rounds, present an opportunity for each player to learn a predictive model of their opponent based on past behavior (Akata et al., 2023; Aumann, 1985; Camerer, 2011; Mertens, 1990). While repeated games have been critical to theories of human cooperation (Axelrod, 1984; Mertens, 1990; Rand & Nowak, 2013), they may also inform our understanding of adversarial opponent modeling. For example, Guennouni and Speekenbrink (2022) find evidence that people not only infer an opponent's recursion depth over repeated interactions but use this information to predict their later actions in a similar game. Further, paired with a stable opponent in games with varying reward structures (for example, Prisoner's Dilemma and Stag Hunt), people infer abstract motives such as *greed* and *risk-aversion* that allow them to predict their opponent's moves across games (van Baar, Nassar, Deng, & FeldmanHall, 2022). Thus, when trying to predict an opponent's behavior in *repeated* adversarial interactions, people may draw on generalizable features of their prior actions (such as their recursion depth or greediness) that support prediction in the current context.

However, abstractions such as recursion depth are not the only source of predictive information available to reasoners trying to *outwit* an opponent. In repeated games, a key source of predictive structure may come from sequential patterns in an opponent's actions themselves. Consider the example at the outset in which Andre Agassi learned to predict the direction of Boris Becker's serve using a reliable cue in the moments before serving. In fact, efforts to predict an opponent using sequential patterns in their behavior—and corresponding counter-measures to evade such prediction—are widespread at the highest levels of competition: baseball players look for patterns in how a pitcher will throw the ball based on their previous pitches, poker players make inferences about other players' hands from their nonverbal behaviors, and hockey players watch for telltale signals that players on the other team may be nursing injuries based on the way they skate. In any setting where behavior can be predicted using patterns in prior actions, efforts to detect and exploit this signal in an opponent's behavior while avoiding it in one's own actions may be central to adversarial reasoning.[2]

This ability to learn from structured, sequential patterns has been studied in a range of domains outside adversarial reasoning about others. For instance, early language acquisition is thought to be supported by statistical learning of transition patterns in phonemes beginning as infants (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999); adults exhibit similar learning in more abstract auditory and visual domains (Frost, Armstrong, & Christiansen, 2019; Turk-Browne, Jungé, & Scholl, 2005). Given repeated practice, people can learn motor patterns consisting of as many as 10 items without advance knowledge that the stimulus contains a repeating pattern (Clegg, DiGirolamo, & Keele, 1998; Nissen & Bullemer, 1987). Most importantly, our commonsense understanding of everyday behavior seems to draw in part on observable sequential patterns. First, people have an intuitive understanding of when an action is best explained by *habit* that relies on the frequency of the behavior and whether it is adaptive to the context at hand (Gershman, Gerstenberg, Baker, & Cushman, 2016). Beyond tracking mere repetition of behavior, people are able to predict others' actions and emotional states on the basis of previous actions and emotions (Thornton & Tamir, 2017, 2021). Thus, the ability to predict actions and events in the world around us using learned sequential patterns has been

[2] A similar idea in economic policy has been expressed as *Goodhart's Law* (Courakis, 1981), which describes the phenomenon in which policies based on observed statistical regularities bring about the collapse of that regularity through exploitation; see Manheim and Garrabrant (2018) for discussion of this phenomenon in adversarial contexts.

richly investigated in a range of cognitive domains. However, the role of such inferences in adversarial reasoning remains poorly understood. What kinds of sequential patterns can people use to exploit an opponent and how well can people avoid such patterns in their own actions?

Mixed strategy equilibrium (MSE) games offer an ideal paradigm for investigating the ability to adapt to an opponent based on their sequential actions (Brockbank & Vul, 2021a; Guennouni & Speekenbrink, 2022; Shachat & Todd Swarthout, 2004; Spiliopoulos, 2013). In MSE games, each strategy or action is dominated by another; a player who persists predictably with any one action or sequence of actions can be exploited based on that predictability. For this reason, Nash Equilibrium (Nash, 1950) play in repeated MSE games entails choosing among available moves *randomly*; any non-random dependency in one's behavior is exploitable by a rational opponent. The most well-known MSE game, and the one we focus on in the current work, is the children's game of rock, paper, scissors (RPS), in which "rock" loses to "paper", "paper" loses to "scissors", and "scissors" loses to "rock". In repeated games of RPS, a player's task is to develop a predictive model of their opponent; if their opponent chooses randomly, this is (by definition) impossible in the long run. But against any non-random opponent, the ability to win systematically depends on how well the player can identify exploitable patterns in their opponent's moves while avoiding similar exploitation based on their *own* choices (Brockbank & Vul, 2021a; Guennouni & Speekenbrink, 2022).

A large body of work has sought to outline the kinds of patterned regularities people exhibit when attempting to behave randomly. When asked to generate example sequences of random events like coin tosses or evaluate sequences for their randomness, people appear to rely on simple biases like an over-representation of alternations relative to repeats (Bar-Hillel & Wagenaar, 1991; Lopes & Oden, 1987; Tversky & Kahneman, 1972). These biases of *subjective randomness* are so ingrained they arise even in decisions by professional athletes, who are highly incentivized to avoid such predictability (Palacios-Huerta, 2003; Walker & Wooders, 2001). Given this finding, it is perhaps unsurprising that in repeated MSE games, people show the same underlying biases in move selection (Budescu & Rapoport, 1994; Rapoport & Budescu, 1992). In fact, recent work has shown that the patterns exhibited by human players paired with other humans over many rounds of rock, paper, scissors extend far beyond those associated with subjective randomness; people's moves show predictable regularity based not only on their own previous moves, but on their opponent's moves and on previous outcomes (Batzilis, Jaffe, Levitt, List, & Picel, 2019; Brockbank & Vul, 2020). Because people exhibit robust sequential patterns in rock, paper, scissors and other MSE games, this offers an ideal venue in which to explore the corresponding ability to adapt to and exploit such patterns in adversarial interactions.

However, determining which patterns people detect in others' actions or their own can be hard to isolate in play between humans because of the dynamic nature of dyadic play (Spiliopoulos, 2013). Instead, repeated matches between humans and algorithmic *bot opponents* that exhibit stable patterns in their moves provide a controlled environment for testing people's ability to exploit such patterns (Brockbank & Vul, 2021a; Zhang, Moisan, & Gonzalez, 2021). For example, when paired with rock, paper, scissors opponents that favor a particular move (e.g., playing "rock" in 70% of rounds), people typically learn to exploit them so long as the bias is sufficiently strong (Kangas, Berry, Cassidy, Dallery, Vaidya, & Hackenberg, 2009; Lie, Baxter, & Alsop, 2013). Recent work has shown that people can also exploit opponents that exhibit patterns in their choices based on their own previous move or their human opponent's previous move (Guennouni & Speekenbrink, 2022; West & Lebiere, 2001). Beyond previous moves alone, people are sensitive to the role that the prior *outcome* plays in determining an opponent's moves (Dyson, Steward, Meneghetti, & Forder, 2020; Zhang et al., 2021), though adaptation to these patterns in prior work has been somewhat limited. Finally, this behavior appears to reflect *ongoing* adaptive reasoning, even as opponent strategies or the games themselves change (Guennouni & Speekenbrink, 2022; Stöttinger, Filipowicz, Danckert, & Anderson, 2014). Taken together, these results indicate that in repeated MSE games like rock, paper, scissors, people exhibit flexible strategic reasoning aimed at exploiting patterns in opponent behavior.

However, these findings paint an incomplete picture of the scope of people's adaptive reasoning about sequential opponent behavior. For one, prior research using bot opponents in MSE games has addressed different questions from the one we focus on here, such as whether responses to gains and losses show different neural and behavioral signatures (Dyson et al., 2020; Dyson, Sundvall, Forder, & Douglas, 2018; Forder & Dyson, 2016). For such questions, it has not been necessary to compare behavior against a broad swath of bot strategies and existing work has employed only one or two distinct opponent strategies. Furthermore, findings across these studies are based on varied experimental conditions, making comparison difficult. Thus, prior work has not systematically explored the range of sequential opponent behaviors that a player might adaptively respond to. Finally, exploiting patterns in an opponent's behavior is only part of the puzzle. A central challenge for players in repeated MSE games is avoiding any detectable patterns in their *own* moves. Once again, this question can be fruitfully investigated by pairing participants with bot opponents that exploit patterns in the *participants'* moves (Brockbank & Vul, 2021b; Eyler, Shalla, Doumaux, & McDevitt, 2009; Moisan & Gonzalez, 2017; Spiliopoulos, 2013; West & Lebiere, 2001), yet prior work using such a paradigm has been limited and has not explained how such behavior informs broader questions about adversarial reasoning.

The current work uses rock, paper, scissors to develop a systematic and comprehensive account of how people adapt to sequential patterns in an opponent's behavior and their own. Which patterns in their opponent's actions can they successfully learn and which ones are out of reach? And how well can people avoid being similarly exploited? Rock, paper, scissors represents an idealized environment for addressing these questions. The patterns that a player exhibits in their moves can be precisely spelled out, the complexity of these patterns can be formally described, and the extent to which a player *exhibits* any given pattern in their moves can be quantified (Brockbank & Vul, 2021a; Dyson, 2019). Further, unlike other adversarial games in which people may exhibit exploitable patterns in their sequential behavior (e.g., chess), RPS involves little expertise. Instead, a player's success at exploiting patterns in an opponent's moves is a result of adaptive reasoning about the causes of their behavior within the immediate interaction context. Despite these advantages, no prior work has provided a systematic account of the patterns people can and cannot recognize and adapt to in this setting.

Here, we aim to address this shortcoming by investigating adaptive behavior over repeated rounds of rock, paper, scissors against an algorithmic "bot" opponent. In experiment 1, we pair participants with one of seven stable bots, each of which exhibits a different sequential dependency in its move choices. These dependencies vary in their underlying complexity, allowing us to precisely assess the degree to which people exploit different behavioral patterns dictating an opponent's moves. We find that people are highly adaptive against opponents that exhibit simple *transition* patterns but show minimal adaptation to more complex opponents. In experiment 2, we ask whether these same limits hold for avoiding exploitable patterns in one's *own* behavior. Participants were once again paired with a bot opponent, but this time each bot chose its moves by trying to exploit a unique pattern in the participant's moves. Here, we examine people's ability to adapt to their adaptive opponent. We find that people are successful against bots that track simple transition patterns in participant moves, but show little flexibility otherwise. Together, our results suggest that the *hypothesis space* of behavioral patterns people draw on in this setting to understand their opponent's moves or their own is limited, but that adaptive reasoning is flexible within these limits.

2. Experiment 1: Stable bot opponents

Experiment 1 pitted participants against predictable bot opponents that chose their moves by following one of seven increasingly complex sequential dependencies. If participants can reliably beat a bot opponent that exhibits a stable pattern in its moves, this suggests people are able to adapt to that particular dependency in an adversarial setting. We investigate the level of behavioral complexity that people can detect and exploit across repeated interactions with a bot opponent.

