Humans fail to outwit adaptive rock, paper, scissors opponents

Erik Brockbank (ebrockbank@ucsd.edu)
Department of Psychology, 9500 Gilman Ave.
La Jolla, CA 92130 USA

Edward Vul (evul@ucsd.edu)
Department of Psychology, 9500 Gilman Ave.
La Jolla, CA 92130 USA

Abstract

How do humans adapt when others exploit patterns in their behavior? When can people modify such patterns and when are they simply *trapped*? The present work explores these questions using the children's game of rock, paper, scissors (RPS). Adult participants played 300 rounds of RPS against one of eight bot opponents. The bots chose a move each round by exploiting unique sequential regularities in participant move choices. In order to avoid losing against their bot opponent, participants needed to recognize the ways in which their own behavior was predictable and disrupt the pattern. We find that for simple biases, participants were able to recognize that they were being exploited and even counter-exploit their opponents. However, for more complex sequential dependencies, participants were unable to change their behavior and lost reliably to the bots. Results provide a quantitative delineation of people's ability to identify and alter patterns in their past decisions.

Keywords: adaptive behavior; exploitation; behavioral game theory; rock-paper-scissors

Introduction

The ability to counteract scenarios in which others seek to trap or exploit us is central to adaptive behavior and intelligence. Reasoning in these scenarios forms the basis for games like squash and chess, in which players must avoid being placed in situations from which they are unable to escape. However, it also underlies complex decision making at a larger scale. Spies are taught to avoid giving away possible signs of their status. A familiar feature of modern IT trainings involves detecting attempts at "phishing", in which a malicious actor obtains login credentials or money through seemingly benign requests. Such examples highlight two important features of human adaptive behavior. First, the stakes of avoiding attempted exploitation can often be high, as when U.S. Democratic National Committee servers were breached in 2016 through a phishing campaign, and second, people often struggle to recognize or counteract exploitative behavior, whether on a chess board or at the hands of scammers. In the current work, we attempt to better understand how people respond adaptively to exploitative behavior over repeated interactions, and when they fail to do so.

Though a broad range of behaviors might qualify as exploitative (e.g., deception), we focus on actions that capitalize on a particular regularity or pattern in another's behavior which makes them predictable. Consider for example the ways in which software is exploited or, as in the previous examples, human behavioral tendencies make their responses in chess or phishing scams more easily anticipated. Here, we investigate the levels of complexity at which people are able to represent exploitation by another agent, and how well they can in turn avoid or even counter-exploit these situations. In other words, how richly can people represent patterns in their own behavior which might form the basis for exploitation? And how easily can they change such behavior?

Prior work on human *subjective randomness* offers insights into how people represent and counteract exploitable patterns in their own behavior. When asked to judge the randomness of a sequence or produce one themselves, people are subject to systematic biases which typically cause them to favor certain sequences over other equally (or more) random ones (Bar-Hillel & Wagenaar, 1991). To illustrate, when prompted to produce a sequence of simulated coin flips, people often produce sequences that (*i*) have an equal number of heads and tails, (*ii*) under-represent "runs" (e.g., HHH) and (*iii*) over-represent alternations (HTH) (Lopes & Oden, 1987). This tendency is so robust that Rapoport and Budescu (1992) note, "Cognitive psychology has engendered few examples of so much support for and agreement among researchers about the prevalence of a cognitive bias."

A natural question is whether people can counteract such patterns in their own behavior. In one study, participants who were repeatedly given feedback about random binary signals they had generated—using measures of signal strength, alternations, and conditional dependence—eventually produced sequences that were indistinguishable from computer generated ones with respect to the measures they had been shown (Neuringer, 1986). Similarly, studies with Matching Pennies, a two-player game in which the Nash Equilibrium strategy (Nash, 1950) is to choose moves randomly, found that in an adversarial setting, people produced more random sequences of events than when they were asked to generate individual sequences of moves (Rapoport & Budescu, 1992; Budescu & Rapoport, 1994). Finally, analyses of the direction of tennis serves and soccer penalty kicks suggests that professional athletes are able to produce highly randomized event sequences (Walker & Wooders, 2001; Palacios-Huerta, 2003).

