Flexibility of reinforcement biases and reaction times in competitive zero-sum games

Dissertation submitted to the
University of Sussex for the degree of
Doctor of Philosophy

Jukka Sundvall

September 2019

ACKNOWLEDGEMENTS

Writing a sequence of thanks for people who have helped me in one way or another feels strange in a situation where I am not even sure my work is actually finished. This thesis will be submitted, examined, I will be questioned, and then, maybe, it will all be done, unless it is deemed that I need to revise. By that point, I will have moved away from this country and started another job. But maybe that just reflects on how science, or really any piece of creative work, is written in general: by the time it's out, you are no longer there.

The last three years have been frustrating and full of setbacks, and I am honestly surprised to be here. First, I would like to thank my supervisor Dr. Benjamin Dyson for helping me get through all of this. He has been amazingly supportive and surprisingly caring – amidst horror stories of extremely cold or ruthlessly demanding supervisors, especially as I took a big leap into the dark by moving into another country to undertake this PhD, Dr. Dyson gave me a soft landing, and has been amazingly understanding of me even during darker times. Dr. Dyson was willing to actually debate me on the minutiae of statistical analyses and inferences, and I feel this has been extremely productive in shaping the end product. He also helped me get used to Brighton and the UK in general, and has both been able to push me to work harder and to rest when I was pushing myself too hard. Him having been willing to continue supervision even after moving away from the UK in the middle of my project and having weekly chats with me just goes to show how dedicated he has been. I would also like to thank my assessor-turned-official-supervisor Dr. Daniel Campbell-Meiklejohn for providing needed feedback, ideas on where to go during the halfway point of my experimental work, and helping keep things running here at Sussex. Prof. Thomas Ormerod, who took the mantle of assessor as Dr. Dyson moved away, has also helped me keep my head straight in the last year of the PhD, and given me ideas for an academic or non-academic future. I would also like to thank the School of Psychology and the Osk. Huttunen Foundation for funding my work.

There are too many people in my everyday life in the UK to thank, but I will try and make it comprehensive. In no particular order, I would like to thank my office mates Jolyon Miles-Wilson and Petra Marcotti for providing company, de facto group therapy sessions, and good excuses for having pizza; my housemate and partner in unpublishable humour, Petar Raykov, for general fun times and care-free living under the same roof; Jo Cutler, Jessica Lunn and Jenny Morris for peer support and ideas during the last months and actually bothering to take part in the writing sessions I wanted to organize; Bojidar Boiadjiev for dragging me half-willing to try bouldering which turned out to be the best idea of the last three years; Roshni Kurien for making my first year in a strange country much more fun and social in an introvert-suitable way; James Livermore, Lina Skora, Bence Palfi, Mateo Leganes Fonteneau, Lucy Somers, Nora Andermane, Reny Baykova and probably several I forget for interesting discussions and general good times.

Back home in Finland, I would like to thank Dr. Michael Laakasuo, who was a key factor in me landing this post in the first place. Dr. Laakasuo deserves special thanks for having been able to keep in touch despite both of us being hard to reach and reclusive. I would like to thank my parents for support and a door that was always open for me when I was back in Finland, and constantly helping me with practicalities. I would like to thank my brother Tuomo for taking the time to visit me during my final crunch: you made it much more bearable and gave me a needed break. To the friends who have bothered to stay in personal contact – Julia Kylmälä, Anton Laihi, Anton Kunnari, Arto Kekkonen, Martti Karvonen, and others I surely forgot again, thank you for sticking around. I would also like to give special thanks to Matias Heikkilä, Matti Keränen, and Joonas Hautaniemi, my people from the woods, for providing perspectives from outside my bubble and keeping me sane. I might still be here without you, but I would not be as well.

Finally, I would like to thank Vilja Helminen. Thank you for keeping my head above the water when I myself could not. Thank you for being there. Thank you, love. I would certainly not be here without you.

STATEMENT

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:………………………………………

Jukka Sundvall

University of Sussex

Jukka Sundvall

Doctor of Philosophy

Flexibility of reinforcement biases and reaction times in competitive zero-sum games

SUMMARY

In competitive zero-sum games with mixed equilibria, two rational players should make each of their game choices randomly, with no contingencies between their choices. However, people often deviate from this equilibrium by following a reinforcement heuristic of repeating moves that won on the previous round *(win-stay)* and avoiding the repetition of moves that did not win *(lose-shift)*. In this thesis, I examine the flexibility of these reinforcement biases, and the speed of decision-making: under what circumstances do people make biased choices, and under what circumstances do people choose quickly or stop to deliberate? In Chapter 1, I review the current state of knowledge on how well people can produce or detect randomness, how reinforcement biases influence decision-making, and how processing speeds might differ between different game situations. In Chapters 2 and 3 I present four experiments where I examined performance in the games Rock, Paper & Scissors (RPS; Chapter 2, Experiments 1 and 2) and Matching Pennies (MP; Chapter 3, Experiments 3 and 4). Surprisingly, I found no reinforcement biases in RPS, but consistent reinforcement biases in MP. Additionally, participants made slower decisions after losses when they succeeded in the game due to finding an appropriate strategy to exploit an opponent's pattern (Chapter 2), but not when they succeeded no matter what they did (Chapter 3). In Chapter 4, I present two experiments (Experiments 5 and 6) directly comparing performance in RPS and MP, designed to replicate the findings from Chapters 2 and 3, and examine why the previous studies only found reinforcement

biases in MP. The results of these two last experiments suggest that reinforcement biases differ between RPS and MP due to different cognitive demands, and that there is considerable variability in reinforcement biases both between individuals and between the two types of bias. In Chapter 5, I discuss the contributions of the findings on the wider literature on bias and randomness detection, the generality of the reinforcement biases, and present some suggestions for future studies.