2.1. Participants

Participants were a convenience sample of 218 University of California, San Diego undergraduate students who received course credit for their participation. One student's data was removed due to technical issues during data collection which prevented completion of the experiment. Our sample size was chosen to have a minimum of 30 participants in each condition (i.e., against each bot opponent). This gave us 90% power to detect an effect size of $d = 0.61$ in our estimate of participant win rates against each bot; under a conservative assumption of uniformly distributed win rates, this effect size amounts to an average win rate of approximately 51%. Informed consent was obtained from all participants in accordance with the Institutional Review Board's approved protocol. Participants completed the experiment in a web browser online.[3]

2.2. Task overview

Participants began by clicking through a set of instructions introducing the game of rock, paper, scissors and noting that they would be playing 300 rounds against a fixed opponent (they were told it would be the same opponent the whole time but were not told anything about the opponent's identity). Upon completion of the instructions, participants were randomly assigned to one of seven bot opponent conditions, described in detail below. In each round, they were shown a set of clickable "cards" with rock, paper, and scissors icons and instructed to choose a move (Fig. 1). They were given 10 s to choose their move each round. Once a participant had chosen a move, they could not change their selection. After selecting a move, participants were shown a results screen indicating their own move, their opponent's choice, the results of the round, and the points each player received for that round. Participants were given 3 points for a win, 0 points for a tie, and −1 points for a loss. These values were chosen to maintain engagement by giving participants greater opportunity for positive accumulation of points.

Rock, paper, scissors is often formulated as offering 1 point for a win, 0 points for a tie, and −1 points for a loss. Past work has modified this reward structure to disentangle Nash Equilibrium *mixed strategy* play from random move selection (Zhang, Moisan, & Gonzalez, 2020; Zhang et al., 2021), to understand evolutionary dynamics of repeated play (Hoffman, Suetens, Gneezy, & Nowak, 2015), or to separately evaluate responses following gains and losses (Forder & Dyson, 2016). The goal of the present study is to understand the impact of different structured opponent move patterns on people's decisions, a question that does not strongly rely on any particular reward structure. In this vein, the allocation of points across outcomes in the current study did not impact the optimal strategy for participants, since there was a single move every round that would give them a high probability of winning.

Throughout the game, participants were shown a graphic illustrating each move's relation to the others, a tally of rounds completed towards the total, and the cumulative points each player had accrued so far (see Fig. 1). The displayed scores and rounds completed served to motivate participants to best their opponent over the 300 rounds, while the graphic of each card's relationship to the others minimized the risk of them misunderstanding or forgetting the rules. Participants could view the results of each round for as long as they wanted before proceeding to the next round, allowing time for them to evaluate potential patterns in their opponent's moves. Most participants completed the RPS game in under 30 min (mean time from round one to round 300: 693 s, SD: 259 s). Following completion of the game, participants completed a brief post-experiment questionnaire with a free response prompt asking about how they had chosen their moves and a set of five slider scale questions further assessing their strategic decision-making during the task. An overview of these responses is included in the supplementary materials and the response data is available at the github repository linked above.

[3] The data and code for this experiment and for all analyses reported here can be found on github at: https://github.com/erik-brockbank/rps-bot-manuscript-public.

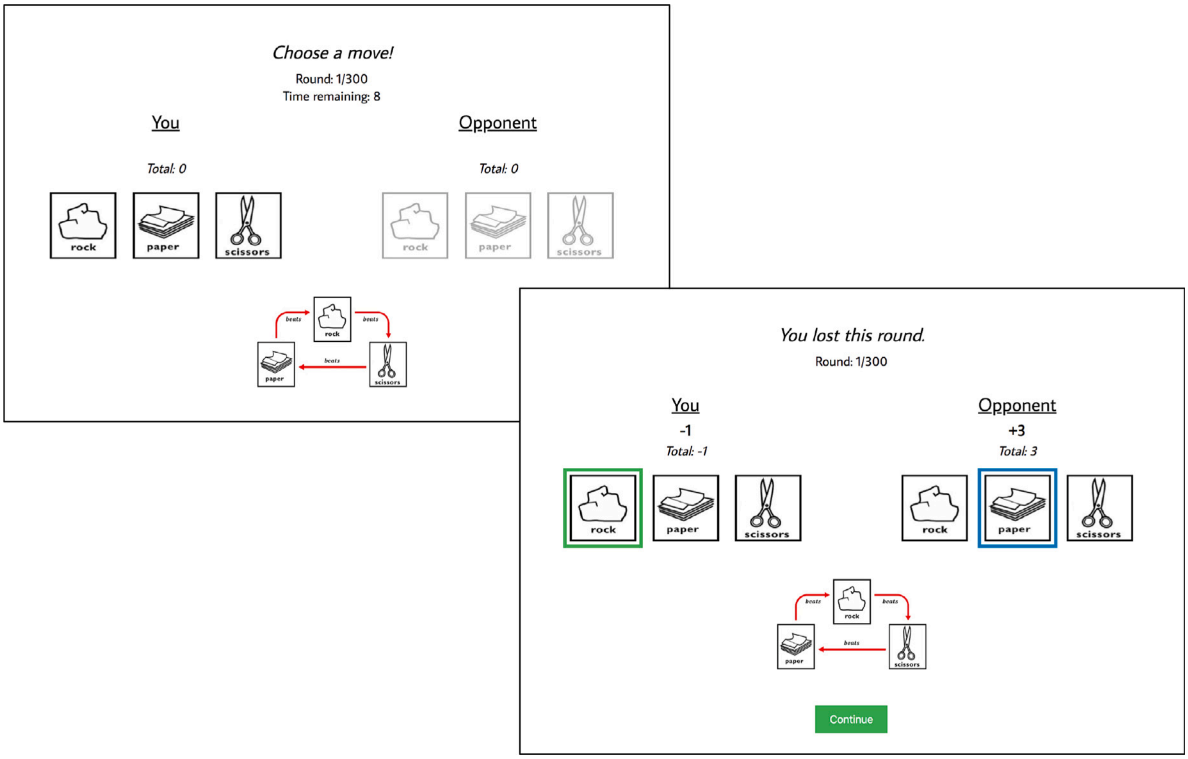


Fig. 1. The stages of each rock, paper, scissors round. Top left: participants had 10 s to select their move by clicking one of three "cards." Bottom right: participants were shown the results, along with updated points for each player, before proceeding to the next round.

2.3. Bot opponent behavior

Participants were paired with one of seven bot opponents that exhibited a distinct sequential pattern or "strategy" in its move choices. Concretely, bot strategies dictated a particular move each round based on events from the preceding round(s). Bots chose this move 90% of the time and each of the other (non-strategic) moves 5% of the time. While the 10% probability of an unpredictable move introduced a small amount of noise into the bot's behavior, each of the bots was highly predictable subject to a particular dependency or pattern. This predictability provided near-optimal conditions for participants to adapt to their bot opponent. In this way, we obtain a clear measure of the flexibility and limits of people's adversarial behavior in this setting: the categories of structured patterns people can exploit in an opponent, and the ones that are too complex to adapt to.

The complexity of the bot strategies can be described in terms of the previous events that dictated their move choice each round; the bot strategies form a hierarchy with an increasing number of events predicting their actions (see Fig. 2). In this experiment, we tested bots across three distinct levels in this hierarchy. For the simplest bots, their move choices were determined only by their previous move; they favored a particular *transition* from one move to the next and participants needed only recognize this transition pattern. An equivalent class of bots chose their move instead based on their (human) *opponent's* previous move. To increase the complexity of these *transition-based* bot opponents, we introduced a second set of bots whose preferred transitions were not static but varied depending on the *outcome of the previous round*. In this way, their moves were not merely dependent on their own or their opponent's prior move, but the result of that move (in essence, reflecting *both* players' previous moves). "Win-stay, lose-shift" behavior, in which a player is more likely to repeat moves that were successful and change moves when they produce unfavorable outcomes, is one variant of these strategies. Finally, a third class of bot opponents further increased strategy complexity by choosing a different move each round based on the combination of the previous outcome and their move two rounds before.

The bot strategies were chosen to encapsulate a broad range of patterns that human players might plausibly detect and exploit in an opponent. The patterns have been formally described in prior work (Brockbank & Vul, 2021a; Dyson, 2019) and have been shown to arise in repeated games of rock, paper, scissors among human dyads (Batzilis et al., 2019; Brockbank & Vul, 2020; Zhang et al., 2020). Further, prior work suggests that people are sensitive to some of these patterns in an algorithmic opponent's behavior, such as simple transition biases (Dyson et al., 2018; Guennouni & Speekenbrink, 2022; Stöttinger et al., 2014) and outcome-dependent transition strategies (Dyson et al., 2020, 2018; Zhang et al., 2021). However, previous studies have not systematically tested the full range of move patterns explored here. To illustrate the relationship between these patterns, we walk through this "strategy space" in more detail below (see Fig. 2). We first describe the categories that bot strategies belonged to, and then briefly outline the specific bots tested here, which represent a maximally diagnostic subset of the possible strategies.

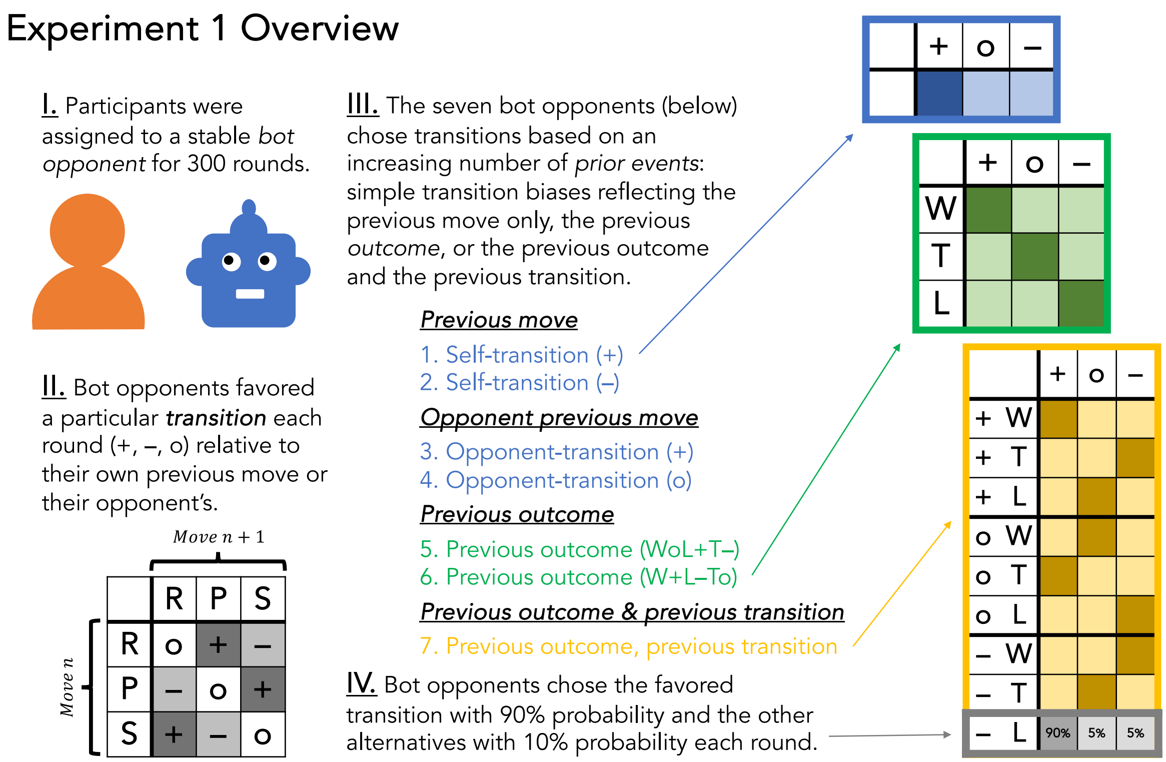


Fig. 2. Experiment 1 overview.