Broadly, these findings suggest that people can, with certain levels of feedback or skill, and notably in adversarial settings, counteract robust patterns in their sequential behavior. Here, we seek a broader, formalized account of how this process works. What kinds of behavioral regularities or biases

can people successfully "undo", and what representational capacities underlie this process? We analyze behavior in a domain which has a wider range of behavioral dependencies or biases than the simple patterns found in subjective randomness. In the current study, we explore behavior in the children's game of rock, paper, scissors (RPS). As with Matching Pennies and other *mixed strategy equilibrium* games, the equilibrium strategy in RPS is to select moves randomly; any predictable pattern could be exploited by the opponent. However, unlike coin flips or Matching Pennies, prior work has found evidence for an array of sequential regularities in people's decision making in repeated games of RPS. For example, people with schizophrenia may be more likely to choose the *Cournot Best Response* (Cournot, 1838), i.e., playing the move which will beat what one's opponent just played (Baek et al., 2013). Similarly, prior work has found evidence for sub-conscious imitation of an opponent when playing the game in person (Cook, Bird, Lünser, Huck, & Heyes, 2012), though results are mixed (Aczel, Bago, & Foldes, 2012). The most widely studied behavioral pattern is *win-stay, lose-shift*, a strategy that has applications in game theory (Nowak & Sigmund, 1993), reinforcement learning (Erev et al., 2010), and Bayesian estimation (Bonawitz, Denison, Gopnik, & Griffiths, 2014). Win-stay, lose-shift responding has been observed in RPS play in shuffled groups (Wang, Xu, & Zhou, 2014), and more recent work has explored the basis for this pattern in adversarial settings (Dyson, Wilbiks, Sandhu, Papanicolaou, & Lintag, 2016). Building on these results, we have shown in previous work that in repeated games of RPS with human dyads, people exhibit a range of stable regularities in their move choices, including the ones outlined above, as well as a number of others combining player previous moves, opponent previous moves, and previous round outcomes (Brockbank & Vul, 2020). Notably, these findings quantify how much people exhibit a particular dependency, allowing for robust comparison of different ways in which people display patterned behavior.

The present work leverages these results to investigate how people respond to exploitation. Given the various regularities in people's decision making, which ones can they avoid or counteract when those dependencies are being exploited? The results described previously leave these questions unanswered, since it cannot be determined from dyad play which of the dependencies in players' move choices their opponents were aware of or were using to determine their own moves. How then might we expect people to behave when reliable patterns in their behavior are being exploited? Prior research on sequence learning and word segmentation suggests that in some domains, humans are highly sensitive to regularities in their own and others' behavior. For example, given repeated practice, people can reliably learn patterns consisting of as many as 10 items in a range of motor and visuospatial domains (Nissen & Bullemer, 1987; Clegg, DiGirolamo, & Keele, 1998). In a similar vein, Saffran, Johnson, Aslin, and Newport (1999) showed that adults and infants are sensitive to statistical co-occurrence among 11 non-linguistic tones, suggesting that our ability to learn word boundaries recruits more domain-general pattern learning. If reasoning about patterns in one's own or another's behavior recruits cognitive processes like those underlying sequence learning or word segmentation, we might expect people to be responsive to patterns in their actions which lead to exploitation. However, it is not clear that this ability extends to more abstract decision making in adversarial interactions like rock, paper, scissors. Here, we address this challenge by pairing participants with bot opponents that exploit various behavioral dependencies observed in human RPS play. We find that people can easily outwit opponents that exploit some behavioral dependencies, but that they are reliably beaten when opponents leverage others; these differences align closely with the complexity of the dependencies. In this way, we provide a precise account of sequential patterns for which people are able to counteract exploitation, and establish clear limits in people's ability to represent and alter regularities in their own behavior.

Experiment

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. The experiment was coded following guidelines for synchronous game play outlined in Hawkins (2015) and is available along with all data and analyses on github at: `https://www.github.com/erik-brockbank/rps`.