CONTRIBUTIONS

The thesis conforms to an article format in which the empirical chapters consist of pieces written in a style that is appropriate for publication in peer-reviewed journals in the field. The first chapter is a review of the relevant literature setting out research questions, and the final chapter is a general discussion of the results reported in this thesis in the context of the wider literature. I am the main author of the chapters that form this thesis and take responsibility for the analysis and write-up of the research. The experiments presented in the empirical chapters were designed jointly with my supervisor Dr. Benjamin Dyson, who offered his expertise and advice. Dr. Dyson took the lead in the design of Experiments 1 and 2 with input from myself on hypotheses, appropriate choice patterns for the computer opponent and questionnaires. Experiments 3 and 4 were designed jointly with equal contributions from myself and Dr. Dyson. Experiments 5 and 6 were designed primarily by me, with design input from Dr. Dyson. I also took lead in how the statistical analyses were structured starting from Chapter 3 and programmed the experiments (Experiments 1-3 in Presentation based on templates provided by Dr. Dyson, Experiments 4-6 in MatLab independently) except for the working memory and executive control tasks used in Chapter 2.

The reaction time data from Chapter 1 (Experiments 1 and 2), separate from the other results, have been published in

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, 1483-1487.

(As Experiments 2 and 3 in the article, respectively.)

CONTENTS

LIST OF FIGURES AND TABLES

Table 5.1: The presence of overall (group-level) win-stay and lose-shift biases in the six experiments in the random/unexploitable conditions

CHAPTER 1: General Introduction

1.1 Are People Bad at Being Random?

Pattern recognition is crucial to survival. In a competitive environment, an individual must be able to notice the frequencies of events as well as possible contingencies between events to be able to maximize their rewards and minimize losses. That is, individuals must aim to exploit their opponents, and avoid being exploited. In order to avoid being exploited by an opponent, individuals should also be able to act unpredictably. That is, they should be able to recognize a situation where there is no pattern to the actions of the opponent and themselves avoid producing a pattern that the opponent could exploit. Falsely interpreting a random sequence of events as a pattern may lead to an attempt of exploiting that pattern, which, in turn, means that the player who drew the false inference would behave in a systematic way and open themselves up to exploitation. These dynamics can be formalized in two-player zero-sum games with a mixed strategy equilibrium (Nash, 1950) – that is, if both players are choosing each alternative randomly, with equal probability, neither have an incentive to change their strategy. However, this incentive changes when either player has any kind of bias or pattern to their strategy. Even outside competitive situations, at least some skill in deducing whether events in the environment are the result of some contingencies or simply random seems necessary for an organism. However, humans are rarely capable of fully "rational" behaviour: instead, we are boundedly rational (Arthur, 1994) and often rely on heuristics that may not lead to strictly optimal results (Gigerenzer et al., 1996).

Several early studies on subjective randomness indicated that people have problems in both producing and recognizing randomness (see Bar-Hillel & Wagenaar, 1991; Wagenaar, 1991, for reviews). In sum, when explicitly asked to produce "random" strings of different elements, participants tend to produce too many alternations and avoid runs of a single element, even though a random process can and does produce such runs. Similarly, when asked to recognize “randomness” in different kinds of series presented to the participants, they often indicate strings with higher-than-chance probabilities of alternation as the most random (see Section 1.3). However, when comparisons were possible, there was little correlation in performance between randomness production tasks and randomness recognition tasks (Bar-Hillel & Wagenaar, 1991), suggesting that the biases in the two domains stem from different factors. In any case, at least for recognition tasks, it seemed that people had an idea of a prototypically “random” sequence as having few runs and more alternation than what an actual random process would produce. Bar-Hiller and Wagenaar (1991) proposed that the biased idea of what a “truly random” sequence would look like is in itself the reason people find it difficult to learn away from the bias. That is, if a person believes that random sequences do not have streaks in them, exposing that person to random sequences with streaks may be useless as they would interpret these sequences as non-random. Thus, subjectively, it would not be an experience of a random sequence that would break the person’s expectations and aid in learning about what true randomness would look like.