2.3.1. Previous move dependencies

The bot strategies in the current experiment amount to favoring a particular *transition* between moves on the basis of prior events. In the rock, paper, scissors game, move transitions can be expressed in three possible ways. A "positive" or "up" transition (+) involves selecting the move that would beat the previous move (e.g., playing "paper" following "rock"). A "negative" or "down" transition (−) involves instead choosing the move that would *lose* to the previous move, for example transitioning from "rock" to "scissors". Finally, a "stay" transition (0) is one in which the same move is repeated. The simplest bots in this experiment exhibited a stable bias towards particular transitions, for example selecting the move each round that represented a positive (+) transition from their previous move (Fig. 2). Critically, such transition biases can be exhibited relative to one's own previous move or an *opponent's* previous move. Thus, a bot that frequently repeats the same move can be described has having a "stay" (0) bias in its *self-transitions*, while a bot that most often repeats its opponent's previous move can be described as having a "stay" bias in its *opponent-transitions*.

2.3.2. Previous outcome dependencies

The previous move dependencies above can be expanded by including additional prior events that impact a player's transitions. Rather than always favoring a particular transition, more complex bots in the current experiment were biased towards different transitions depending on the previous game's *outcome*. For example, this might involve repeating the same move (a stay transition) when it was successful (after a win) but transitioning up and down after a tie or loss respectively. Critically, while a simple transition bias can be described as occupying one of three possible states (favoring up, down, or stay relative to a previous move as described above), *outcome-based* transition biases involve nine distinct states: three possible transition biases (up, down, or stay) for each previous outcome (win, tie, or loss), thus illustrating their increased complexity (see Fig. 2). Because an outcome is a function of the player's previous move and their opponent's, outcome-based transitions can be equivalently stated in terms of *self-transitions* and *opponent-transitions* (e.g., a positive self-transition after a win is equivalent to a negative opponent-transition after a win). We describe all outcome-based bot dependencies below in terms of their self-transition bias after each outcome.

2.3.3. Previous outcome, previous transition dependencies

In addition to the previous round's outcome, a player's transition from one round to the next might show a further dependency on their *previous transition* (i.e., the one they made two moves ago to the previous move). For example, if they transitioned up, down, or stay and then *won* in the previous round, this might lead them to make the same transition again. Here, the bias towards a particular self-transition in each round can reflect nine unique previous event combinations (a previous up transition and a win, a previous up transition and a tie, …), leading to 27 possible transition biases. Dependencies at this level are therefore significantly more complex than the simplest ones outlined above (Fig. 2).

Although further expansions of the move dependencies described above are possible, the patterns outlined here represent a principled approach to capturing a broad range of structured behavior that people might adapt to in an opponent's moves: choosing based on either player's previous move (transition dependencies), the combination of both players' previous moves (outcome-based dependencies), and a further inclusion of the opponent's choice *two* moves prior (previous outcome, previous transition). Below we describe the bot opponents in the present experiment, which spanned these behavioral complexity levels.

2.3.4. Bot strategies

The seven bot opponents that participants faced in the current experiment were chosen to encapsulate the range of behavioral dependencies described above. The bot strategies are illustrated in Fig. 2 and outlined below in increasing order of complexity. This set of strategies is not an exhaustive list of all the patterns that fall under the dependency categories above; instead, the bots below were chosen to be maximally *diagnostic* of the increasingly complex sequential patterns people might be able to exploit in an opponent's moves. After outlining each of the strategies tested in the current experiment, we briefly describe the criteria used to exclude additional candidate strategies.

Previous move:

- **Self-transition (+)** bot most often chose the move that constituted a positive (+) or upward transition relative to its own previous move, i.e., the move that would beat what it had previously played.
- **Self-transition (−)** bot favored the move each round that constituted a negative (−) or downward transition relative to its own previous move, i.e., the move that would lose to what it had previously played.
- **Opponent-transition (+)** bot favored the move each round that constituted a positive transition relative to its *opponent's* previous move. In other words, it typically chose the move that would beat what its opponent had just played.
- **Opponent-transition (0)** bot primarily chose the move each round that constituted a stay transition, once again relative to its *opponent's* previous move, essentially copying its human opponent from one round to the next.

Previous outcome:

- **Previous outcome (W0L+T−)** bot favored stay self-transitions after a win, positive self-transitions after a loss, and negative self-transitions after a tie.
- **Previous outcome (W+L−T0)** bot made positive self-transitions after a win, negative self-transitions after a loss, and stay self-transitions after a tie.

Previous outcome & previous transition:

- **Previous outcome, previous transition** bot favored distinct self-transitions based on each unique combination of its *previous* self-transition and the resultant outcome of the previous round.

The seven bot strategies in this experiment represent a strategic sampling of all possible *previous move* dependencies, *previous outcome* dependencies, and *previous outcome & previous transition* dependencies described above that can be enumerated. Though many more such biases are possible within each of these categories, we have chosen only those which satisfy several additional criteria. First, we removed from consideration any bots which participants might successfully exploit without giving any thought to the underlying pattern in their opponent's behavior. These are transition bots that simply play the same move over and over (a stay *self-transition*) or could be reliably beaten by playing the same move over and over (a downward *opponent-transition*).

For this same reason, we also excluded bots which exhibit a simple bias towards a particular move, e.g., choosing rock on 90% of trials (this is conceptually similar to favoring a stay self-transition). Prior work suggests that people can often exploit opponents that exhibit stable move biases (Danckert, Stöttinger, Quehl, & Anderson, 2012; Kangas et al., 2009; Lie et al., 2013; Sepahvand, Stöttinger, Danckert, & Anderson, 2014; Stöttinger et al., 2014). One critical finding from these results is that the strength of the adaptive response is fairly sensitive to the strength of the opponent's bias; participants are less likely to exploit an opponent that only *sometimes* favors a particular move. Consistent with these results, we expect that the bot strategies outlined here (or those excluded) could be more difficult to exploit if the bias to favor a move or transition were lower (e.g., 70% or even 50%). Though it remains unclear what kind of psychometric function describes this change in adaptation to increasingly subtle biases, here we focus instead on the sequential patterns people can adapt to when such patterns are highly reliable but increasingly complex.

In this vein, a final criterion for selecting among potential bot strategies was that the bot strategies above cannot be described at a simpler complexity level. For example, a *previous outcome* bot that transitions up after every previous outcome is no different from the simpler `Self-transition (+)` bot. Similarly, it is possible for some previous outcome strategies (even those with a different transition after each outcome) to be re-cast as simple opponent-transition dependencies (see Dyson, 2019 for further discussion). The bot strategies above were selected to avoid this possibility, in essence utilizing the full *expressiveness* of the dependency that dictated their moves (Brockbank & Vul, 2021a). The seven bot strategies in this experiment therefore include both a broad sampling of the patterns people might adapt to, as well as a selective or *diagnostic* set of patterns for understanding people's ability to adapt to their bot opponent.

2.3.5. Measuring adaptation to bot strategies

To understand how sensitive people were to stable patterns in their opponent's behavior, we measured participants' *win percentage* against each bot opponent. If participants fail to adapt to the sequential pattern exhibited by their bot opponent, we expect them

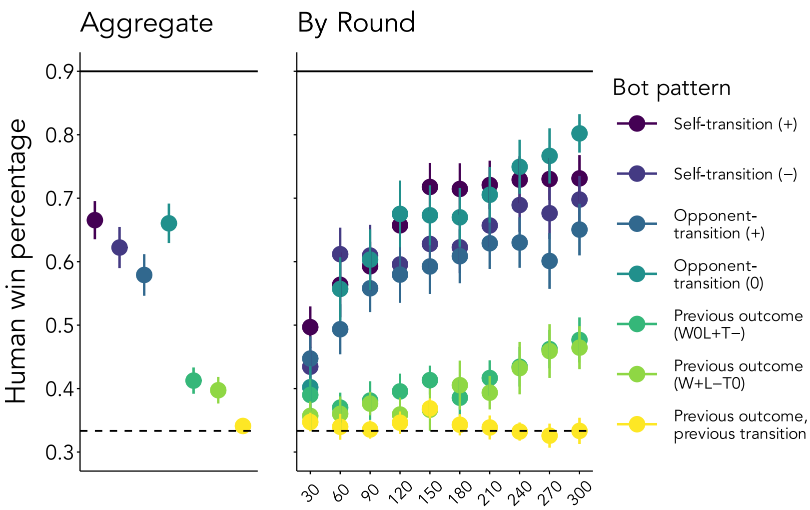


Fig. 3. Performance against RPS bot strategies. (Left) Overall participant win percentage against each bot. (Right) Participant win percentage against each bot over the course of the experiment. Error bars reflect standard error of the mean (SEM) across participants. Dashed lines indicate chance performance, while solid lines indicate optimal performance (bots chose strategy-consistent moves on 90% of rounds and each of the other two moves roughly 5% of the time, so participants could win at best 90% of rounds on average). Participants were highly successful at detecting and exploiting transition dependencies but showed little adaptation to more complex behavior patterns.

to perform at chance, winning 1/3 of rounds in a given block of trials. Win rates greater than 33% indicate successful exploitation by participants. We compare participants' win percentages across each of the seven bot conditions to better understand the level of behavioral complexity that people can adapt to over repeated interactions.

Prior work has sought to characterize the *representations* underlying people's adaptive behavior by estimating the depth of their recursive reasoning about their opponent in adversarial games (Camerer et al., 2004; Stahl & Wilson, 1995); results applying these models to rock, paper, scissors play have suggested that people seek to exploit patterns in their opponent's prior moves (*level-1* responding; see Batzilis et al., 2019; Guennouni & Speekenbrink, 2022) but that deeper levels of recursion are less frequent and may not be useful, since the few available actions in RPS quickly introduce cycles into recursive reasoning (De Weerd, Verbrugge, & Verheij, 2013). The bot strategies described above can all be considered *level-0* strategies since they are not adaptive to participants' own moves. The question of how participants perform against the bots is not one of recursive depth, but of adapting to increasingly sophisticated sequential patterns.

2.4. Results

2.4.1. Adaptation to bot strategies

How well did people detect and adapt to behavioral regularities of varying complexity? Participants were highly successful at exploiting the four simple *previous move* bots (`Self-transition (+)`, `Self-transition (-)`, `Opponent-transition (+)`, and `Opponent-transition (0)`; Fig. 3 Left). Average win percentages ranged from 57.9% (SE = 3.28%) to 66.5% (SE = 3.01%) against these opponents. Not surprisingly, participant win percentages were significantly higher than chance in these conditions (`Self-transition (+)`: $t(31) = 11.05$, $p < .001$; `Self-transition (-)`: $t(30) = 8.91$, $p < .001$; `Opponent-transition (+)`: $t(30) = 7.49$, $p < .001$; `Opponent-transition (0)`: $t(30) = 10.52$, $p < .001$). In contrast, participants showed only moderate success against the *previous outcome* strategies, `Previous outcome (W0L+T-)` and `Previous outcome (W+L-T0)`. Average win rates against these strategies were 41.3% (SE = 2.07%) and 39.7% (SE = 2.09%), respectively. Though participants performed above chance overall in these conditions as well (`Previous outcome (W0L+T-)`: $t(29) = 3.82$, $p < .001$; `Previous outcome (W+L-T0)`: $t(30) = 3.06$, $p = .005$), they were far less successful than those matched with the simpler *previous move* bots. Finally, participants performed poorly against the most complex `Previous outcome, previous transition` strategy; average win rate was 34.1% (SE = 0.7%) and did not differ significantly from chance ($t(30) = 1.05$, $p = .30$). In this way, the seven strategies tested here decompose adaptive reasoning into patterns that can be exploited *strongly*, *partially*, and *not at all*.

Participants' learning trajectories against each bot tell a similar story. For the *previous move* bots (`Self-transition (+)`, `Self-transition (-)`, `Opponent-transition (+)`, and `Opponent-transition (0)`), participants rapidly detected the opponent's strategy and exploited it for the majority of the experiment; learning rates and maximum performance are similar against all four of these transition dependencies (Fig. 3 Right). Meanwhile, people's ability to exploit the *previous outcome* bots (`Previous outcome (W0L+T-)` and `Previous outcome (W+L-T0)`) arose only in the last 100 rounds of the experiment, and never reached performance levels comparable to the transition strategies. Finally, consistent with aggregate performance, participants never succeeded above chance against the `Previous outcome, previous transition` bot.