Procedure

Participants completed the experiment in a web browser on their home computers. Participants began by clicking through a set of instructions explaining the rules of rock, paper, scissors and the format of the experiment. After reading the instructions, participants were randomly assigned to play one of eight bot opponents. Participants were not told anything about the identity or strategy of their opponent during the instructions or during gameplay.

The experiment consisted of 300 rounds of RPS played against the bot opponent. In each round, participants were first shown a screen with a clickable "card" illustrating each move choice, as well as a matching panel used to illustrate their opponent's choice once the round was complete. At the bottom of the screen, an illustration of the rules remained throughout the duration of the game to avoid any possibility of moves chosen due to misunderstanding. Participants had 10 seconds to choose a move each round; a countdown timer at the top of the screen showed their time remaining. If participants did not select a move in this time, they lost the round.

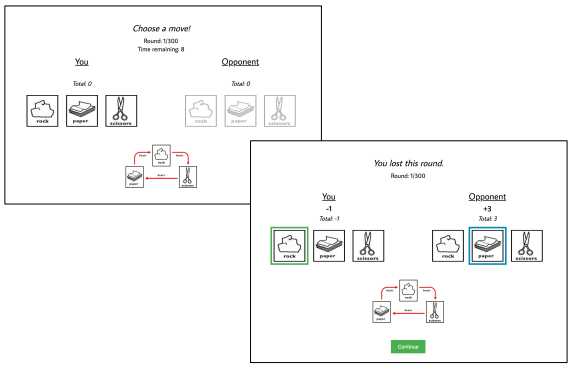

Figure 1: The stages of each rock, paper, scissors round. Top left: participants had 10s to select their move. Bottom right: participants were shown the results of each round, along with updated points for them and their opponent, before clicking to proceed to the next round.

Once participants had chosen a move, they were immediately shown the results of that round. Their opponent's move was highlighted next to their own, and a message at the top indicated the outcome, as well as the points awarded to the participant and their opponent. In each round, the winner received 3 points, the loser received -1 point, and in the event of a tie, both players received 0 points[1]. After viewing the results of a given round, participants clicked a button to proceed to the next round. Once they clicked "Continue", the next round began immediately. Throughout the duration of the experiment, participants were shown their cumulative points and their opponent's cumulative points to motivate them to win over repeated rounds, as well as the current round index out of 300. The phases of each round are shown in Figure 1.

After completing all 300 rounds against their bot opponent, participants were taken to a post-game questionnaire. Here, they were first given a free response prompt which asked them to describe their strategy (*"In the text box below, please describe any strategies you used to try and beat your opponent"*). Then, they were prompted with a series of statements about their game play and asked to respond on a seven point Likert scale ranging from Strongly Disagree to Strongly Agree (for example, *"I paid attention to my opponent's moves in order to try and predict their next move"* and *"There were noticeable patterns in my opponent's moves that allowed me to predict their next move"*).

To understand how responsive people were to an opponent that was exploiting patterns in their behavior, we measured the average *win count differential* for the bots using each strategy. Each bot's win count differential is the number of wins it obtained over 300 rounds, minus the number of wins for its opponent. The bot's win count differential in a game provides a measure of how much the bot was able to exploit its (human) opponent. Two opponents selecting moves randomly would each be expected to have a win count differential of zero; an average win count differential greater than zero indicates that the bot was able to exploit participants successfully.

Adaptive Bot Strategies

Participants were paired with one of eight adaptive bot opponents for the duration of the experiment. Each bot had a policy of choosing the move that would beat whatever move they determined was most likely for their human opponent in the next round. In the event that multiple moves were considered equally probable, the bot sampled one at random and chose the move that beat the sampled move[2]. The bots differed in how they determined their opponent's most likely move each round, relying on distinct sequential dependencies their opponent might exhibit. To illustrate, a naïve approach would involve simply tracking the participant's cumulative proportion of *rock*, *paper*, and *scissors* choices, then selecting a move each round that would beat whichever opponent move is most likely. A more complex bot might instead track the ongoing sequences of a participant's last five moves and choose the move each round that beats whatever choice is most probable given the participant's previous four moves. In repeated dyadic RPS games with a human opponent, people exhibit a number of stable dependencies in their move choices (Brockbank & Vul, 2020). The eight adaptive bots in the current experiment each exploited one of the dependencies outlined in Brockbank and Vul (2020), choosing their move each round based on the particular regularity they were exploiting. We outline each of these bot strategies below.

