Paper-Rock-Scissors: an exploration of the dynamics of players' strategies

Hanshu Zhang
Carnegie Mellon University

Frederic Moisan
University of Cambridge

Cleotilde Gonzalez
Carnegie Mellon University

This research contributes to the understanding of dynamic decision making behavior in adversarial repeated interactions. Using a well-known competitive game, Rock-Paper-Scissors in a two-player experiment, we collected data of repeated play in pairs over many trials. We design a payoff matrix that allows us to distinguish the optimal (Nash) behavior from random behavior. Our analyses indicate that participants do not play in agreement with Nash or random. We also do not find evidence of the cyclic behavior suggested in the literature. Interestingly, human behavior is very heterogeneous. While some players follow the common "Win-Stay/Lose-Shift" heuristic, many others also follow a "Win-Shift/Lose-Stay" heuristic. We summarize our conclusions for the study of the dynamics of behavior in adversarial situations.

INTRODUCTION

Almost everyone have settled disputes by playing a simple game called Paper-Rock-Scissor (PRS). The rule for winning a one-shot play of this game is simple: rock crushes scissors, scissors cut paper, and paper covers rock. But in addition to being a fun game to resolve disagreements, RPS is also a serious game used by game theorists and psychologists to study competitive behavior in naturalistic settings, such as security, terrorism and war (Fisher, 2008). Because none of the strategies (P, R, or S) is absolutely better than either of the other two, PRS is an interesting research paradigm to study adversarial behavior and dynamics of player's strategies. For example, rock can beat paper but at the same time rock can be beaten by scissors. However, most previous work focused on one-shot games involving models and humans or two humans but with re-shuffling of the players rather than observing repeated plays of the same pair (but see Hoffman, Suetens, Gneezy, & Nowak, 2015, for exceptions). In this paper we investigate the dynamics of players strategies in a setting that involves many sequential trials of the same pair.

In the traditional form of the PRS game a win gives 1 point to the winner and takes 1 point from the loser and a tie gives 0 points to both players. With this setting, a player who plays RPS randomly has a 1/3 chance of winning in any round. Importantly, this is also the optimal strategy — i.e., *Nash Equilibrium* strategy (the strategy where no player's deviation is beneficial): in each trial, if player 1 chooses each strategy 1/3 of the time, player 1's payoff is 0 regardless of player 2's strategy. Otherwise, player 2 could find a combination strategies that would win the game with a positive expected value. Thus, in all previous behavioral research with this traditional zero-sum symmetrical payoff design, it is not possible to distinguish whether players are in agreement to Nash or random strategy. Some researchers have found evidence of strategies that are consistent with Nash, suggesting that experienced players who use information of previous plays strategically are more likely to win (Batzilis, Jaffe, Levitt, List, & Picel, 2019). However, most research observed that there is considerable deviation from equilibrium play (e.g. Eyler, Shalla, Doumaux, & McDevitt, 2009; Hoffman et al., 2015). This latter observation is more in agreement with the well-known gap between human behavior and rational solutions. Indeed, most humans would follow a *satisfying* (i.e., good enough) strategy, rather than an optimizing strategy (Simon, 1956). In this paper we employ the specific payoff setting designed to detect whether participants play in agreement to Nash or random strategy.

Since Nash Equilibrium appears to fail to describe players' behavior in the PRS game, many studies have focused on the investigation of simple heuristics that are more "psychologically plausible" for boundedly rational humans (Wang, Xu, & Zhou, 2014). For example, Eyler et al. (2009) found that players often repeat the same action 3 consecutive times (PPP, RRR, SSS) or mix the actions in any order (e.g., PRS, RSP, SPR). In agreement with these findings, West and Lebiere (2001) concluded that a model with a "lag-2" strategy (i.e., players attended to the opponent's last 2 actions) provided a close representation of human PRS play. This "cycling" behavior is also suggested by others (e.g. Dyson, Wilbiks, Sandhu, Papanicolaou, & Lintag, 2016; Forder & Dyson, 2016), which suggests that the strategies depend on the outcomes. Specifically, following P or S strategies, players were more likely to switch than select the same action again; and a loss or a draw prompted more switching than staying.