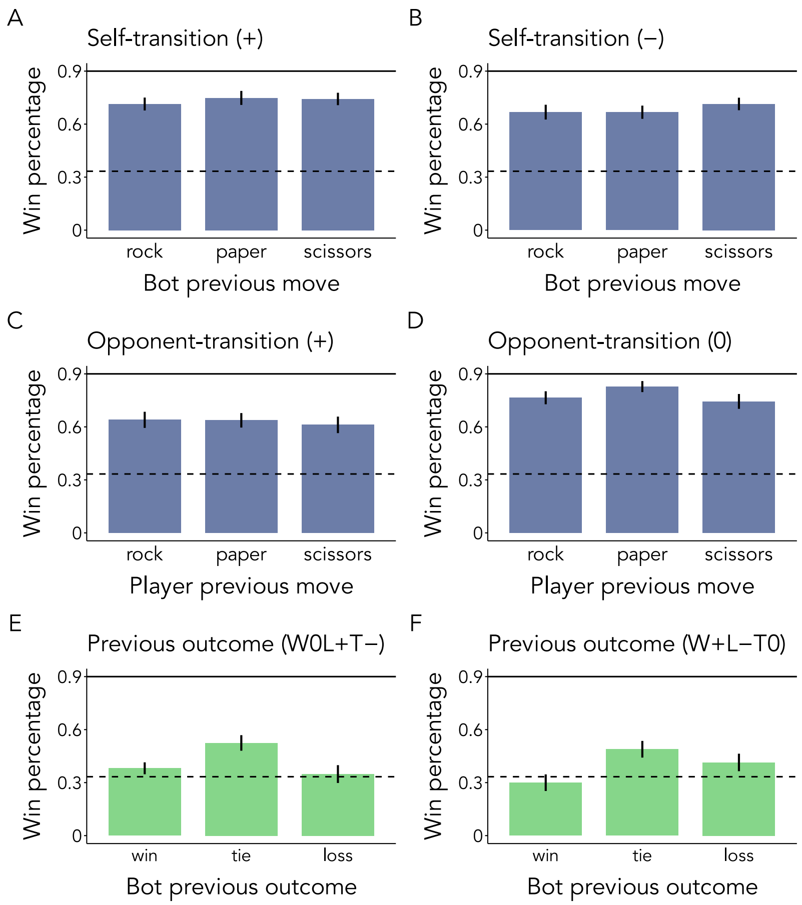


Fig. 4. Participant win percentage against bot opponents following each of the prior events that dictated the bot's move choice. (A)–(D) Conditional win rates against the transition strategies based on each previous bot or participant move. (E)–(F) Conditional win rates against the more complex strategies based on each previous bot outcome. Error bars reflect standard error of the mean across participants. Dashed lines indicate chance performance, while solid lines indicate optimal performance (bot opponents chose the strategy-consistent move following each move or outcome 90% of the time and chose each of the other two moves 5% of the time; the expectation for optimal human win rate is therefore 90%). People showed uniformly successful adaptation against the previous move dependencies; meanwhile, partial adaptation to the previous outcome dependencies was primarily based on exploiting individual contingencies in each bot's strategy (behavior after ties).

2.4.2. Decomposing adaptive behavior

What explains participants' failure to adapt to the *previous outcome* dependencies as effectively as the simpler *previous move* bots? Lower win rates against the more complex bot strategies could have arisen from selective learning of the dependency, wherein only part of the opponent's contingency structure was effectively used against them, or degraded overall learning such that participants exploited the full scope of the dependency, albeit noisily. These constitute distinct accounts of the learning process underlying adversarial decision making in this setting. Against the four simple *previous move* strategies (Fig. 4A–D), participants' win percentage in the last 100 rounds was significantly greater than chance following all bot or participant previous moves (all *ps* $< .001$). Further, win rates following each previous move were not significantly different from *each other* for three of the four *previous move* opponents (`Self-transition (+)`: $F(2, 31) = 0.73$, $p = .49$; `Self-transition (-)`: $F(2, 30) = 1.73$, $p = .19$; `Opponent-transition (+)`: $F(2, 30) = 0.30$, $p = .74$; `Opponent-transition (0)`: $F(2, 30) = 4.40$, $p = .02$). In other words, participants in the *previous move* conditions largely exploited the full contingency structure of their opponent's behavior.

However, against the two *previous outcome* strategies (Fig. 4E–F), participants only won reliably after a tie, indicating that they learned *individual components* of the bot's strategy. To illustrate, win percentages differed significantly following each outcome in these conditions (`Previous outcome (W0L+T-)`: $F(2, 29) = 3.77$, $p = .03$; `Previous outcome (W+L-T0)`: $F(2, 30) = 3.29$, $p = .04$). These non-uniform win percentages were driven by the fact that for both of these strategies, win percentages differed significantly from chance following a tie (`Previous outcome (W0L+T-)`: $t(29) = 4.34$, $p < .001$; `Previous outcome (W+L-T0)`: $t(30) = 3.30$, $p = .003$) but *not* following a win or loss. This selective adaptation suggests that for more complex opponent behaviors, participants failed to represent the full scope of the opponent's strategy, instead exploiting specific behaviors *within* the more complex strategies.

2.5. Discussion

In this experiment, we explored the basis for adaptive adversarial reasoning in a simple mixed strategy equilibrium game. Participants played 300 rounds of rock, paper, scissors against a bot opponent that exhibited one of seven distinct patterns in its move choices. These patterns increased in complexity from simple transition biases to intermediate patterns where the transition biases varied across previous outcomes (e.g., "win-stay, lose-shift"), to the most complex transition dependencies contingent on both the previous outcome and the bot's own previous transition. We examine how well participants exploited their bot opponents to better understand the scope of their adversarial reasoning.

Participants learned rapidly and were highly successful against the simplest *previous move* opponents, showed some success against intermediate *previous outcome* bots (primarily in the final third of the game), and performed at chance against the complex *previous outcome, previous transition* strategy. Further, a close examination of participants' conditional win rates following each outcome shows that people were not adapting to the *previous outcome* strategies uniformly. Instead, their partial success appears largely isolated to exploiting individual transitions in the bot's strategy. These results are consistent with prior work suggesting that people can successfully exploit bots that exhibit a stable self-transition or opponent-transition bias (Dyson et al., 2018; Guennouni & Speekenbrink, 2022; Stöttinger et al., 2014) but show incomplete exploitation of opponents with outcome-dependent transition biases (Dyson et al., 2020; Zhang et al., 2020, 2021). However, prior work has not offered an account of this pattern of results.

Intriguingly, these results may arise from similar adaptive behavior by participants. First, in all four *previous move* conditions, optimal exploitation of the bots can be achieved by implementing self-transition biases in one's own moves (for example, after a win against the `Self-transition (-)` bot, shifting down from one's previous move has a 90% chance of winning the next round). However, this strategy of assuming a stable self-transition bias seems to explain the results in the *previous outcome* conditions as well. To illustrate, one adaptation to the `Previous outcome (W0L+T-)` strategy is to favor repeating the same move over and over (a *stay* self-transition); doing so can induce a state of alternating wins and ties for the participant as the bot shifts *down* after a tie and then back *up* after a loss. Discovery of this simple strategy could explain participants' high win percentage following a *tie* against this opponent (Fig. 4E). Indeed, in an exploratory analysis of participant self-transitions in the last 100 rounds against this bot, the average proportion of stay self-transitions was 45.4% (SE = 5.97%). This number is offset by a small number of participants (N = 6) who favored an alternative strategy of repeating *positive* self-transitions, which can instead trap the bot in a pattern of continual losses (since it also transitions up after a loss). These participants chose positive self-transitions on 70.4% of the final 100 rounds (SE = 7.14%) and had an average win percent of 72.2% (SE = 4.78%).

Against the `Previous outcome (W+L-T0)` bot, we see a similar set of results. First, exhibiting a positive self-transition bias by cycling between moves could trap the opponent in a cycle of alternating ties and losses as they repeated the same move following a tie (leading to a loss) and then shifted down after a loss (leading to a tie). In the last 100 rounds, participants averaged 38.5% positive transitions (SE = 4.39%). Here too, this number was offset by a small group of participants (N = 7) who instead favored *negative* self-transitions, which could trap the bot in a sequence of losses. These participants averaged 75.2% negative transitions in the final 100 rounds (SE = 4.95%), achieving an average win percentage of 72.6% (SE = 2.87%). These results likely explain the higher win percentage after bot *ties* and *losses* shown in Fig. 4F.

Taken together, the current results suggest that people have a limited *hypothesis space* for representing sequential structure in an opponent's moves, but will flexibly exploit the simplest structures when they are present (and highly salient, i.e., appearing in roughly 90% of opponent moves). In particular, they may do so by implementing transition biases in their own moves when this behavior is adaptive. These findings raise the possibility that participants' adversarial reasoning relies heavily on self-monitoring of adaptive patterns or structure in their *own* actions. The current experiment provides only a partial measure of this ability, since participants were able to exploit the stable bots using similar self-transition biases in nearly all cases. *A complete account of adaptive adversarial reasoning should therefore consider people's ability to detect and modify a full range of sequential patterns in their own moves.* Experiment 2 addresses this second aspect of strategic decision-making.

3. Experiment 2: adaptive bot opponents

In this experiment, we test whether sophisticated adaptive reasoning is driven primarily by the ability to modulate predictable patterns in one's own moves. The more complex behaviors displayed by the bots in experiment 1 arise in people's move choices against human and bot opponents (Brockbank & Vul, 2020; Dyson et al., 2018; Dyson, Wilbiks, Sandhu, Papanicolaou, & Lintag, 2016; Wang, Xu, & Zhou, 2014), suggesting that people are vulnerable to exploitation of these patterns. The current experiment asks whether people are able to minimize complex patterns in their actions, even though they fail to exploit these same regularities in others.

We test this question by once again evaluating people's behavior against a bot opponent over 300 rounds of rock, paper, scissors. However, rather than choosing their moves according to a fixed dependency structure as in experiment 1, the bots in the current experiment *adapted to the human participants*. Concretely, each bot tracked a distinct sequential regularity in its human opponent's moves over the course of the game and tried to exploit this regularity. On every round, the adaptive bot estimated its human opponent's most likely move based on the particular pattern it was tracking, then chose its own move accordingly. Thus, for participants to succeed against their adaptive bot opponent required reducing the degree to which they exhibited the pattern the bot was exploiting. The patterns that the adaptive bots relied on to predict their opponents mirrored the structure of those exhibited by the stable bots in experiment 1. This allows us to precisely measure people's capacity to minimize increasingly complex patterns in their own decisions.