Transition baserate: A participant's move in a given round can be thought of as reflecting a particular *transition* from their previous move (Dyson, 2019). Any time they play a move which beats the previous move (e.g., *paper* after *rock*), this is a "positive" transition or shift up, denoted here with a $+$. Meanwhile, any time they play a move which loses to the previous move, this represents a "negative" transition or shift down ($-$), and any time they repeat the previous move, this is a "stay" transition (0). The `Transition baserate` bot tracks participant transitions and chooses the move each round which beats their opponent's most likely transition.

Opponent transition baserate: The notion of a transition between moves can just as easily be described relative to an *opponent's* previous move (Dyson, 2019). The `Opponent transition baserate` bot chooses a move based on the participant's most likely transition relative to their *bot opponent's* previous move.

Transition given player's prior choice: This bot keeps track of 1-back sequential dependencies in player moves: Does the player tend to play *rock* after playing *paper*? This is similar to the `Transition baserate` strategy but allows for the possibility that a participant's most likely move after *rock*

[1]The imbalance in points allocated to wins and losses was chosen to make the game more engaging for participants, so that even with an equal number of wins and losses, participants would maintain a positive score.

[2]In this way, the bots maximized expected win count, but did not maximize expected *win differential*. This means bots were indifferent between moves that had a 50/50 change of a win or tie, and those that had a 50/50 chance of a win or loss.

is, e.g., *paper* (a + transition), but their most likely move following *paper* may be *paper* again (a 0 transition). This bot therefore tracks every sequence of two moves rather than assuming that transition rates are independent of prior move.

Transition given opponent's prior choice: This bot is identical to the `Transition given player's prior choice` bot, except that it exploits any pattern in participant move choices based on their *bot opponent's* previous move rather than their own previous move.

Transition given prior outcome: This bot tracks a participant's most likely transitions conditioned on each possible previous outcome. Win-stay, lose-shift behavior (Wang et al., 2014; Dyson et al., 2016) will be exploited by this bot.

Choice given player's prior choice & opponent's prior choice: This bot tracks player move choices given each *combination* of their own and their bot opponent's previous move.

Choice given player's prior two choices: This bot chooses the move which beats their human opponent's most likely move given the participant's move choices in each of the previous two rounds.

Transition given prior transition & prior outcome: This bot exploits any dependency participants exhibit on 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 will detect such a pattern and exploit it.

Adaptive Bot Complexity

Intuitively, the eight bot strategies described above differ in the *complexity* of the behavioral regularity they are exploiting. We formalize the complexity of a bot's strategy based on the memory demands of executing it. The simplest transition bots (`Transition baserate` and `Opponent transition baserate`) need only store and update three counts in memory: a 1x3 matrix with the number of +, −, and 0 transitions. Meanwhile, the three intermediate strategies (`Transition given player's prior choice` and `Transition given opponent's prior choice` and `Transition given prior outcome`) all maintain a 3x3 (9-cell) matrix of transition counts based on additional information from the previous round (i.e., the participant's previous move, the bot's previous move, or the previous outcome). Finally, the most complex bots (`Choice given player's prior choice & opponent's prior choice` and `Choice given player's prior two choices` and `Transition given prior transition & prior outcome`) rely on a 9x3 (27-cell) matrix which tracks unique combinations of previous events (two previous moves or a previous transition and previous outcome) to choose their moves. While this is not the only way to formalize the complexity of these strategies, the memory requirement offers an intuitive description of why some of the regularities the bots exploit may be easier to counteract than others. We explore how well the complexity of a regularity (and the corresponding bot's strategy) maps onto people's ability to adapt to each bot.