Ayton, Hunt and Wright (1989) and Rapoport and Budescu (Rapoport & Budescu, 1992) noted several issues with the usual randomness production and recognition paradigms. First, studies of subjective randomness often contained explicit instructions regarding what randomness should look like, e.g. that a random string of letters would not likely contain recognizable English words or alphabetical sequences such as “ABC” (Baddeley, 1966). Instructions such as this may have been taken by participants as an explicit warning against producing patterns and thus lead to an alternation bias (Ayton et al, 1989; Rapoport & Budescu, 1992). Second, the definition of a "random sequence" is itself questionable, as randomness is not strictly speaking a property of the sequence itself but of the process generating it, and only an infinite sequence would contain enough information about the stochasticity of the process. The randomness of a finite sequence could be tested in several different ways, none of which can overcome the aforementioned problem of the test requiring infinite sequences, which are impossible for people to either produce or observe (Ayton et al., 1989; Rapoport & Budescu, 1992). Third, it is difficult to incentivize participants to produce or detect the kinds of sequences that the experimenters consider sufficiently random without biasing the participants (Rapoport & Budescu, 1992). Fourth, production tasks especially are quite artificial, as there is seldom if ever a need for an individual to *produce* random sequences (at least without feedback; see Section 1.2), even if the *judgment* of whether an event was caused by chance alone or not may be common (Rapoport & Budescu, 1992). Thus, it may simply be that participants of randomness production experiments have failed simply because the way randomness production was operationalized in these studies is a skill that people rarely if ever practice. This suggestion was supported by an earlier study by Neuringer (1986), who had participants produce 60 sequences of 100 binary choices for a total of 6000 choices, either with or without feedback on the statistical properties of the sequences the participants had produced. Participants in the group who received feedback after the production of each sequence eventually learned to behave significantly more random-like than participants who did not receive any feedback.

Rapoport and Budescu (1992) proposed two-player (or *dyadic*) zero-sum games with a mixed strategy equilibrium as a better way of assessing people's skills in producing random behaviour. As mentioned above, the mixed strategy equilibrium essentially leads to a situation where two rational players should end up playing the game randomly. Thus, in a game paradigm, there is no risk of instructional biases (players can simply be told to play the game and try and win as much as they can), the participants can be incentivized more easily than in pure randomness production tasks, and the task is less artificial (there are plausible real-life situations where two parties in competition need to be able to act unpredictably). The mathematical issue of measuring randomness still applies to the zero-sum game paradigm, and thus Rapoport and Budescu (1992) simply measured different indices of randomness as in earlier studies, by comparing participants' behaviour to that of the expected behaviour of a Bernoulli process (i.e. a string of binary outcomes each with an independent probability of 50%).

In their study, Rapoport & Budescu (1992) compared the randomness of participants playing 150 rounds of a two-choice zero-sum competitive game (Matching Pennies) against one another to two control conditions: one where the participant was simply asked to produce a sequence that simulated the toss of a fair coin for 150 times, and one where the participant was asked to produce a binary sequence of 150 iterations that would then later be used as their moves in a series of rounds of the zero-sum game. Note that in each of these tasks the ideal results would look similar, but only the game condition provided trial-to-trial feedback (information of the opponent's moves and outcomes of rounds). Assessing a number of statistical markers of random-like behaviour such as the distribution of choice types, the number of runs, the distribution of response patterns of different lengths (m-tuples), and the conditional probabilities of a response as a function of a prior response or responses, Rapoport and Budescu (1992) showed that participants in the competitive game condition consistently had either approximated randomness better than participants in the other conditions, or behaved equally non-randomly. The results thus supported the notion that people's difficulties in producing random sequences in prior experiments was significantly affected by the task design itself, and that the functional production of randomness (randomness in order to avoid exploitation in a competitive game) comes more naturally.

What about zero-sum games allows people to avoid deviations from randomness better than pure production tasks? Wagenaar (1991) suggested that a random process requires at least i) a fixed set of alternatives ii) a "memoryless" selection procedure and iii) a selection procedure with no preference for any of the alternatives (see also Bar-Hillel et al., 2014; Mehta et al., 1994). Wagenaar (1991) argued that people's errors in randomness production tasks are due to having a memory that can not be simply turned off, and due to having preferences. That is, people might have initial preferences for certain kinds of items or patterns in their production of random sequences, and memory of their previous choices biases them towards avoiding items they chose previously as they try to achieve randomness (alternation bias). In response to Wagenaar (1991), Rapoport and Budescu (1992) argued that pitting players against one another interferes with the process that leads to deviations from randomness in pure production tasks. Specifically, they argued that game feedback (wins and losses) removes any preference the players may have, and that having to track the opponent's moves interferes with the players' memory of their own past moves (thus leading to a reduction in alternation bias). However, this claim is undermined by very clear and robust associations between prior outcome and subsequent choice as expressed by reinforcement learning principles.