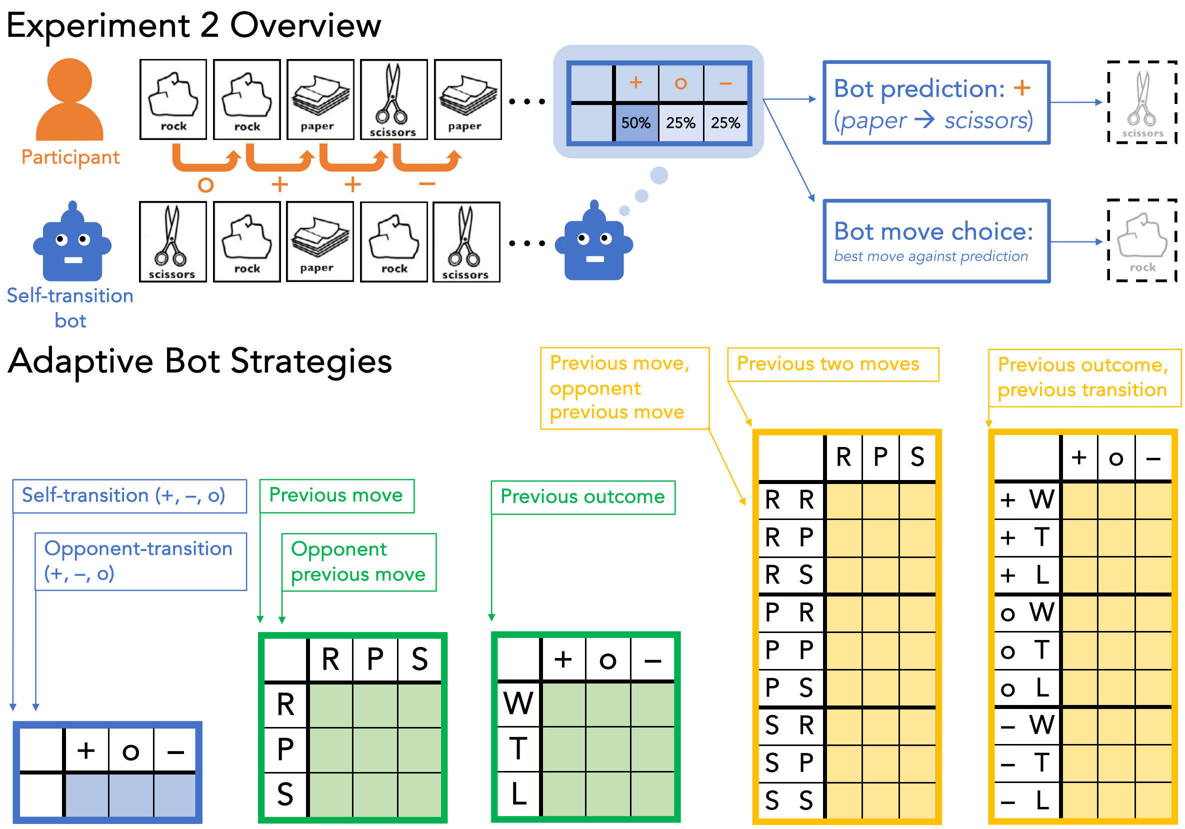


Fig. 5. Experiment 2 overview.

3.1. Participants

Participants were 194 undergraduate students who received course credit for their participation. One participant was removed because of technical error and a second was excluded due to clear evidence of not trying (i.e., choosing the same move to their own detriment in the vast majority of rounds). Participants were randomly assigned to one of eight adaptive bot conditions, described in further detail below. Our sample size was chosen to have a minimum of 20 participants in each bot condition. This gave us 90% power to detect an effect size of $d = 0.77$ in our estimate of bot win rates against each participant (or, symmetrically, participant win rates against each bot), an average win rate of roughly 55% assuming uniformly distributed win rates. Informed consent was obtained from all participants in accordance with the Institutional Review Board's approved protocol. Participants completed the experiment in a web browser online.

3.2. Task overview

The procedure for the experiment was identical to experiment 1 with the notable exception that participants were now paired with an *adaptive* bot opponent. This opponent chose its moves in an effort to maximize expected win probability in each round, using the participant's decisions on prior rounds to predict their next move (see Fig. 5). As before, participants were not told anything about their opponent's identity. All additional aspects of the experiment were identical to experiment 1, including the way points were allocated to wins (3), losses (−1), and ties (0). As in experiment 1, this point allocation does not impact participants' optimal strategy in the task. We compare human performance against each of the bot opponents to understand the complexity of patterns participants can modify in their *own* moves.

3.3. Bot opponent behavior

Participants were paired with one of eight adaptive bot opponents for the duration of the experiment. Each bot had an identical decision policy of choosing the move that would beat whatever move it estimated was most likely for its human opponent in the next round. In the event that multiple opponent moves were considered equally probable, the bot sampled one at random and chose the move that beat the sampled move.[4] What differentiated the bots across conditions was *how* they determined their human opponent's most likely move over consecutive rounds—each bot relied on a particular sequential dependency in its opponent's choices to predict their next move. To illustrate, a naïve approach might involve simply tracking a participant's cumulative proportion of *rock*, *paper*, and *scissors*, then selecting the move each round that would beat whichever choice had been most frequent. Given the simplicity of this dependency and the fact that people are unlikely to show a strong ongoing bias towards a particular move, we would not expect such a bot to perform particularly well against human opponents. However, tracking more complex patterns presents an opportunity for more successful prediction and exploitation.

The eight adaptive bots in the current experiment tested this idea by exploiting a broad range of dependencies in their human adversary's moves. These were chosen to align with the full scope of patterns exhibited by the stable bots in experiment 1. This allows for direct comparison between the complexity of patterns in opponent behavior that people successfully exploited and those they are able to minimize in their own moves. Further, these eight dependencies have been previously observed in games among human dyads (Brockbank & Vul, 2020), so all of them are good candidates for potentially exploiting participants over many rounds. Each of the eight adaptive bot strategies are described below in order of increasing complexity (see Fig. 5).

- **Self-transition** bot tracked participant *self-transition* rates and chose the move each round which beat its opponent's most likely self-transition (+, −, 0).
- **Opponent-transition** bot chose a move based on the participant's most likely *opponent-transition* (+, −, 0), i.e., the participant's most likely transition relative to the bot's previous move.
- **Previous move** bot tracked the co-occurrence of every pair of participant moves from one round to the next. For example, did the participant play *rock* most often after playing *paper*? This is similar to the `Self-transition` bot but allows for the possibility that a participant's most likely self-transition may vary depending on their previous move (i.e., if they choose *rock* most often after *rock*, but *paper* most often after *scissors*, these represent self-transition biases that differ across prior moves).
- **Opponent previous move** bot was identical to the `Previous move` bot above, except that it exploited any pattern in participant moves based on their *bot opponent's* previous move rather than their own.
- **Previous outcome** bot tracked a participant's most likely transition conditioned on each previous outcome. "Win-stay, lose-shift" behavior or any other *outcome-transition* dependency can be exploited by this bot.
- **Previous move, opponent previous move** bot tracked player move choices given each *combination* of their own previous move and their bot opponent's previous move. This bot relied on the same information as the `Previous outcome` bot but once again encoded the dependency more richly by tracking all unique combinations of the two players' prior moves and the participant's subsequent move.
- **Previous two moves** bot chose the move which beat its human opponent's most likely move given the participant's move choices in each of the previous two rounds.
- **Previous outcome, previous transition** bot exploited any dependency participants exhibited in their transitions each round, given both the outcome of the previous round and the transition they made in the previous round. For example, if participants were more likely to shift *up* after a round in which they shifted *up* and *won*, this bot would exploit such a pattern. These patterns align with the *previous outcome, previous transition* bot in experiment 1.

The adaptive bot strategies above capture a broad range of prior events that might impact a participant's move choices in predictable ways: their own previous move (`Self-transition` and `Previous move`), their opponent's previous move (`Opponent-transition` and `Opponent previous move`), the two combined (`Previous outcome` and `Previous move, opponent previous move`), the participant's previous *two* moves (`Previous two moves`), or the participant's previous two moves alongside the opponent's previous move (`Previous outcome, previous transition`). By comparing participant behavior against each bot, we obtain a precise measure of people's ability to strategically modify their own actions to avoid being exploited on any of these dimensions. Prior work investigating people's adversarial reasoning against *adaptive* algorithmic opponents has contrasted adaptive behavior with simpler strategies such as random or minimax action selection (Moisan & Gonzalez, 2017), or has limited the space of adaptive opponent strategies to dependencies based on either one or two previous moves (West & Lebiere, 2001). The current work therefore offers a comprehensive evaluation of how people perform against adaptive opponents at varying levels of strategic sophistication.

3.4. Adaptive bot complexity

Just as in experiment 1, the adaptive bots described above differ in the *complexity* of their strategies. We can quantify the complexity of a bot's strategy based on the *memory demands* of tracking the dependency it uses to predict its opponent (see Fig. 5), much as we described the complexity of the bots in experiment 1 in terms of the number of states needed to describe their strategy (Fig. 2). The simplest adaptive transition bots (`Self-transition` and `Opponent-transition`) need only store and update three counts in memory: a 1×3 matrix with the number of +, −, and 0 transitions they observe. Meanwhile, the next three strategies above (`Previous move`, `Opponent previous move`, and `Previous outcome`) have an *intermediate* complexity because they

[4] Bots were indifferent between moves that had a 50/50 chance of a win or tie, and those that had a 50/50 chance of a win or loss. In this way, the bots maximized expected *win count*, but did not maximize expected *win count differential* (e.g., by favoring ties over losses).

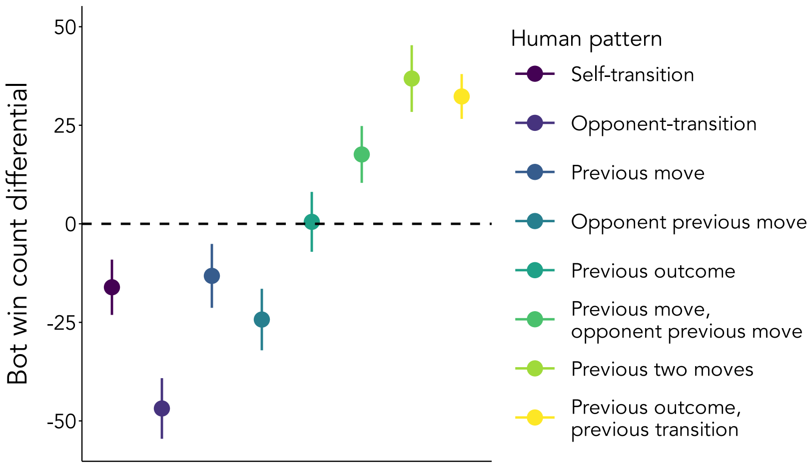


Fig. 6. Success of adaptive bot strategies against human opponents, shown in order of bot complexity. Dashed line indicates chance performance; error bars show standard error of the mean. Bot win count differentials greater than chance reveal strategies that reliably outsmart human opponents, while values less than chance indicate successful counter-exploitation by human players. Participants were unable to eliminate complex regularities in their own behavior (far right), but effectively counter-exploited the simpler bot strategies (far left).

instead maintain *nine* counts in memory using a 3×3 matrix of transition or move frequencies based on additional information from the previous round (i.e., each possible previous move or previous outcome). Finally, the most complex bots (`Previous move, opponent previous move`, `Previous two moves`, and `Previous outcome, previous transition`) rely on a 9×3 matrix which tracks opponent moves or transitions given nine possible previous event combinations (two previous moves or a previous transition and previous outcome) to estimate their opponent's most likely move.

The memory required to generate opponent predictions offers a straightforward description of how the strategies vary, which could account for differences in people's success against the bots. Because it is similar to the format for illustrating stable bot strategies in experiment 1, we can also compare results across experiments based on bot complexity. We explore how people's ability to adapt to each bot changes with the complexity of the dependency the bot exploits.

3.5. Results

As in experiment 1, results reported below along with the accompanying data can be found on github at: https://github.com/erik-brockbank/rps-bot-manuscript-public. The results of the post-experiment questionnaire, which are not analyzed here, can be found in the supplemental materials.