This type of strategy, reflecting a common "win-stay/lose-shift" (WSLS) heuristic has received most empirical support. In WSLS a participant would keep the same strategy after a success and shift to another strategy after a failure. This is a simple, cognitively plausible heuristic, well supported by principles of behaviorism such as Thorndike's Law of Effect (Thorndike, 1911) and the matching law (Herrnstein, 1961), where the proportion of responses matches the degree of reinforcement. Wang et al. (2014) found that participants implement WSLS in the RPS leading to successful play. Dyson et al. (2016) found that participants were more likely to switch their item selection at trial n + 1 following a loss or draw at trial n, revealing a strategic vulnerability of individuals following the experience of negative rather than positive outcome. Similarly, Forder and Dyson (2016) found greater reliance on "lose-shift" than on "win-stay".

To summarise, we study the dynamics of strategies used by pairs of players playing the competitive RPS repeatedly. Our design allows us to investigate whether humans play randomly or Nash by making Nash a different solution from random play. We also investigate evidence for the commonly supported WSLS strategy.

EXPERIMENT

Participants

Ninety-six players on Amazon Mechanic Turk (MTurk) participated and completed the study (Age: [18,64], Female = 36). It took 14 minutes on average for participants to finish the task. Participants who finished the task received an average payoff of \$1.5. Four pairs were excluded from data analysis because least one player in the pair chose the same action over 50% of the trials. This left 45 pairs (90 participants) in the final data analysis.

Design

We developed a web application where available Mturk workers who had accepted to participate in our study, were paired up with another available participant to play the PRS game.

Importantly, we designed a novel payoff matrix Table 1, in which the Nash equilibrium is different from the Random strategy. The Nash strategy is a mix of 1/4, 1/2, 1/4 for rock, paper, and scissors respectively, resulting as follows: In a two-player PRS game, player 1 chooses rock, paper, scissor with the probability $i = \{p_{R1}, p_{P1}, p_{S1}\}$ and player 2 chooses $j = \{p_{R2}, p_{P2}, p_{S2}\}$. Thus in our experiment design, the expected payoff for player 1 of playing rock $E_{Rock} = 2 \times p_{R2} + 1 \times p_{P2} + 4 \times p_{S2}$ equals the expected payoff of playing paper $E_{Paper} = 3 \times p_{R2} + 2 \times p_{P2} + 1 \times p_{S2}$ and playing scissor $E_{Scissor} = 0 \times p_{R2} + 3 \times p_{P2} + 2 \times p_{S2}$ when player 2 choosing $j = \{p_{R2} = 1/4, p_{P2} = 1/2, p_{S2} = 1/4\}$. Because it is a symmetric game, player 1 should also have the same choice probability so that players 2's expected value for choosing each strategy is same as well, which gives us the solution for Nash Equilibrium. The random strategy continues to be 1/3, 1/3, 1/3.

In addition, to avoid the effect of real losses (Forder & Dyson, 2016), we use only positive numbers in the payoff matrix.

Procedure

Participants were asked for informed consent according to the protocol approved by the Institutional Review Board at Carnegie Mellon University. Then all players read the same general task instructions, before they were redirected to fill in a brief demographic survey about their age, gender, residency, and education level.

After finishing the survey, participants were matched with another online participant available in the MTurk pool. A participant waited to be matched to another participant for as long as 10 minutes. If no other participant was found within this time frame, then the participant was thanked and paid a basic amount for the waiting time. If another participant was available, a match was done. Participants were not given any information about the player they were matched with. After the match, they played the PRS for 100 trials. After each of the participants in a pair made a choice (P, R, or S), they were notified of the points obtained from choices they made that trial. Participants were not provided with the payoff matrix ahead of time, but rather they "discovered" the outcomes through feedback according to the actions taken by both players, and as identified in Table 1.

Table 1. *The payoff matrix table*

	Rock	Paper	Scissor
Rock	(2,2)	(1,3)	(4,0)
Paper	(3,1)	(2,2)	(1,3)
Scissor	(0,4)	(3,1)	(2,2)

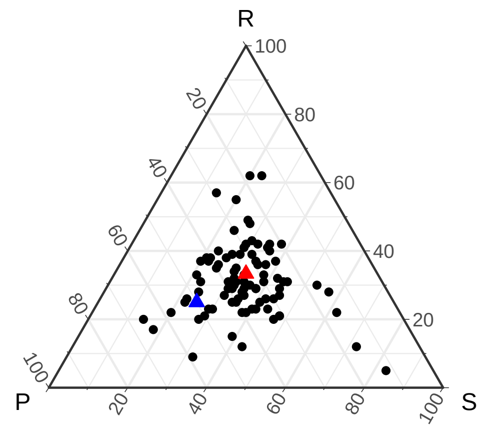


Figure 1. The ternary plot that describes individual choices. The red dot represents the random choices and the blue dot represents the Nash Equilibrium action.