1.2 Reinforcement Biases

In their simplest form, reinforcement learning principles can be viewed as a heuristic of repeating actions with positive outcomes ("*win-stay*") and avoiding the repetition of actions with negative outcomes ("*lose-shift*"). This is essentially the (simplified) logic of behaviourist principles of learning (c.f. Thorndike, 1911, *law of effect*). In the context of games with mixed equilibria, *win-stay* and *lose-shift* represent examples of predictable performance that is non-random and hence exploitable. As such they represent suboptimal performance if an agent acts according to them regardless of the type of learning or game task they are faced with (see e.g. Achtziger & Alós-Ferrer, 2013; Achtziger et al., 2015; Alós-Ferrer & Ritschel, 2018; Scheibehenne et al., 2011; Wilke & Barrett, 2009; Wilke et al., 2014). Thus, outcome-dependent deviations from randomness are important to examine when measuring subjective randomness. Rapoport & Budescu's (1992) analysis of how often a player repeated the opponent's move from the previous round is essentially an analysis of the sum score of *win-stay* and *lose-shift* (hence *stayshift*) choices in their game task. The reason the two analyses are the same is because in Matching Pennies, a participant repeating the opponent's move is either repeating a move that they matched in the previous round (i.e. won) and thus also repeating their own move, or repeating a move that they did not match in the previous round (i.e. lost) and thus also shifting from their own previous move (see Dyson, 2019, on isomorphism between different dependencies in zero-sum games). In this analysis, they found only 13 out of the 66 participants in their dyadic condition had a significant bias towards repeating the opponent's move. However, crucially, the analysis lacks a comparison group, as there was only one experimental group that played a game with trial-by-trial feedback. The feedback this group received was essentially created by the two players themselves as a dynamic, coupled system (see West & Lebiere, 2001). Thus, an aspect missing from Rapoport and Budescu's (1992) examination is a test of the role of *different types* of feedback dynamics on the potentially present reinforcement effects.

West and Lebiere (2001) provided a simulated model of a Rock, Paper & Scissors game with two players that attempt to exploit each other's predictable behaviour with a limited memory, and found it aligns to games played by actual people. By learning sequential dependencies and attempting to exploit them, the two players end up creating random-like behaviour. This is a plausible model of what may have happened with the players in Rapoport and Budescu's (1992) experiment: it is likely that both human players tried to exploit any systematic behaviour they perceived on the other player's part. This would, in turn, lead each player to change any pattern they previously had expressed as they noticed the opponent exploiting it (i.e. receiving negative feedback). In the long run, this could lead to both players adopting different kinds of patterns for short runs, followed by a change in the pattern caused by the opponent trying to exploit the pattern, and the cycle beginning again. This is a very specific type of feedback for the players to receive, and the fact that most players did not exhibit a *stayshift* bias in this game environment does not automatically mean that they would avoid it in a zero-sum game with an opponent that acted differently. Thus, this opens up the question of what kind of a dynamic is needed between the two players for them to avoid deviations from randomness.

The role of the dynamic producing game feedback becomes apparent by looking at studies where that dynamism is removed. Such cases might be where the participants play zero-sum games against opponents that play completely randomly, with no learning or predictable reaction to the participant's responses. In these experiments, the game task still has the properties highlighted by Rapoport and Budescu (1992) as important in reducing deviations from randomness, namely feedback and tracking the opponent's actions. However, the task lacks an opponent that would punish a player for playing predictably, though it does not reward this predictability either. As a result, the expected number of wins and losses for the player is equal no matter how they play. In two studies of rhesus monkeys playing a Matching Pennies (Lee et al., 2004) or a Rock, Paper & Scissors game (Lee et al.,, 2005) against a randomly playing computer opponent, the monkeys exhibited a *stayshift* bias. For human players, Scheibehenne et al. (2011) found that participants playing a binary choice game on a simulated slot machine (resembling a Matching Pennies game) exhibited a significant *stayshift* bias. This result was replicated in a similar paradigm by Wilke et al. (2014), who found that both habitual gamblers and non-gambling controls misapplied the *stayshift* rule when playing on a machine with random outcomes. Dyson et al. (2016) and Forder and Dyson (2016) provided similar results of human players playing Rock, Paper & Scissors against a computer opponent that played the game randomly, though the results of both papers show only a significant *lose-shift* bias but not a significant *win-stay* bias.

Thus, it seems that people express a *stayshift* bias against an opponent that plays randomly. This suggests that a zero-sum game scenario in and of itself is not enough for people to discard this deviation from randomness. The fact that *stayshift* could *theoretically* be exploited does not reduce the bias. As such, *stayshift* could be considered an example of a *trivial bias* (see McKay & Efferson, 2010) in this context since the bias does not cause the players to lose any more than they would choosing any other strategy. The *stayshift* bias is then a bias only in the technical sense that it is not the theoretical optimal strategy, but neither does it lead to exploitation. An argument could be made that an agent would only change their strategy from an initially biased position if they are incentivized to do so. This incentive could be in the form of being exploited by an opponent that notices their bias (i.e. negative feedback), and/or if there is a pattern to the opponent’s play that could be exploited by adopting a different decision rule (i.e., positive feedback). McKay and Efferson (2010) posit that a bias is non-trivial only if people express it even in situations where it actively leads to decreased success. This leads to the next question of how easily players learn away from the bias when incentivized.