3.5.1. Performance against adaptive bots

How successful was each of the adaptive bot strategies against their human opponents? In experiment 1, we used participant win percentages to illustrate how well participants exploited the stable patterns in their bot opponent's moves (Fig. 3). Here, we instead examine bot *win count differentials* across conditions. A bot's win count differential is the number of times it beat its human opponent minus the number of times the human opponent won. Unlike one-sided win rates, the win count differential considers bot wins minus losses, thus allowing us to distinguish between ties and losses; this is important for the current experiment, as bot losses represent *counter-exploitation* by their human opponents. Win count differentials greater than zero for a given bot suggest that participants were reliably exploitable in that condition; meanwhile, values less than zero result from participants successfully counter-exploiting their bot opponent. The bot win count differentials illustrate a clear relationship between each bot's strategy complexity and how effectively that bot exploited participants (Fig. 6).

First, the three bot strategies which tracked the most complex dependencies in human move choices (27-cell memory) were able to consistently beat participants (three rightmost points in Fig. 6). Bot win count differentials were significantly greater than zero for all of these strategies (`Previous move, opponent previous move`: $t(24) = 2.44$, $p = .02$; `Previous two moves`: $t(19) = 4.36$, $p < .001$; `Previous outcome, previous transition`: $t(25) = 5.68$, $p < .0001$). For the two bot strategies with the highest average win count differentials, only 4 out of 20 (`Previous two moves`) and 4 out of 26 (`Previous outcome, previous transition`) participants had win count differentials greater than or equal to zero; this is significantly fewer than would be expected by chance (binomial test $p = .01$ and $p < .001$, respectively) and suggests that most people paired with these bots would not be expected to come out ahead over many rounds. In short, participants exhibited the most complex regularities enough for the adaptive bots to exploit them, and further, people were essentially trapped in these behavior patterns, indicating a clear limit to the structure they can minimize in their own moves.

In the three *intermediate* (9-cell memory) bot conditions, participants were far less exploitable. Two of the bots obtained win count differentials which were not significantly different from zero (`Previous move`: $t(23) = -1.63$, $p = .12$; `Previous outcome`: $t(21)$

= 0.07, p = .95) while bot win count differentials for the third were significantly *less* than zero (`Opponent previous move`: $t(27)$ = −3.11, p < .01). Intriguingly, given the evidence from prior work that people often exhibit "win-stay, lose-shift" behavior over repeated rock, paper, scissors rounds (Baek, Kim, Kim, Choi, Lee, Lee, Hahn, & Jeong, 2013; Batzilis et al., 2019; Brockbank & Vul, 2020; Dyson et al., 2018, 2016; Wang et al., 2014), the current results suggest that this pattern of outcome-dependent responding may be tempered when it is being actively exploited by an opponent. Participants' apparent success at evading exploitation in this way is also interesting in light of their minimal ability to adapt to these same patterns in an opponent's moves in experiment 1. More generally, these results present a clear contrast with those of the 27-cell strategies above; participants managed to avoid any kind of systematic exploitation based on the previous outcome or their own prior move, leading to chance performance over the 300 rounds in these conditions.

Finally, bot win count differentials in the two simplest (3-cell memory) conditions (along with the 9-cell `Opponent previous move` and, to a lesser degree, `Previous move` conditions) suggest highly successful adaptive behavior by *participants*. Bot win count differentials were significantly *less* than zero for the transition bots (`Self-transition`: $t(20)$ = −2.30, p = .03; `Opponent-transition`: $t(25)$ = −6.09, p < .01; Fig. 6, two leftmost points). At an individual level, bot win count differentials were negative against 17 out of 21 (`Self-transition`) and 22 out of 26 (`Opponent-transition`) participants in these conditions, significantly more than would be expected by chance (binomial test, p < .01 and p < .001, respectively). These results suggest that participants reliably *outwitted* bot strategies that exploited simple transition regularities. The only way to achieve this pattern of results is for participants to have discovered a way of counter-exploiting the bot opponents. *How did they accomplish this?* In the next section we present exploratory analyses aimed at better understanding this question.

3.5.2. Strategic responding to adaptive bots

In experiment 1, successfully exploiting the stable *previous move* bots required that participants incorporate self-transition biases into their own moves. In addition to participants' success against these simple bots, we find evidence that their *partial* exploitation of the *previous outcome* strategies resulted from incorporating higher rates of particular self-transitions into their moves. Might a similar strategy have allowed participants to counter-exploit the simplest adaptive bots in the current experiment? Here, we explore whether participants exhibited higher levels of self-transition and opponent-transition biases in conditions where they successfully outwitted their bot opponent. We test the hypothesis that participants may have systematically beaten the `Opponent-transition` and `Opponent previous move` bots through adaptive *self-transition* biases, and that their (more modest) success against the `Self-transition` and `Previous move` bots was a result of adaptive *opponent-transition* biases.

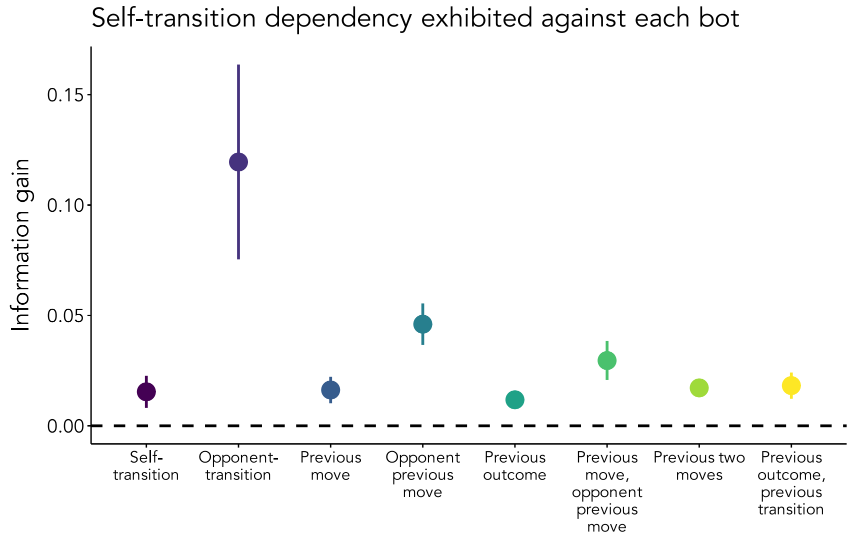
Quantifying participants' use of transition biases. The extent to which participants demonstrated a particular dependency over 300 rounds can be precisely quantified in terms of the *information gain* for that dependency in their moves (Brockbank & Vul, 2021a). Briefly, information gain measures how much the distribution of moves for a given dependency deviates from uniform. We use this measure to compare how much participants expressed self-transition and opponent-transition biases against different adaptive bot opponents. To illustrate, given a sequence of participant move choices X, the probability distribution over self-transitions $T = \{+, -, 0\}$ will be the number of times each self-transition appeared in X divided by $|X|$. Intuitively, the more uneven this distribution is, the more participants exhibited a reliable self-transition bias in X. The *Shannon entropy* H (Shannon, 1948) of the distribution formalizes this intuition: The *lower* the H value (less entropy) for a particular dependency, the *more* evident that dependency is in the underlying move sequence. The information gain $I(T)$ is simply the distance between the Shannon entropy of a uniform distribution over transitions and the entropy of the empirical distribution, $H(T)$:

$$I(T) = -\log(1/3) - H(T) = \log(3) + \sum_{i \in T} p(T_i)\ \log(p(T_i))$$

This value will be larger the more non-uniform the distribution of transitions T is; in short, I quantifies how much a participant's moves were predictable via a given dependency, with larger values indicating that people demonstrated that regularity more.

How much did participants display stable self-transition and opponent-transition biases against each of the adaptive bots? Overall, participants showed very little self-transition regularity, particularly against bot opponents that would have exploited this pattern (`Self-transition` and `Previous move` bots; Fig. 7 Top). However, participants exhibited the highest self-transition information gain in their moves against bots that tracked *opponent-transition* dependencies: `Opponent-transition` and `Opponent previous move` (and, to some degree, `Previous move, opponent previous move`). There is a significant difference in information gain for participants' self-transition dependency between the `Self-transition` and `Opponent-transition` bots ($t(45)$ = −2.10, p = .04), which are otherwise matched in the complexity of their strategies, as well as between the `Previous move` and `Opponent previous move` bots ($t(50)$ = −2.57, p = .01) which are similarly matched. Thus, participants paired with *opponent-transition*-exploiting bots appeared to rely on increased *self-transition* biases as a means to counter-exploit them.

Do participants exhibit a similar effect for opponent-transition dependencies, namely a greater opponent-transition bias against the bots that tried to exploit *self-transition* patterns? Participants displayed little opponent-transition bias against the bots that exploited this pattern (`Opponent-transition` and `Opponent previous move`; Fig. 7 Bottom). In contrast, they exhibited greater opponent-transition dependencies against bots that tried to exploit *self-transition* biases (`Self-transition`, `Previous move`, and `Previous two moves`). Information gain for the opponent-transition dependency differed significantly between the symmetrical `Self-transition` and `Opponent-transition` bots ($t(45)$ = 4.63, p < .001), as well as between the `Previous move` and `Opponent previous move` bots ($t(50)$ = −3.77, p < .001). In sum, participants showed increased *opponent-transition* dependencies against bots that sought to exploit their *self-transition* biases, mirroring the increased self-transitions described above. Despite the similar overall pattern, the magnitude of participants' elevated opponent-transition biases is reduced relative to the



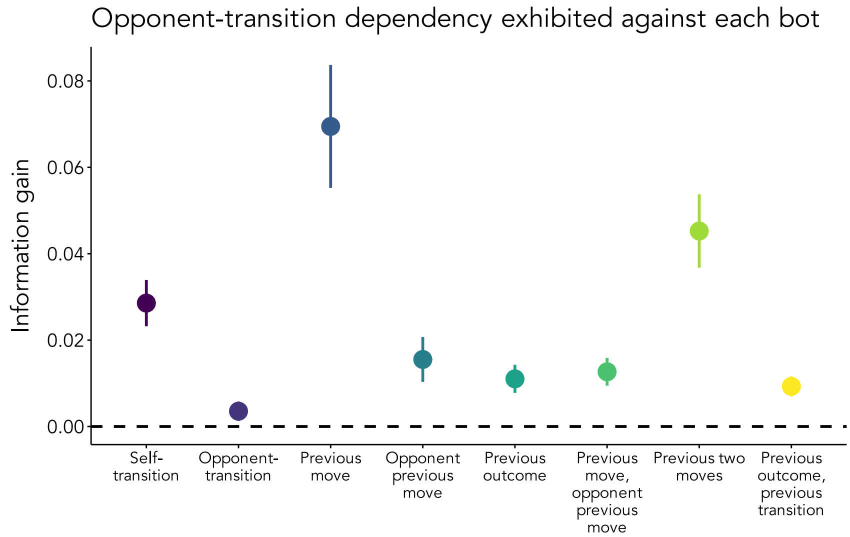


Fig. 7. (Top) Information gain for self-transition dependencies in participant move choices against each bot. (Bottom) Information gain for opponent-transition dependencies in participant move choices against each bot. Dashed lines show chance performance and error bars reflect standard error of the mean. Participants showed evidence of modulating their self-transition and opponent-transition dependencies to reduce exploitability and counter-exploit adaptive bot opponents.

self-transition biases. One reason for this may be that opponent-transitions were less adaptive against the bots; participants did not counter-exploit the self-transition tracking bots as effectively as participants paired with the opponent-transition tracking bots (Fig. 6).

Simulating counter-exploitation. To better quantify the degree to which particular transition biases were adaptive against the bots, we simulated stable self- and opponent-transition biases against the transition-exploiting bots. We simulated 1000 participants each playing 300 rounds against the `Self-transition` and `Opponent-transition` bots. In our simulations, participants chose exclusively a single self-transition or opponent-transition for all 300 rounds. We ran separate sets of 1000 simulations for each of the two transition-exploiting bots and each possible transition bias (two bots x six transition biases for a total of 12 sets of simulations). For each set of simulations, we recorded the average win count differential for participants exhibiting the particular bias against their bot opponent.