Results

At a high level, our results indicate that there is considerable variance in the degree to which people were able to avoid exploitation by their adaptive bot opponents. We first show that bot win count differentials are highly correlated with the *expected win count differentials* obtained in Brockbank and Vul (2020); the extent to which people exhibit a given dependency in dyad play is closely aligned with their (in)ability to counteract such a pattern when it is being exploited. Next, our results offer insights into *why* people are able to recognize and alter some patterns in their behavior but not others. We show that increases in bot win count differentials follow increases in the *complexity* of the dependency the bots exploit, with certain complexity levels reliably exploiting human opponents and others frequently *counter-exploited* by participants.

Patterns of exploitability against bots align with dyad play

We first explore whether the dependencies people exhibited most in dyadic play (Brockbank & Vul, 2020) are in turn the ones that led to the highest win count differentials for exploitative bots. While this is an intuitive hypothesis, it need not be the case; against a stable human opponent, participants may have exhibited certain dependencies to a greater degree *precisely because their opponents were not exploiting these dependencies*. Without any pressure to disrupt such regularities, they may have persisted more than they would against an exploitative opponent. Therefore, it remains an open question whether the dependencies people exhibited most in dyad play are the ones they are least able to recognize and counteract.

For a given dependency over move choices, Brockbank and Vul (2020) calculate an *expected win count differential* for each participant based on how much they exhibit that dependency in game play. This represents an approximation of how much a hypothetical opponent could be expected to win if they were to simply maximize against this dependency when playing each participant; the higher the expected win count differential for a given dependency, the more participants exhibited that dependency in dyad play. Figure 2 shows expected win count differentials from Brockbank and Vul (2020) on x and average win count differentials for each adaptive bot in our results on y. Each point therefore indicates how successfully the bot opponents were able to exploit a given dependency, as a function of how much people exhibited that same dependency in dyad play. The correlation between average *expected* win count differentials and the matched average *bot* win count differentials for each strategy is 0.958 ($t(6) = 8.16$, $p < 0.001$). As the amount that people exhibited a particular dependency in dyad play increases, so too does the win count differential obtained by exploiting this dependency. This suggests that people's tendency to exhibit some regularities more than others when making strategic decisions extends to their inability to *adapt* their play when such regularities are being exploited.

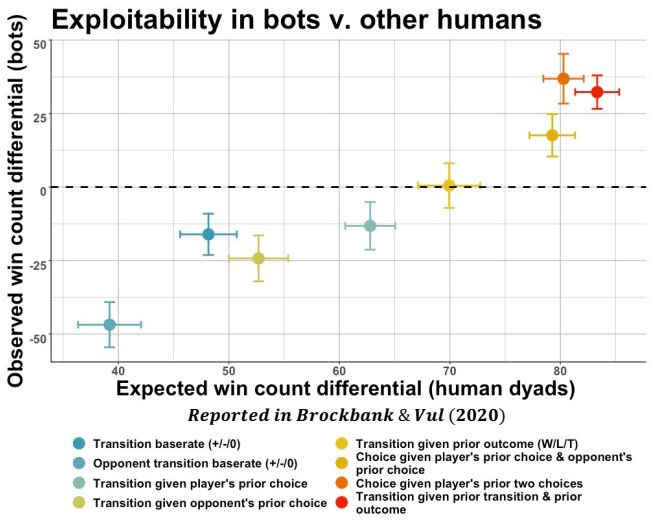

Figure 2: Relationship between expected win count differentials from human dyad play reported in Brockbank and Vul (2020) and bot win count differentials in our data. Each point represents a behavioral pattern exhibited in dyad play and exploited by one of the eight bots. The dashed line indicates chance performance.