Finally, all participants were instructed to complete a general short survey about their strategy. Participants received a payment based on the cumulative points at the end of the study (2 point equals to 1 cent), in addition to a base payment of 50 cents.

RESULTS

Of the 90 participants 22.2% started with rock, 40% with paper, and 37.7% with scissors. This result does not confirm a nonuniform preference for rock as observed in past studies (Eyler et al., 2009). On average, over 100 trials we found 31.8% selections of rock, 34.8% paper, 33.3% scissors, respectively.

To evaluate the choices at the individual level, we calculated the proportions of each individual's choices over 100 trials, and they are illustrated in a ternary plot of Figure 1. The figure also locates the Nash (Blue dot) and the Random (Red triangle) for reference. Figure 1 also indicates a large variability in individual strategies.

To identify whether individual participants played consistent with Nash (i.e., 1/2, 1/4, 1/4) or to Random (i.e. 1/3, 1/3, 1/3), we calculated the Euclidean distance from the average of each participant's choice proportions over 100 trials (represented by each point in Figure 1) to the Nash strategy as follows:

$$D_{Nash} = \sqrt{(p_{paper} - 0.5)^2 + (p_{rock} - 0.25)^2 + (p_{scissor} - 0.25)^2)}$$

We also calculated the distance to the Random strategy as follows:

$$D_{Random} = \sqrt{(p_{paper} - \frac{1}{3})^2 + (p_{rock} - \frac{1}{3})^2 + (p_{scissor} - \frac{1}{3})^2)}$$

.

A distance closer to 0 would provide evidence of the similarity of participants' strategies to the Nash or Random strategies. A one-sample t-test of the distance between players' strategies and the Nash strategy ($M_{D_{Nash}} = 0.21, SD_{D_{Nash}} = 0.12$) was significantly different from 0 ($t = 18.34, p < .001$), suggesting that participants did not play in agreement to Nash. Also, although participants' strategies were closer to Random ($M_{D_{Random}} = 0.14, SD_{D_{Random}} = 0.11$), the test also indicated that this distance was significantly different from 0 ($t = 14.03, p < .001$), providing evidence that participants did not play in agreement to the Random strategy.

An analysis of the dynamics of the proportions of R,P,S actions over the trials indicated that participants did not learn to play randomly or in agreement to Nash over the course of the 100 trials, despite the fact that they were given explicit feedback regarding their actions and payoffs and those of the other player.

Win-Stay/Lose-Shift

Here we test for evidence of a heuristic that is well-documented in the literature: "win-stay/lose-shift" (WSLS; Dyson et al., 2016; Forder & Dyson, 2016; Wang et al., 2014).

All payoffs are positive and thus the losses are relative in our PRS game. The lowest payoff is 0 (scissors are crashed by rock) and the highest payoff is 4 (when rock crashes scissors). In agreement with previous studies Dyson et al. (2016), we annotate two different cyclic directions of the *shift*: *upgrade* to refer to the subsequent $t + 1$th choice that beats the previous tth choice (e.g. rock-paper) and *downgrade* to refer to the subsequent $t + 1$th choice that is beaten by the previous tth choice (e.g. paper-rock), based on each player's self choice. Therefore, each outcome can be followed by three different strategies: stay, upgrade, and downgrade. Figure 2 illustrates the upgrade and downgrade strategies. Thus, a choice in a trial can be followed by a stay (e.g., paper beats rock, select paper again), upgrade (e.g., paper beats rock, shift to scissors), or downgrade (e.g., paper beats rock, shift to rock).

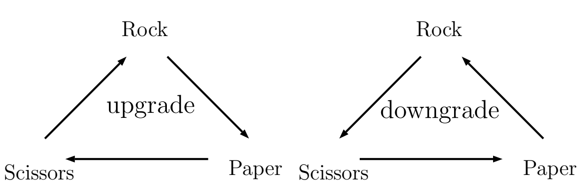


Figure 2. The coded cyclic strategy of *upgrade* and *downgrade* that represent consecutive choices.