Our results show a clear *theoretical* benefit for favoring self-transitions against the `Opponent-transition` bot. For roughly 36% of simulated participants, a stable self-transition bias yielded a win count differential of 0.2 per trial against the `Opponent-transition` bot. Over the 300 rounds, this corresponds to a total win count differential between 40 and 80 (this result was similar for all three self-transition biases). Thus, while favoring a particular self-transition was not *always* adaptive against the `Opponent-transition` bot, doing so could sometimes maneuver the bot into move sequences that led to a substantially greater number of wins and ties than losses for participants, providing a plausible mechanism by which participants obtained positive win count differentials against this opponent (Fig. 7 Top). However, in simulated rounds against the `Self-transition` bot, we do not find the same advantage for exhibiting a stable bias towards *opponent-transitions*. For a similar proportion of simulated subjects, win count differentials were close to 0 per trial, but were never positive. Thus, while favoring a particular opponent-transition against the `Self-transition` bot might have allowed some participants to *avoid exploitation*, it is not clear that such a strategy could have enabled them to successfully *counter-exploit* the bot as they did (Fig. 6).

3.6. Discussion

In the current experiment, we tested people's ability to minimize sequential patterns in their move choices when being exploited by an adaptive opponent. Participants played 300 rounds of rock, paper, scissors against a bot that tried to predict the participant's next move. We tested eight bot opponents that varied in the complexity of the patterns they relied on to predict their human adversary; these encapsulate the range of dependencies exhibited by the bots in experiment 1, allowing for a symmetrical comparison between the patterns people can adapt to in an opponent and those they can revise in their own moves.

Against the most sophisticated adaptive bots, participants lost reliably, suggesting that they exhibited sequential patterns in their moves which they could not reduce. However, paired with bots that merely tried to exploit simple transition dependencies in their opponent's actions, participants showed evidence of successfully *outwitting* them. Our second key finding concerns the nature of this counter-exploitation. We find that participants exhibited increased *self-transition* dependencies when paired with bots that exploited *opponent-transition* biases, and vice versa with bots that exploited self-transition biases. Participants' success against the simpler adaptive bots reflects their ability to modify self-transition and opponent-transition patterns in their own moves. This is consistent with prior work finding that people can succeed against an adaptive opponent whose strategy is based on patterns at a *lag* of one prior move but not two (West & Lebiere, 2001); however, earlier results have not offered an account of the *mechanism* by which people outwit such adaptive opponents.

Broadly, these results suggest that people are flexible in their use of *self-transition* and *opponent-transition* patterns against adaptive opponents. And their use of these patterns is unlikely to be random or thoughtless; our simulation results reveal that doing so is not always advantageous. However, this adaptive ability appears limited to a choice among these relatively simple transition-level behavioral dependencies. Changes to one's own behavior in this adversarial setting enabled people to effectively counter-exploit simple opponents, but lacked the scope needed to adapt to more complex opponents.

4. General discussion

In this work, we address the question of how people perform adaptive reasoning in an adversarial setting. Specifically, we ask what kind of *sequential patterns* people can exploit in an opponent's behavior and avoid in their own actions against an adaptive opponent. Across two experiments using the game of rock, paper, scissors, participants demonstrated a highly selective ability to exploit patterns in their opponent's moves and revise identical dependencies in their own moves. These adaptive responses exhibited a consistent reliance on detecting and modifying *transitions* from one move to the next, but little ability to generalize to more complex patterns. In this way, results paint a clear picture of people's sequential, adaptive reasoning as flexibly utilizing a well-defined but surprisingly limited set of behaviors.

In experiment 1, we measured participants' ability to exploit a range of stable patterns exhibited by a bot opponent. Participants were highly successful against *transition-level* patterns over 300 rounds but struggled to adapt to more complex opponent behaviors. In experiment 2, we examined whether this limited adaptive behavior extended to revising patterns in one's *own* actions. Results from experiment 2 reinforce and complement the findings from experiment 1. Successfully exploiting the transition bots in experiment 1 required that participants implement self-transition biases. We also find in experiment 1 that the limited success against the *outcome-based* bots resulted from adaptive use of particular self-transitions. Results from experiment 2 show that people also manipulate transition biases in their moves as a means of counter-exploiting an adaptive opponent. Collectively, findings suggest that in repeated rock, paper, scissors interactions, people successfully outwit their opponent when doing so involves detecting or modifying patterns in their transitions from one move to the next. However, this flexibility is restricted to a constrained set of such patterns.

Critically, these findings likely represent an *upper bound* on people's ability to adapt to patterns in their opponent's and their own actions. The bots tested in both experiments were highly predictable, choosing pattern-consistent moves in roughly 90% of rounds. This is a lower level of noise in their patterned behavior than most prior work using algorithmic opponents (Dyson et al., 2020; Stöttinger et al., 2014; Zhang et al., 2021). It is likely that increasing the noise in the bots' sequential patterns would only further degrade people's ability to adapt; future work should consider the functional form of this signal detection problem. Given that the current results may be considered close to the upper bound of such a function, they are striking in the limitations they reveal about people's ability to detect sequential structure in an opponent's or their own actions. *What might these limitations reveal about behavior in other games and in adversarial settings more generally?*

Prior work using the rock, paper, scissors game suggests that adaptations to an opponent's strategy may generalize to new opponents (Stöttinger et al., 2014) and to new games with the same opponent (Guennouni & Speekenbrink, 2022). This latter finding has been shown in other economic games as well (van Baar et al., 2022). However, the current results suggest there are clear limits to what can be learned about an RPS opponent in the first place; to what degree does this finding hold in other adversarial contexts? People's ability to adapt to algorithmic opponents has been studied in a range of adversarial games outside RPS (Devaine, Hollard, & Daunizeau, 2014; Goodie, Doshi, & Young, 2012; Hedden & Zhang, 2002; Moisan & Gonzalez, 2017; Sher, Koenig, & Rustichini, 2014; Spiliopoulos, 2013; Zhang, Moisan, Aggarwal, & Gonzalez, 2022). For example, in signaling games, decisions about whether to lie or trust a bot partner reflect inferences about the underlying credulity and truthfulness of the partner (Zhang et al., 2022). Meanwhile, over repeated rounds of an attacker-defender game against adaptive bot opponents, people are vulnerable to exploitation of their most common action from previous rounds (Moisan & Gonzalez, 2017). Broadly, these findings suggest that in other economic games, people infer predictive features of an opponent's behavior but are vulnerable to exploitation based on their own action history. In some of these contexts, bots exhibit or exploit *biases* in action selection; these findings are similar to results observed in RPS play against biased opponents (Kangas et al., 2009; Lie et al., 2013; Stöttinger et al., 2014). Fewer of these results have explored adaptation to *sequential* patterns in opponent or participant moves (Hedden & Zhang, 2002; Spiliopoulos, 2013), but those that have obtain similar results to ours, suggesting that the sequential patterns people can exploit or revise in their own behavior is limited in these settings as well. More generally, it seems unlikely that rock, paper, scissors is entirely unique among adversarial games in the limitations it surfaces in people's ability to reason about sequential behavior.

In this vein, the generalizability of the present findings likely depends on the *source* of participants' limited adaptive reasoning. Here, we consider several possible explanations for participants' behavior in the present context and what this means for generalizing our findings to other contexts. First, it may be that people's adaptive reasoning about their opponent in RPS reflects limited cognitive resources. For example, limits in working memory might prevent them from recalling past moves needed to detect sequential patterns in their opponent's play or their own (Rapoport & Budescu, 1997). While working memory limits might in part account for participants' failure to exploit or evade the most complex bot opponents, we find this an unlikely explanation for their failure to exploit the intermediate complexity *outcome-transition* bots, since these required only remembering the result of the previous round. Further, if this were the primary bottleneck for adaptive behavior, we might expect at least partial exploitation of the most complex bot in experiment 1 driven by what limited information participants could hold in memory. However, it remains possible that other cognitive factors such as *interference* from the small set of available actions might have contributed to the challenges of adapting to sequential patterns. Insofar as limits in memory or processing ability account for our results, it suggests that similar findings might arise in any game or adversarial context that involves repeated interactions over a limited action space. This further suggests that successful adversarial reasoning may rely heavily on *scaffolding* that supports exploitation of sequential structure in an opponent's behavior by reducing cognitive demands.

However, another reason to be skeptical that the present results arise from resource limitations is that people reason effectively about complex sequential patterns in other adversarial settings (consider for example the array of chess openings an experienced player can recognize). To this end, it may be that people do not have access to the *abstractions* needed to reason about sequential behavior in RPS; in other words, maybe they simply did not know what patterns to look for in their opponent's actions or their own. Indeed, it seems likely that participants would have fared better in our experiments if they had received information ahead of time about the contingencies to watch for in their own moves or their opponent's (Bransford & Johnson, 1972). A natural question then is *why* people failed to detect the relevant action patterns. It may be that detecting complex sequential behavior patterns is easier in more grounded or naturalistic decision contexts. For example, people's ability to perform certain logical reasoning operations is limited when the operands are *arbitrary* (much like RPS moves), but they may succeed at isomorphic problems when they have a naturalistic schema for such reasoning (Cheng & Holyoak, 1985; Griggs & Cox, 1982). Recent work in this vein suggests that behavior in game settings that more closely emulate real-world behavior provides insight into the computational processes underlying coordination and planning (McCarthy, Hawkins, Wang, Holdaway, & Fan, 2021; van Opheusden et al., 2023; Wu et al., 2020), tool use and physical reasoning (Allen, Smith, & Tenenbaum, 2020), and a host of other cognitive abilities (Allen et al., 2023). If the limitations observed in the present results arise in part from the small and arbitrary action space employed in rock, paper, scissors, such limitations may be less likely to arise in game settings with more naturalistic action spaces. If this is true, it would suggest that the ability to successfully reason about sequential structure in others' behavior is *context-dependent*; to the degree that adversarial reasoning draws on domain-general capacities shared by RPS play and behavior in more realistic adversarial settings, such capacities may be limited.

However, it is also possible that people's failures to exploit sequential patterns of behavior in an opponent's actions or revise them in their own is a more general feature of adversarial reasoning that arises even in game settings with richer action spaces. How then, might we account for these more general limitations? A key feature of the rock, paper, scissors game is that it is fairly restrictive with regard to canonical theory of mind reasoning. Inferring an opponents' beliefs requires recursive reasoning which quickly becomes circular due to the few moves available (De Weerd et al., 2013). Instead, a player's primary recourse is to search for exploitable patterns in their opponent's moves across repeated rounds (Brockbank & Vul, 2021a). One account of the current results is that people's strategic sophistication is limited in the absence of the robust theory of mind that they are accustomed to in everyday contexts. In other words, without the opportunity to infer different possible goals or beliefs an opponent might have from their actions, we learn little from the sequential structure in their behavior alone. If our results arise from a general challenge of reasoning about others in settings that do not easily admit theory of mind inferences, then we might expect the limitations observed in the current results to arise in a wide range of interactions beyond adversarial games. For example, when someone's actions are driven by habit (Gershman et al., 2016), scripts (Schank & Abelson, 1977), or common behaviors (Thornton & Tamir, 2017), people may show a limited ability to recognize contingencies or complex sequential structure underlying these behaviors. More generally, reasoning about others might involve considering multiple causal models that can account for observed behavior (Burger & Jara-Ettinger, 2020); the current results shed light on the limitations of sequential action patterns as a basis for such reasoning.