Whether players can avoid reinforcement-based biases seems to depend greatly on the type of game task and the dynamics of the feedback. The evidence of a *stayshift* bias reducing due to negative feedback in competitive zero-sum games is somewhat mixed. In the above-mentioned studies of rhesus monkeys (Lee et al., 2004; Lee et al., 2005), the *stayshift* bias was reduced once the monkeys' computer opponent started exploiting the monkeys' biases. In other words, the monkeys seemed to start with an initial bias that could be exploited, and only reduced this bias when it in fact was exploited. Similarly, in a study of reinforcement biases in human participants, Ivan et al. (2018) found that participants did not exhibit a significant stayshift bias when playing Matching Pennies against a computer opponent that punished the players for predictability. Given that the *stayshift* bias has been commonly observed in earlier studies where participants played against randomly playing opponents, it would seem that the bias is an initial default (or, the result of random feedback) that is discarded once it starts leading to more losses than chance.

On the other hand, Scheibehenne et al. (2011) reported data suggesting a more nuanced pattern in human players. In their experiment, Scheibehenne et al. (2011) had participants choose to bet on one of two slot machines, one of which had a completely random binary pattern while the other one had a pattern that could either be exploited by a *stayshift* rule (*win-stay* and *lose-shift*) or the opposite *shiftstay* rule (*win-shift* and *lose-stay*) to varying degrees. This exploitability was achieved by manipulating the autocorrelation (likelihood of repetition or "clumpiness") of the series the slot machine produced:

increased autocorrelation is exploitable by *stayshift*, whereas decreased autocorrelation is exploitable by *shiftstay*. Note that if a machine is exploitable by the *shiftstay* rule, applying the *stayshift* rule when playing will lead to increased negative feedback. When the non-random slot machine could be exploited by the *stayshift* rule, the participants' likelihood of choosing to bet on that machine increased throughout trials regardless of how strongly exploitable it was (i.e. how high the autocorrelation was). However, when the non-random slot machine could be exploited by the opposite *shiftstay* rule, participants only increased choices for this machine when it was highly exploitable (80% of the time, i.e. only 20% likelihood of repetitions), and for lower rates of exploitability they actually decreased their choices throughout trials. In sum, when participants played against the exploitable machines, their rate of optimal choices was much higher when the machine could be exploited with the *stayshift* rule than when it could be exploited with the *shiftstay* rule. Additionally, the rate of optimal choices for games played on the machine that could only be exploited with *shiftstay* rule, and thus punished *stayshift* responding, only increased above 50% of trials when the machine was highly exploitable (80% of the time). Lower levels of exploitability yielded chance or sub-chance optimal responding, suggesting that humans may need relatively strong negative feedback for their *stayshift* behaviour in order to learn away from it in the task. Wilke et al. (2014) provided a partial replication of these results with a similar design, finding that neither habitual gamblers or control participants increased their choice frequency on a negatively autocorrelated slot machine throughout several trials. This suggests that the bias is not completely trivial, as people seem to have difficulties learning away from it even when incentivized, or indeed noticing the incentive.

The *stayshift* bias can also lead to suboptimal play choices in game situations that do not require the players to randomize their choices. In a direct test of the effect of reinforcement, Achtziger et al. (2015) used a two-step binary choice task, where the optimal strategy after the first decision step, based on Bayesian updating, was either *shiftstay* or *stayshift.* For the case where the *shiftstay* rule was optimal, there was a direct conflict with the reinforcement heuristic of repeating winning moves. In the two-step task, the first decision is essentially random, but the outcome of that decision, whether a win or a loss, gives the participant information about the state of the world – that is, whether they are more likely to win by shifting or staying. The participants were made aware of all the probabilities of winning and losing in two different states of the world, but they could only deduce which state of the world they were in by making an initial choice. However, the knowledge of the outcome and the associated reward itself could override this information. In the situation where an initial loss suggested that *stay* was the subsequent optimal response, or an initial win informed them that they should *shift*, participants made the opposite, erroneous choice, on 58.3% of the trials on average. In the situation where the optimal responses aligned with *stayshift,* the error rate was only 9.29% on average. Additionally, participants were equally likely to make these errors regardless of whether the financial incentive for correct play was high or low. In a second experiment, Achtziger et al. (2015) removed the win and lose feedback and monetary reward from the initial decision, but still provided information that the participants could use the deduce the state of the world for the second decision. Here, the overall error rate fell significantly, from 58.3% of suboptimal *shiftstay* responding in the first experiment to 25.1% in the second experiment. Additionally, participants now made significantly more optimal choices when financial incentives were high than when they were low, suggesting that concrete rewards may only encourage rational choices when reinforcement is inhibited. These results broadly replicate earlier studies using the same paradigm (Achtziger & Alós-Ferrer, 2013;

Charness & Levin, 2005), and again suggest that the *stayshift* bias is not trivial (see McKay & Efferson, 2010), as people seem to apply it in a situation where the exact opposite behavioural rule would be optimal. Given that in most games in the wild or in the lab, feedback, reward and information about a trial are intertwined (see Losecaat Vermeer & Sanfey, 2015) and immediately obvious to the player, the reinforcement bias is likely ubiquitous.