Given the example that "paper beats rock", the WSLS strategy made a specific assumption that the paper player *stays* with paper while the rock player *shifts* to paper (upgrade) or scissor (downgrade). Table 2 shows the empirical transition probabilities averaged across participants. However, these probabilities indicate that the average group behavior did not reflect the WSLS. Again, this might be due to the large individual strategy variability that we observed in Figure 1.

As Dyson et al. (2016) noted that participants who upgraded were more likely to continue upgrading and participants who downgraded were more likely to continue downgrading, it is possible that some participants decide which strategy to use while ignoring whether they win or lose. Intuitively, even though RPS is a dynamic two player game, some participants may only focus on their own selections without considering others. To test if participants' decisions of strategy are independent from the outcome, we performed chi-square test on each individual's choices to evaluate whether there was independence between the strategy and the outcome.

The individual chi-square test results indicated that 36 out of 90 participants selected strategy in each trial depending on each trial's outcome (*dependence* group), suggesting the other 54 participants' strategy was independent from the trial's outcome (*independence* group). Using these two categories we calculated again the transition probabilities for the two groups. However, again, none of the two groups showed a behavior that matched the WSLS heuristic.

Win-Stay/Lose-Shift based on payoffs

Given that the "loses" are relative in our special non zero-sum PRS design, it is possible that participants interpret "loses" differently (e.g., a 0 is worse than a 1), and they make effort to obtain 4 points with rock and avoid 0 points with scissors. To explore this possibility, we looked at the strategies (stay, upgrade, downgrade) at trial $t + 1$ after each particular payoff was experienced at trial t to examine whether WSLS only applies to a more specific payoff pattern.

Table 3 gives the empirical transition probability for each payoff averaged across all 90 participants. The empirical transition probabilities are very similar across the various outcomes. We performed non-parametric two sample Kolmogorov-Smirnov (KS) on each two proportions of strategy within the same outcome (e.g. stay followed 0 and 1 and followed 3 and 4). The KS test tells whether samples are from the same distribution and the results indicate that only the proportion of downgrade followed by 0 points and 1 points were statistical significant different ($D = .24, p = .01$). The payoff of 0 is special as that is the only payoff combination where participants earn no points (and not money). Specifically, it indicated that losing with scissors when facing rock stimulated more subsequent selections of rock than other losing scenarios. This pattern was more prominent in the *independence* group

Table 2. *The empirical transition probability for outcome*

Previous Outcome	Stay	Shift	
		Upgrade	Downgrade
Lose	0.33	0.33	0.33
Tie	0.34	0.32	0.34
Win	0.33	0.33	0.34

than the *dependence* groups. However those comparisons were not statistically significant based on the performed KS test.

To summarise, we did not find consistent evidence with the previous literature suggesting that participants follow WSLS in the PRS. Instead, we found some evidence for the way participants interpret "win" and "lose" based on the payoffs. Participants are particularly sensitive to the loss outcome of 0 as scissor player tends to switch to rock after losing with 0 points while the rock players get 4 points. Additionally, we found the evidence for two categories of participants: a group of players whose choices of strategy were independent from the outcomes and another group that acted independently from the outcomes. We continue to explore this individual variability in the following section.

Cluster Analyses

We used hierarchical clustering within each of the *dependence* and *independence* groups. This methods allows us to systematically capture the similarities and dissimilarities in the strategies among participants. We employ a basic Ward agglomerative clustering method in which the similar clusters are merged based on their proximity until all clusters form a single cluster. In practice, we check the last one or two operations before merging into the final single cluster to cut cluster trees and decide the number of clusters representing the proximate group behavior.

Independence group clustering Figure 3 shows the average proportion of strategy (stay, downgrade, upgrade) for the *independence* group clustered into three sub-groups. Contrary to the previous study that participants keep "upgrading" or "downgrading" in a cyclic manner (e.g. paper-rock-scissor or scissor-rock-paper; Dyson et al., 2016), we found that participants' proportions of strategies were comparatively located in the middle of the ternary plot with some participants tending to stay more (Blue cluster), and some tend to switch more and upgrade/downgrade (Grey cluster). The middle Yellow group choose a slightly more balanced strategy combination. Our data observation indicates that most of the participants in the *independence* group adopted a mixed strategy thus increasing the chances to appear unpredictable to the other player in the game. For example, a participant can move from paper to rock (upgrade) and then back to paper (downgrade) again. Therefore, the participants did not make choices that were dependent on winning or losing, they produced more "complex" sequential choices by changing between upgrading or downgrading instead.