4.0.1. Conclusion

In sum, this work explores how people adapt in repeated adversarial interactions. What kind of structured behavioral patterns can people predict and exploit in others? And how well can they modify these patterns in their own actions to avoid similar exploitation? By observing decision-making across repeated rounds of mixed strategy equilibrium game play, we obtain a clear and precise account of the limits of people's adaptive adversarial behavior.

CRediT authorship contribution statement

Erik Brockbank: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Edward Vul:** Conceptualization, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Data availability

I have included links to all data and code in the manuscript.

Acknowledgment

This research was partly supported by UCSD Research Award #A139946 awarded to Edward Vul.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cogpsych.2024.101654.

References

Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., & Schulz, E. (2023). Playing repeated games with large language models. arXiv preprint arXiv:2305.16867.
Allen, K. R., Brändle, F., Botvinick, M., Fan, J., Gershman, S. J., Griffiths, T. L., et al. (2023). Using games to understand the mind. PsyArXiv.
Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2020). Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning. *Proceedings of the National Academy of Sciences*, *117*(47), 29302–29310.
Aumann, R. J. (1985). Repeated games. In *Issues in contemporary microeconomics and welfare* (pp. 209–242). Springer.
Axelrod, R. (1984). *The evolution of cooperation*. New York, NY: Basic Books.
van Baar, J. M., Nassar, M. R., Deng, W., & FeldmanHall, O. (2022). Latent motives guide structure learning during adaptive social choice. *Nature Human Behaviour*, *6*(3), 404–414.
Baek, K., Kim, Y. T., Kim, M., Choi, Y., Lee, M., Lee, K., et al. (2013). Response randomization of one-and two-person rock–paper–scissors games in individuals with Schizophrenia. *Psychiatry Research*, *207*(3), 158–163.
Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 0064.
Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.
Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, *12*(4), 428–454.
Batzilis, D., Jaffe, S., Levitt, S., List, J. A., & Picel, J. (2019). Behavior in strategic settings: Evidence from a million rock-paper-scissors games. *Games*, *10*(2), 18.
Binmore, K. (1987). Modeling rational players: Part I. *Economics & Philosophy*, *3*(2), 179–214.
Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717–726.
Brockbank, E., & Vul, E. (2020). Recursive adversarial reasoning in the rock, paper, scissors game. In *Proceedings of the annual meeting of the cognitive science society*: *42*.
Brockbank, E., & Vul, E. (2021a). Formalizing opponent modeling with the rock, paper, scissors game. *Games*, *12*(3), 70.
Brockbank, E., & Vul, E. (2021b). Humans fail to outwit adaptive rock, paper, scissors opponents. In *Proceedings of the annual meeting of the cognitive science society*: *43*.
Budescu, D. V., & Rapoport, A. (1994). Subjective randomization in one- and two-person games. *Journal of Behavioral Decision Making*, *7*(4), 261–278.
Burger, L., & Jara-Ettinger, J. (2020). Mental inference: Mind perception as Bayesian model selection. In *Proceedings of the annual meeting of the cognitive science society*: *42*.
Camerer, C. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.
Camerer, C., Ho, T., & Chong, K. (2003). Models of thinking, learning, and teaching in games. *American Economic Review*, *93*(2), 192–195.
Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *Quarterly Journal of Economics*, *119*(3), 861–898.
Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*(4), 391–416.
Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. *Trends in Cognitive Sciences*, *2*(8), 275–281.
Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, *69*(5), 1193–1235.
Courakis, A. S. (1981). *Inflation, depression, and economic policy in the West*. Rowman & Littlefield.
Danckert, J., Stöttinger, E., Quehl, N., & Anderson, B. (2012). Right hemisphere brain damage impairs strategy updating. *Cerebral Cortex*, *22*(12), 2745–2760.
De Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence*, *199*, 67–92.
Dennett, D. C. (1989). *The intentional stance*. MIT Press.
Devaine, M., Hollard, G., & Daunizeau, J. (2014). The social Bayesian brain: Does mentalizing make a difference when we learn? *PLoS Computational Biology*, *10*(12), Article e1003992.
Dyson, B. J. (2019). Behavioural isomorphism, cognitive economy and recursive thought in non-transitive game strategy. *Games*, *10*(3), 32.
Dyson, B. J., Steward, B. A., Meneghetti, T., & Forder, L. (2020). Behavioural and neural limits in competitive decision making: The roles of outcome, opponency and observation. *Biological Psychology*, *149*, Article 107778.
Dyson, B. J., Sundvall, J., Forder, L., & Douglas, S. (2018). Failure generates impulsivity only when outcomes cannot be controlled. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(10), 1483.
Dyson, B. J., Wilbiks, J. M. P., Sandhu, R., Papanicolaou, G., & Lintag, J. (2016). Negative outcomes evoke cyclic irrational decisions in rock, paper, scissors. *Scientific Reports (Nature Publisher Group)*, *6*(1), 20479.
Eyler, D., Shalla, Z., Doumaux, A., & McDevitt, T. (2009). Winning at rock-paper-scissors. *The College Mathematics Journal*, *40*(2), 125–128.
Forder, L., & Dyson, B. J. (2016). Behavioural and neural modulation of win-stay but not lose-shift strategies as a function of outcome value in rock, paper, scissors. *Scientific Reports (Nature Publisher Group)*, *6*(1), 33809.

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128.
Gershman, S. J., Gerstenberg, T., Baker, C. L., & Cushman, F. A. (2016). Plans, habits, and theory of mind. *PLoS One*, *11*(9), Article e0162246.
Goodie, A. S., Doshi, P., & Young, D. L. (2012). Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, *25*(1), 95–108.
Goodman, N. D., Baker, C. L., Bonawitz, E. B., Mansinghka, V. K., Gopnik, A., Wellman, H., et al. (2006). Intuitive theories of mind: A rational approach to false belief. In *Proceedings of the twenty-eighth annual conference of the cognitive science society*: *vol. 6*, Cognitive Science Society Vancouver.
Gopnik, A., & Wellman, H. M. (1992). *Why the child's theory of mind really is a theory*. Blackwell Publishing Ltd.
Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*(3), 407–420.
Guennouni, I., & Speekenbrink, M. (2022). Transfer of learned opponent models in zero sum games. *Computational Brain & Behavior*, *5*(3), 326–342.
Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, *85*(1), 1–36.
Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-beauty contests". *The American Economic Review*, *88*(4), 947–969.
Hoffman, M., Suetens, S., Gneezy, U., & Nowak, M. A. (2015). An experimental investigation of evolutionary dynamics in the rock-paper-scissors game. *Scientific Reports*, *5*(1), 8817.
Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604.
Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The naive utility calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, Article 101334.
Kangas, B. D., Berry, M. S., Cassidy, R. N., Dallery, J., Vaidya, M., & Hackenberg, T. D. (2009). Concurrent performance in a three-alternative choice situation: Response allocation in a rock/paper/scissors game. *Behavioural Processes*, *82*(2), 164–172.
Kleiman-Weiner, M., Ho, M. K., Austerweil, J. L., Littman, M. L., & Tenenbaum, J. B. (2016). Coordinate to cooperate or compete: abstract goals and joint intentions in social interaction. In *CogSci*.
Lie, C., Baxter, J., & Alsop, B. (2013). The effect of opponent type on human performance in a three-alternative choice task. *Behavioural Processes*, *99*, 87–94.
Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 392.
Manheim, D., & Garrabrant, S. (2018). Categorizing variants of Goodhart's law. arXiv preprint arXiv:1803.04585.
McCarthy, W. P., Hawkins, R. D., Wang, H., Holdaway, C., & Fan, J. E. (2021). Learning to communicate about shared procedural abstractions. In *Proceedings of the annual meeting of the cognitive science society*: *43*.
Mertens, J.-F. (1990). Repeated games. In *Game theory and applications* (pp. 77–130). Elsevier.
Moisan, F., & Gonzalez, C. (2017). Security under uncertainty: Adaptive attackers are more challenging to human defenders than random attackers. *Frontiers in Psychology*, *8*, 982.
Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, *36*(1), 48–49.
Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32.
Palacios-Huerta, I. (2003). Professionals play minimax. *Review of Economic Studies*, *70*(2), 395–415.
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526.
Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425.
Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, *121*(3), 352.
Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review*, *104*(3), 603–617.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.
Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Routledge.
Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.
Sepahvand, N. M., Stöttinger, E., Danckert, J., & Anderson, B. (2014). Sequential decisions: A computational comparison of observational and reinforcement accounts. *PLoS One*, *9*(4).
Serrino, J., Kleiman-Weiner, M., Parkes, D. C., & Tenenbaum, J. (2019). Finding friend and foe in multi-agent games. *Advances in Neural Information Processing Systems*, *32*.
Shachat, J., & Todd Swarthout, J. (2004). Do we detect and exploit mixed strategy play by opponents? *Mathematical Methods of Operations Research*, *59*, 359–373.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.
Sher, I., Koenig, M., & Rustichini, A. (2014). Children's strategic theory of mind. *Proceedings of the National Academy of Sciences*, *111*(37), 13307–13312.
Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., et al. (2016). Mastering the game of go with deep neural networks and tree search. *nature*, *529*(7587), 484–489.
Spiliopoulos, L. (2013). Strategic adaptation of humans playing computer algorithms in a repeated constant-sum game. *Autonomous Agents and Multi-Agent Systems*, *27*, 131–160.
Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, *10*(1), 218–254.
Stöttinger, E., Filipowicz, A., Danckert, J., & Anderson, B. (2014). The effects of prior learned strategies on updating an opponent's strategy in the rock, paper, scissors game. *Cognitive Science*, *38*(7), 1482–1492.
Thornton, M. A., & Tamir, D. I. (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences*, *114*(23), 5982–5987.
Thornton, M. A., & Tamir, D. I. (2021). People accurately predict the transition probabilities between actions. *Science Advances*, *7*(9), eabd4995.
Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552.
Tversky, A., & Kahneman, D. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454.
Ullman, T., Baker, C., Macindoe, O., Evans, O., Goodman, N., & Tenenbaum, J. (2009). Help or hinder: Bayesian models of social goal inference. In *Advances in neural information processing systems*: *vol. 22*.
van Opheusden, B., Kuperwajs, I., Galbiati, G., Bnaya, Z., Li, Y., & Ma, W. J. (2023). Expertise increases planning depth in human gameplay. *Nature*, 1–6.
Walker, M., & Wooders, J. (2001). Minimax play at Wimbledon. *American Economic Review*, *91*(5), 1521–1538.
Wang, Z., Xu, B., & Zhou, H. J. (2014). Social cycling and conditional responses in the rock-paper-scissors game. *Scientific Reports*, *4*, 5830.
West, R. L., & Lebiere, C. (2001). Simple games as dynamic, coupled systems: Randomness and other emergent properties. *Cognitive Systems Research*, *1*(4), 221–239.
Wu, S. A., Wang, R. E., Evans, J. A., Tenenbaum, J., Parkes, D. C., & Kleiman-Weiner, M. (2020). Too many cooks: Coordinating multi-agent collaboration through inverse planning. In *Proceedings of the annual meeting of the cognitive science society*: *42*.
Zhang, H., Moisan, F., Aggarwal, P., & Gonzalez, C. (2022). Truth-telling in a sender–receiver game: Social value orientation and incentives. *Symmetry*, *14*(8), 1561.
Zhang, H., Moisan, F., & Gonzalez, C. (2020). Rock-scissors: An exploration of the dynamics of players' strategies. In *Proceedings of the human factors and ergonomics society annual meeting*: *vol. 64*, (no. 1), (pp. 268–272). SAGE Publications Sage CA: Los Angeles, CA.
Zhang, H., Moisan, F., & Gonzalez, C. (2021). Rock-paper-scissors play: Beyond the win-stay/lose-change strategy. *Games*, *12*(3), 52.