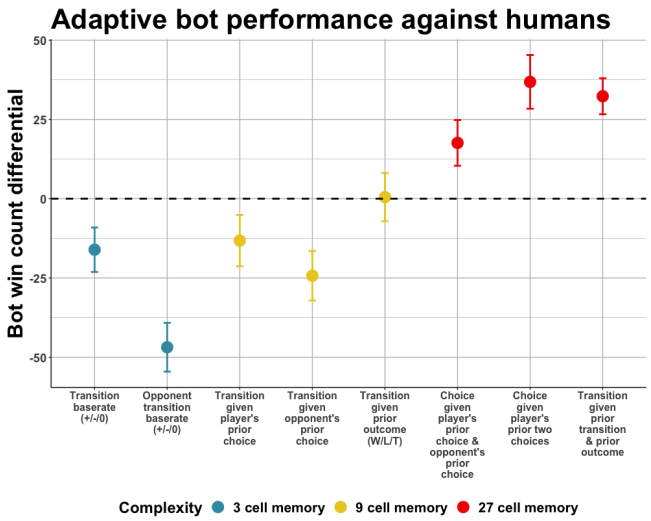

Figure 3: Average win count differentials for each adaptive bot strategy. Positive win count differentials indicate that the bots were able to "outsmart" their human opponents. Adaptive bot complexity falls into three distinct categories. Bots exhibit a clear pattern of exploiting more complex human move dependencies more successfully. Error bars indicate one SEM. The dashed line indicates chance performance.

In sum, findings from Brockbank and Vul (2020) suggest that against other humans, people strongly exhibit the dependencies being exploited by adaptive bots in the present experiment. We therefore expect that people's baseline tendency to display these biases in the current results will be minimally different. The question is whether an agent that exploits such behavioral regularities will lead to adaptive behavior among participants, or whether they will be unable to modify biases in their decisions. Results in Figure 2 suggest that indeed, the more strongly people display a bias in dyadic interaction, the more that bias can be exploited by a calculating opponent.

Game outcomes predicted by strategy complexity

While the *expected* win count differentials reported in Brockbank and Vul (2020) are all positive (reflecting idealistic assumptions about exploitability), a notable feature of our results is that *bot* win count differentials straddle the intercept. In Figure 2, expected win count differentials on x lie roughly between 40 and 90, while average bot win count differentials on y range from -50 to 50. For some bots, not only did participants avoid being exploited by a particular regularity, but they successfully *counter-exploited* the bot opponent, giving the bot a negative win count differential. However, for other dependencies, participants were unable to avoid being exploited, producing average bot win count differentials greater than zero. How might differences between the adaptive bot strategies account for this variance? Figure 3 shows average bot win count differentials by strategy, arranged according to the memory complexity described previously. Results illustrate a clear relationship between a bot's complexity and whether it was able to exploit its human opponents.

First, bot win count differentials are significantly less than zero for the simplest 3-cell memory bots in blue in Figure 3 (`Transition baserate`: $t(20) = -2.30$, $p = 0.03$; `Opponent transition baserate`: $t(25) = -6.09$, $p < 0.01$). Bot win count differentials were negative for 17 out of 21 and 22 out of 26 participants paired with each bot, far fewer than would be expected by chance ($p < 0.01$ and $p < 0.001$). The only way to achieve this pattern of results is for participants to discover a means of counter-exploiting the bot opponents by picking up on the higher-order dependencies exhibited by the bots as they adapt to simpler dependencies. These results suggest that participants reliably recognize and even successfully outwit strategies based on the simplest behavioral regularities.

In contrast, the three 27-cell bot strategies which exploited the most complex dependencies in human move choices were able to consistently beat their human opponents, shown in red in Figure 3. Bot win count differentials are significantly greater than zero for these strategies (`Choice given player's prior choice & opponent's prior choice`: $t(24) = 2.44$, $p = 0.02$; `Choice given player's prior two choices`: $t(19) = 4.36$, $p < 0.001$; `Transition given prior transition & prior outcome`: $t(25) = 5.68$, $p < 0.0001$). For the two bot strategies with the highest win count differentials, only 4 out of 20 and 4 out of 26 participants in each condition had win count differentials greater than or equal to zero; this is far fewer than would be expected by chance ($p = 0.01$ and $p < 0.001$) and suggests that most people paired with these bots would not be expected to come out ahead over many rounds. People are essentially trapped in these complex de-

pendencies, suggesting a clear limit to the regularities they can recognize and adjust in their own behavior.