Table 3. *The empirical transition probability for payoffs*

Previous Payoff		Stay	Shift	
			Upgrade	Downgrade
Lose	0	0.31	0.32	0.37
	1	0.33	0.36	0.32
Tie	2	0.34	0.32	0.34
Win	3	0.33	0.33	0.34
	4	0.30	0.35	0.35

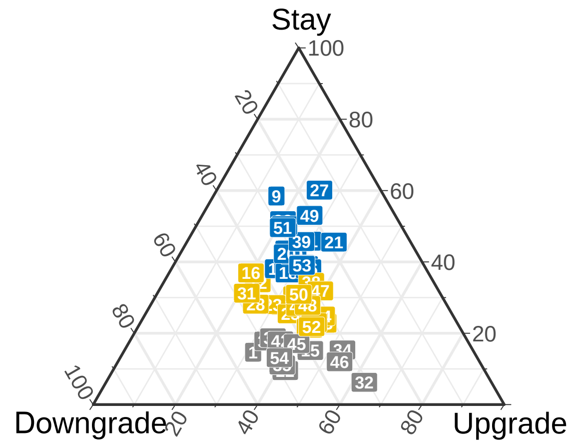
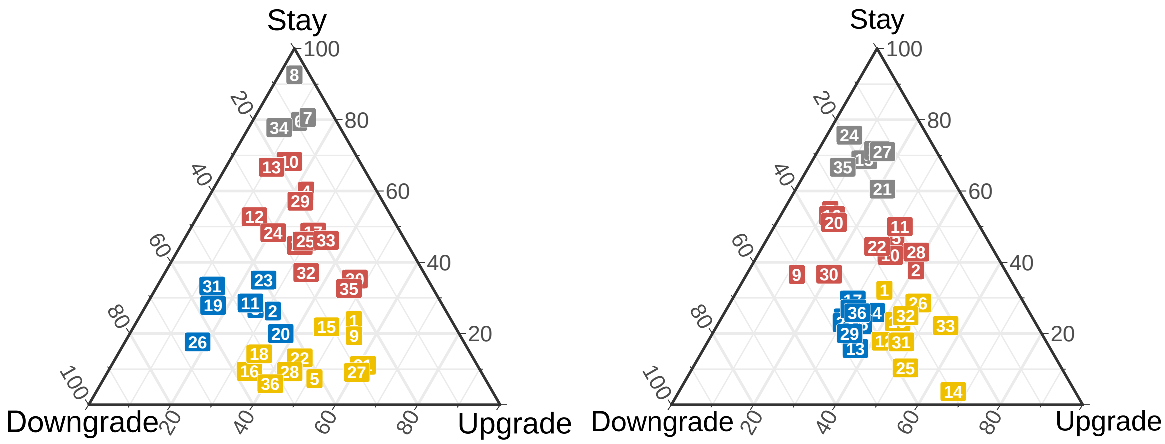


Figure 3. The ternary figure describes participants proportion of different strategies in the *independence* group

Dependence group clustering Since the *dependence* group showed evidence that the selected strategy was dependent on the outcomes, we performed two separate hierarchical clustering for the proportion of strategies for wins (Figure 4(a)) and loses (Figure 4(b)) for the 36 individuals in the *dependence* group.

Figure 4(a) suggests that some participants resemble win-stay (i.e., Red and Grey group), as they "stay" more after a win, but others resemble win-shift, as the Yellow cluster appears to upgrade after a win and the Blue cluster appears to downgrade after a win. Similarly, Figure 4(b) shows a diversity of behaviors after a lose. Participants show behavior that resembles lose-shift (i.e., Blue and Yellow groups); the Yellow cluster slightly upgrades more while the Blue group downgrades more. But some participants also reflect a lose-stay behavior (i.e., the Grey and Red groups). The two different clusters of wins and loses indicate that from a data driven point of view, there are heterogeneous individual variances in the chosen strategy. It is clear that not all participants chose "win-stay" and "lose-shift", which also supports our observation in the previous section that there was no generalized WSLS heuristic across the whole group. But these results also show the existence of "win-shift" and "lose-stay" strategies among individual participants.