More detailed work (e.g. Gruber & Thapa, 2016; Ivan et al., 2018) has also revealed potential differences between the flexibility of win-stay relative to lose-shift, with lose-shift being potentially a more automatic response that is normally suppressed in adults by executive functions. This fits together with Forder and Dyson’s (2016) findings of players exhibiting a lose-shift bias and also making faster decisions following losses than other outcome types: participants exhibited more cognitive control (Mackie et al., 2013) after wins than losses. In sum, a *win-stay*, *lose-shift* bias seems to be a kind of default decision rule in zero-sum games and other kinds of decision tasks for both humans and some animals. While human players may be able to learn away from the bias when incentivized, they may need very frequent negative feedback for their biased choices or an otherwise strong incentive to do so. Interestingly, people seem to be better at avoiding this reinforcement-based deviation from randomness when they are paired with opponents that will punish any kind of non-random play in a zero-sum game than when their opponent or the game simply only punishes the reinforcement bias. The bias does not seem to be a trivial bias, as human players seem unable to decrease the bias to the same degree that they can increase it when incentivized to do so, and less able to notice situations where decreasing the bias is optimal than situations where they should increase it. That is, there is some inflexibility in the *stayshift* bias, which may stem from it being an automatic process (or a "System 1" process; see Evans, 2003, 2008, for reviews), but there are also some differences between the two types of bias in terms of flexibility.

1.3 Link to Hot Hand & Gambler's Fallacy

The *stayshift* bias may in more general terms be considered an example of positive recency, that is, the tendency to expect events of one type to follow each other in a sequence. In binary choice tasks, a bias towards staying after wins implies an expectation of the same choice-outcome pair (or the same game choice from an opponent), and a bias towards shifting after losses implies an expectation that the choice that would have yielded a win previously will yield a win now (or, again, the same game choice from an opponent). In tasks with more than two choice options, a bias towards staying after wins is equivalent to the bias in binary choice tasks, and a bias towards shifting after losses implies at least an expectation that the losing choice is unlikely to yield a win on the next round. A bias towards shifting after losses specifically to the choice that would have yielded a win in the previous round (myopic best reply or Cournot's best response; see Alós-Ferrer & Ritschel, 2018; Dyson, 2019) would imply positive recency more clearly, and there is some evidence of games with more than two options that players are biased in precisely this way (Dyson et al., 2016; Alós-Ferrer & Ritschel, 2018). The positive recency effect is commonly called the Hot Hand Fallacy, as in expecting a basketball player who has scored being more likely to score again i.e. having a "hot hand" (see Gilovich et al., 1985; see also Miller & Sanjurjo, 2018, suggesting the belief may not be a fallacy in the context of actual basketball games). The opposite, negative recency, is the tendency to expect that an outcome is less likely to be followed by another outcome of the same type. This bias is commonly called the Gambler's Fallacy, as in a gambler expecting that after a run of the roulette ball landing on red, it should more likely land on black. The Gambler's Fallacy is observed in real-life gambling situations (Sundali & Croson, 2006), some prediction tasks (Barron & Leider, 2010), and in randomness production and detection tasks (alternation bias; see Section 1.1).

The existence of both the Hot Hand and Gambler's Fallacies raises the question of when people express these biases, as it would be logically impossible for one person to express both at the same time about the same outcome. Why, specifically, do most studies of different kinds of game tasks find biases consistent with the Hot Hand Fallacy so often? Ayton and Fischer (2004) noted that both fallacies have been previously been explained by an appeal to the representativeness heuristic (Gilovich et al., 1985). That is, the gambler's fallacy has been explained as a result of people assuming that the global properties of randomness should also be apparent locally, in every part of a random string (Kahneman & Tversky, 1972). This leads to an assumption that every part of a random sequence should contain equal numbers of all possible outcomes, leading people to judge runs of one outcome as unrepresentative of randomness, and thus expect runs to end. Likewise, the Hot Hand Fallacy has been explained as a result of people judging a streak as unrepresentative, and thus deciding that the underlying process is not actually random, therefore expecting the streak to continue. As Ayton and Fischer (2004) note, this latter explanation raises the question as to why people would not judge that a roulette wheel has become "hot" when it has produced a streak of one type of outcome. Similarly, one could ask why participants playing zero sum games do not form a *win-shift* bias, as a streak of one type of choice by the opponent could be considered unrepresentative of randomness and thus likely to be followed by a loss if the player repeats their choice.

Ayton and Fischer (2004) suggested that the relevant factor in whether exhibit positive or negative recency effects in interpreting or engaging with a sequence of events is their prior knowledge of the type of process they are dealing with. People may reasonably assume that the results of a roulette wheel should be random, and that the results of an athlete scoring points are not random. Thus, people may interpret a streak of successes presented to them as indicating a deviation from noise that is likely to *end* soon if they are told the string represents e.g. roulette outcomes because there is no intentionality in a roulette wheel. On the other hand, the same streak of successes presented to them would indicate a deviation from noise that is likely to *continue* if they are told the string represents e.g. the scoring outcomes of an athlete, because an athlete is an intentional agent, and an intentional agent succeeding several times in a row indicates skill.