Finally, among the intermediate 9-cell memory bots, two obtained win count differentials which were not significantly different from zero (`Transition given player's prior choice`: $t(23) = -1.63$, $p = 0.12$; `Transition given prior outcome`: $t(21) = 0.07$, $p = 0.95$) while the third was significantly less than zero (`Transition given opponent's prior choice`: $t(27) = -3.11$, $p < 0.01$). Binomial tests based on individual win count differentials reflect a similar pattern. Thus, at these intermediate complexities people cannot counter-exploit the adaptive bots, but can avoid being exploited themselves. Even though people show a high degree of exploitability for these dependencies in dyadic play (Brockbank & Vul, 2020), this level of complexity occupies an intermediate point at which participants were able to effectively minimize the degree to which their behavior was exploited, but could not respond strategically to counter-exploit bots with this level of sophistication.

Consistent player motivation Are people equally motivated when playing the different types of bots we pit them against? When paired with other humans, people are sufficiently motivated by the ongoing scores we present to try to outwit one another (Brockbank & Vul, 2020). However, perhaps this does not hold when playing against synthetic agents. In particular, we want to make sure that motivation is consistent across all bot types. Three of the five Likert scale questions on the post-experiment survey addressed effort and motivation: (*i*) *I paid attention to my opponent's moves in order to try and predict their next move*, (*ii*) *I was focused on winning for the entire time I was playing*, and (*iii*) *I was trying to win each round against my opponent*. We asked whether responses on these questions differed as a function of bot strategy. An analysis of variance of these responses as a function of bot strategy did not find any significant effects (*i*. $F(7, 181) = 1.64$, $p = 0.13$; *ii*. $F(7, 181) = 0.52$, $p = 0.82$; *iii*. $F(7, 181) = 0.87$, $p = 0.53$). Thus, while it is perhaps natural that some participants may have experienced frustration at being consistently beaten by a high complexity opponent, it is unlikely that this accounts for the results in Figure 3.

Discussion

The current work explores how people respond to an opponent that seeks to exploit predictable patterns in their actions. This sort of challenge is central to games like chess and tennis, but also underlies sophisticated behavior like online scams or negotiations. Specifically, we ask what levels of patterned complexity people are able to recognize in their own behavior, and how flexibly they can counteract such patterns to avoid exploitation. To address these questions, we analyze move choices in the game of rock, paper, scissors (RPS), a setting in which prior work has found evidence for a range of sequential regularities in people's decision making. Participants played 300 rounds of RPS against one of eight bot opponents, each of which strategically exploited a unique dependency found in earlier results (Brockbank & Vul, 2020). We compare each bot's average win count differential—their total wins minus their (human) opponent's total wins—to understand what kinds of patterns people can reliably avoid, and when they are instead trapped by an exploitative adversary.

The bot strategies tested here captured a wide range of human responses to exploitation; participants were able to counter-exploit and win reliably against some of the bot strategies, while consistently losing to others. To explain these differences, our results offer two novel findings. First, we show that the extent to which the adaptive bots were able to successfully exploit regularties in participant responses is proportional to how much participants exhibited the same regularities against other people (Brockbank & Vul, 2020). This suggests that people's capacity to recognize and alter patterns in their own behavior may be tied to how much they exhibit those patterns naturally. Second, we show that the degree to which people were exploited by a particular dependency is closely aligned with the memory demands of tracking that dependency. In other words, the complexity of the behavioral pattern predicts people's (in)ability to break it.

While the current results offer novel perspectives on the ways people respond to exploitation, other interpretations of these data are worth considering. For one, the *complexity* of the dependencies that the bots exploit may be better captured through other measures; for example, rather than the memory requirements for representing the underlying dependency, bot strategies can be classified according to the number of *state variables* that are needed, e.g., a player's previous move, the opponent's previous move, or the outcome of the previous round. Thus, while it is tempting to blame the memory demands associated with tracking a more complicated dependency structure, the fault may just as well lie in processing constraints on using such structures. In this vein, future work should pull apart the impact of different complexity variables in adaptive behavior. For example, a version of the current task in which the bot's strategy is visible throughout the experiment would isolate the computations required to respond and reveal limits in adaptive reasoning while controlling for the memory complexity of strategy representations.