Taking the advantage of being able to identify each participant's strategies between two different outcomes, we chained the two dendrograms together based on the participants' label identity. Figure 5 helps track how each individual participant falls into wins and loses clusters. The patterns of the connections of individuals across the dendrograms observes an interesting cross pattern. Participants that follow a win-stay strategy (Grey and Red groups in the left panel) often are those that follow a lose-shift strategy (Blue and Yellow groups in the right panel). Similarly, those that follow a win-shift strategy (Yellow and Blue groups in the left panel) often are those that also follow a lose-stay strategy (Red and Grey groups in the right panel). These results suggest two prominent strategies and two



(a) Win

(b) Lose

Figure 4. The ternary figure describes participants proportion of different strategies in the *dependence* group for wins (left) and loses (right).

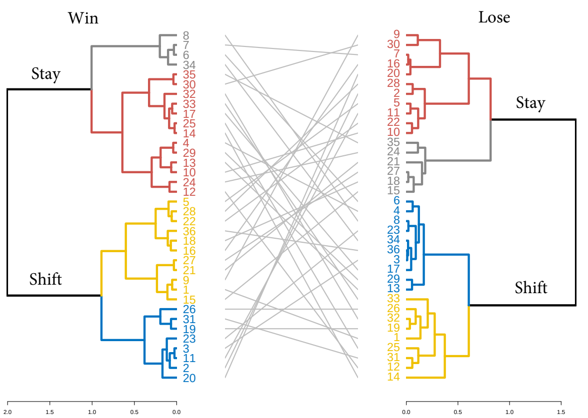


Figure 5. The two dendrograms of win and lose connected by the identical participant. The height of dendogram represents the proximal distrance between clusters (data points).

types of individuals; those that follow a "win-stay" and "lose-shift" strategies, and others that follow "win-shift" and "lose-stay" strategies.

CONCLUSION

In the current study, we designed a PRS game and collected data in an online 2-player experiment. Our results conclude that participants did not play the game optimally (i.e., in agreement to Nash) nor randomly. We also found that humans' behavior are very heterogeneous, and cannot conclusively be described by a commonly claimed heuristic of WSLS.

Instead, we found evidence for different "types" of individuals: one's strategy selection is dependent on outcomes (*dependence* group) and another one revealed independent strategy selection (*independence* group). The cluster analyses further showed that the two groups performed differently as how to increase the chances of being unpredictable. The *independence* group altered their cyclic methods. The *dependence* group, on the other hand, showed individuals that preferred the classical "win-stay" and "lose-shift" heuristics; but also individuals that preferred "win-shift" and "lose-stay" heuristics. A plausible explanation of this apparently "irrational" behavior is that people might follow a more sophisticated reasoning, in the sense that they believe their opponent will expect them to follow the WSLS, and as a best response they decide to do the opposite. More research is needed to follow on this finding. We believe our research contributes to understanding the dynamics of players strategies in an adversarial setting.

ACKNOWLEDGEMENT

This research was sponsored by the Army Research Office and accomplished under Grant Number W911NF-17-1-0431 to Cleotilde Gonzalez. We want to thank previous collaborators: Michael Yu, Pegah Fekhari, and Hui Sun for some initial analyses performed on this data (not reported here), and research assistants at the Dynamic Decision Making Laboratory in supporting data collection for this study.

REFERENCES

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.

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*, *6*, 20479.

Eyler, D., Shalla, Z., Doumaux, A., & McDevitt, T. (2009). Winning at rock-paper-scissors. *The College Mathematics Journal*, *40*(2), 125–128.

Fisher, L. (2008). *Rock, paper, scissors: game theory in everyday life*. Basic Books.

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*, *6*, 33809.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*(3), 267–272.

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*, 8817.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129–138.

Thorndike, E. L. (1911). *Individuality*. Houghton, Mifflin.

Wang, Z., Xu, B., & Zhou, H.-J. (2014). Social cycling and conditional responses in the rock-paper-scissors game. *Scientific reports*, *4*, 5830. doi: 10.1038/srep05830

West, R. L., & Lebiere, C. (2001). Simple games as dynamic, coupled systems: Randomness and other emergent properties. *Cognitive Systems Research*, *1*(4), 221–239.