Ayton and Fischer (2004) tested this prediction first with an experiment where participants were asked to forecast or gamble on outcomes on a simulated roulette wheel with binary outcomes. The participants also indicated their confidence of guessing the next outcome correctly on each round. Consistent with the notion that positive and negative recency effects depend on the assumptions people hold about the process, the participants exhibited negative recency in their guesses, but positive recency in their confidence. That is, they were more likely to predict that a streak of a certain outcome type on the roulette wheel to end, but they were more confident about a guess being correct when they had had a streak of correct guesses. In a second experiment, Ayton and Fischer (2004) asked participants to judge whether a sequence of binary outcomes presented to them was more likely to be caused by human skilled performance or chance (scoring in different sporting contexts or the outcomes of a roulette wheel, coin flip, or die throw). As predicted, when participants were presented with sequences with low alternation rates and more streaks, they were more likely to indicate that the sequence represented the results of human skilled performance. Sequences with high alternation rates were judged as more likely to represent chance processes. Similar results were reported by Burns and Corpus (2004), who found that participants were more likely to predict a streak in a sequence of 100 equally distributed binary outcomes to continue if the sequence was described as a competitive situation or a noncompetitive situation with a skill component relative to a random situation. In a field study of actual roulette bets, Sundali and Croson (2006) found that players who exhibited a Gambler's Fallacy about the actual outcomes of the roulette spin tended to also exhibit a Hot Hand Fallacy about their own betting success, suggesting the two biases are linked but relate to different properties of a task (the likelihood of an event vs. the likelihood of predicting that event correctly).

If negative recency is common in situations where people assume the process producing a sequence to be random, but positive recency is common in situations where people assume the process to be non-random, this implies that players exhibiting a positive recency effect (the *stayshift* bias) in zero sum games more commonly assume the opponent to be playing non-randomly. This is not surprising in and of itself, however, positive recency effects have also been found in game situations where players are not playing against another human or computer player but e.g. playing on a simulated slot machine or predicting natural events (see Scheibehenne et al., 2011; Wilke & Barrett, 2009; Wilke et al., 2014). Positive recency effects seem to depend not only on what participants think of the process behind a sequence but also whether they encounter the sequence trial-by-trial, as in studies where participants play games, or by receiving information about a whole sequence (Barron & Leider, 2010; Tyszka et al., 2008). This helps explain the ubiquity of positive recency, i.e. reinforcement biases, in game choices. Based on the literature, positive recency effects seem to be a common trait of human cognition when predicting what will happen in a sequence.

1.4 Aims

The aims of the experiments reported in this thesis were to examine the flexibility of reinforcement biases under different conditions, the predictors of biased behaviour, and the effect of positive and negative game outcomes on the impulsivity of game decisions. Experiments 1 and 2 (Chapter 2) addressed how well participants could play RPS against predictable computer opponents that required deviations from reinforcement-based decisions to beat, and the separate effects of different game outcomes on the likelihood of reinforcement biases. Experiments 3 and 4 (Chapter 3) addressed the effects of high and low win-rates and different trajectories of positive and negative outcomes (success slopes) on reinforcement biases, the players' confidence of their choices, and their reaction times in MP. Experiments 5 and 6 (Chapter 4) examined behaviour in both RPS and MP, under both conditions of random and predictable opponent behaviour, in order to uncover the reasons for differences in biases between the game types, as well as conflicting results in the general literature. The experiments aimed to shed light on the differences between *win-stay* and *lose-shift* responding in terms of flexibility, frequency, and impulsivity, and the factors that predict biased behaviour.

CHAPTER 2: Reinforcement Biases and Reaction Times in RPS

2.1 General Introduction

In Experiments 1 and 2, I used the game Rock, Paper & Scissors (RPS) to examine reactions to different game outcomes against both *exploitable* and *unexploitable* opponents, under different value conditions. My aim was to examine participants' ability to exploit predictability and to act in a non-exploitable way in a non-predictable situation, and to see whether differences in reward structure affected these abilities. Specifically, the exploitable choice patterns chosen for the opponent were biased towards specific shifts and led to a situation where the optimal strategy against the opponent was not in alignment with reinforcement or myopic best reply.

2.2 Experiment 1

2.2.1 Introduction.

2.2.1.1 Rock, paper, scissors and reinforcement. Assuming two rational players, the Nash Equilibrium (Nash, 1950) of RPS is achieved when both players use a mixed strategy, that is, when they make random choices with replacement from a flatly distributed set of options. This ensures that neither player has a recognizable pattern in their choices and no bias towards any of the choice options (e.g. overplaying rock). However, while RPS provides an environment where the production of random strings is considerably easier than in traditional non-game randomness production tasks (Rapoport & Budescu, 1992), the game structure simultaneously introduces the effects of trial-by-trial reinforcement into the task. The types of non-random choices people make in RPS seem to follow a similar pattern of reinforcement errors as seen in binary choice tasks (see e.g. Achtziger et al., 2015; Scheibehenne et al., 2011; Wilke et al, 2014). Previous studies of behaviour in RPS or similar three-choice games (Alós-Ferrer & Ritschel, 2018; Dyson et al., 2016; Forder & Dyson, 2016; Wang et al., 2014) have found that people tend to follow a *stayshift* heuristic, with shifts being more likely after both losses and draws. The shifting rule could thus be more accurately called "lose/draw-shift" in the context of RPS (see Lee et al., 2004 for a similar strategy in monkeys). While a draw in the game is not strictly speaking a penalty, a player trying to maximize wins will of course prefer to not draw.