The findings presented here raise a number of additional questions about adversarial behavior that merit further investigation. First, our results suggest that *within* a particular complexity level, people vary in their ability to adapt to exploitation. Notably, across all three complexity levels in our data, participants were more successful at counteracting dependencies in their opponent's behavior than their own past actions (e.g., `Opponent transition baserate` versus `Transition baserate`). As far as we are aware, existing work on strategic reasoning does not offer a clear explanation for this pattern. Future work should explore the possible role of *player-relative* and *opponent-relative* dependencies in people's ability to adapt to exploitation. In addition to variability within complexity levels, our results show individual variability within bot strategies, i.e., differential levels of success

responding to a bot's exploitative behavior. This variability likely reflects the use of different cognitive strategies across individuals, for example *model-free* reinforcement learning strategies or *model-based* predictions (Sepahvand, Stöttinger, Danckert, & Anderson, 2014). Future work might examine the time course or win patterns against exploitative opponents to better understand individual strategy differences.

Broadly, the current results speak to the behavioral patterns people can detect and adjust, and thus might be instructive for real world settings where people are expected to revise their behavior to undo potentially complex patterns. For instance, work in "explainable AI" (Gunning et al., 2019) pursues a human-legible description of the behaviors that prompted a particular algorithmic decision, e.g., which aspects of a person's behavior led to a rejected loan request. The goal of this criterion is to allow people to respond positively to an adverse decision by changing their behavior. But what constitutes a sufficiently simple behavioral explanation for people to adjust patterns in their actions? While our results cannot provide a general answer, they suggest limits in the complexity of patterns people can detect and adjust in their own behavior, a key to successful human-AI interactions.

References

Aczel, B., Bago, B., & Foldes, A. (2012). Is there evidence for automatic imitation in a strategic context? *Proceedings of the Royal Society B: Biological Sciences*, *279*(1741), 3231–3233.

Baek, K., Kim, Y. T., Kim, M., Choi, Y., Lee, M., Lee, K., ... Jeong, J. (2013). Response randomization of one-and two-person rock–paper–scissors games in individuals with schizophrenia. *Psychiatry research*, *207*(3), 158–163.

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in applied mathematics*, *12*(4), 428–454.

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-stay, lose-sample: A simple sequential algorithm for approximating bayesian inference. *Cognitive psychology*, *74*, 35–65.

Brockbank, E., & Vul, E. (2020). Recursive adversarial reasoning in the rock, paper, scissors game. In *Proceedings of the 42nd annual conference of the cognitive science society* (pp. 1015–1021). Cognitive Science Society.

Budescu, D. V., & Rapoport, A. (1994). Subjective randomization in one- and two-person games. *Journal of Behavioral Decision Making*, *7*(4), 261–278.

Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. *Trends in cognitive sciences*, *2*, 275–281.

Cook, R., Bird, G., Lünser, G., Huck, S., & Heyes, C. (2012). Automatic imitation in a strategic context: players of rock–paper–scissors imitate opponents' gestures. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 780–786.

Cournot, A. (1838). *Recherches sur les principes mathematiques de la theorie des richesses*. Paris: Hachette. English translation (N. Bacon trans.): Research into the mathematical principles of the theory of wealth.

Dyson, B. J. (2019). Behavioural isomorphism, cognitive economy and recursive thought in non-transitive game strategy. *Games*, *10*(3), 32.

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.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., ... Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, *23*(1), 15–47.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). Xai—explainable artificial intelligence. *Science Robotics*, *4*(37).

Hawkins, R. X. (2015). Conducting real-time multiplayer experiments on the web. *Behavior Research Methods*, *47*, 966–976.

Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 392.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, *36*(1), 48–49.

Neuringer, A. (1986). Can people behave "randomly?": The role of feedback. *Journal of Experimental Psychology: General*, *115*(1), 62.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, *364*(6432), 56–58.

Palacios-Huerta, I. (2003). Professionals play minimax. *The Review of Economic Studies*, *70*(2), 395–415.

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, *121*(3), 352.

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*, 27–52.

Sepahvand, N. M., Stöttinger, E., Danckert, J., & Anderson, B. (2014). Sequential decisions: a computational comparison of observational and reinforcement accounts. *PloS one*, *9*(4).

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.