It has been found that the magnitude of the *win-stay* bias, but not the *lose/draw-shift* bias, can also be altered by different reward values, suggestive of varying degrees of cognitive control as a function of outcome. Cognitive control can be broadly defined as information prioritization in goal-driven behaviour (Mackie et al., 2013) - i.e. the decision does not happen automatically. It is thought as arising in situations of uncertainty or conflicting information in a task, and manifesting as increased reaction times (see Botvinick et al., 2001, for a comprehensive look into the conflict monitoring account of cognitive control). Forder & Dyson (2016) compared game situations with either an emphasis on wins (+2 points for a win, -1 for a loss) or losses (+1 for a win, -2 for a loss) to a baseline condition (+1 for a win, -1 for a loss), when participants played against an *unexploitable* (mixed strategy) opponent. It was found that reaction times for choices made after wins were longer than those for choices made after losses or draws. Additionally, the amount of *lose-shift* behaviour was similar for all conditions, but *win-stay* behaviour was increased when the value of a win was higher. It is worth noting here that the difference in behaviour was observed by simply manipulating a numerical score with no financial or other tangible incentive, suggesting that possibly evolutionary behavioural biases like the *stayshift* heuristic can manifest even in situations with no reward or anything "real" for the player to compete for. In other words, these effects may manifest as response simply to information about wins and losses, not specifically wins and losses tied to rewards.

Together, the results indicate an inflexibility of the *lose/draw-shift* bias and a flexibility of the *win-stay* bias, with slower responses following wins. Note that the longer deliberation time did not necessarily lead to more ”rational” choices, as the increase in *win-stay* behaviour when the scoring for wins was higher did not improve performance against the randomly playing opponent.

Based on the results of Forder & Dyson (2016) it is not completely clear whether the differences in choice behaviour between the conditions were due to the asymmetry between the scores for wins and losses, or merely due to the higher value of wins: the win-heavy condition (+2 points for a win, -1 for a loss), where participants made significantly more *win-stay* decisions than in the other conditions, had the highest point value for wins. It is thus possible that the value of losses or the difference in value between losses and wins played no part in increasing *win-stay* behaviour, and that *win-stay* behaviour only increased as a function of the value of wins. Therefore, in Experiment 1 I decided to examine player behaviour in two different conditions where the penalty and reward were equal to each other: a *high value* condition with +3 for a win and -3 for a loss, and a *low value* or baseline condition with +1 for a win and -1 for a loss (as per Forder & Dyson, 2016). If players were incentivized to stay after a win simply due to a higher point value, I would expect to see more win-stay behaviour in the *high value* condition.

2.2.1.2 RPS and exploitation. It should be noted that for RPS the *stayshift* bias has been observed only in cases where the players play against a computer opponent using a mixed strategy (Dyson et al., 2016; Forder & Dyson, 2016) or in competition with another human unlikely to have a steadily *exploitable* pattern of play (Alós-Ferrer & Ritschel, 2018; Wang, Xu, & Zhou, 2014). What about situations where a player could reliably learn a winning strategy against the opponent? Studies showing frequent reinforcement errors even when an optimal strategy could be learned through trial and error (Achtziger et al., 2015; Scheibehenne et al., 2011; Wilke et al., 2014) differ from a typical RPS scenario. Firstly, the decision tasks in these studies are binary, in contrast to RPS, where the player always has three options. Secondly, the games are not two-player competitive games per se, even though the structure may otherwise be similar (i.e., clear definitions of outcomes, game played over several rounds, need for a strategy). Critically, the player and the game are not competing for points or for resources. Behaviour in a game that is explicitly framed as a zero-sum competition against another player may differ significantly from that in other types of games (see Bornstein et al., 2002; Goodie et al., 2012).

In a study examining participants' responses to different RPS strategies (Stöttinger et al., 2014) participants increased their rate of optimal responses to a frequency biased opponent (an opponent playing Rock more often than any other choice) quicker than to to an opponent whose moves were dependent on the player's previous move during a 100 trial training phase. Moreover, players had in general a higher proportion of optimal plays against the frequency-biased opponent than the player-dependent opponent averaged across the training phase. The authors did not report the patterns of the participants' responses, but it's notable that the participants fared worse against the player-dependent opponent than the frequency-biased opponent. The *stayshift* rule could be beneficial against frequency-biased opponents, as it naturally steers players to shift away from the moves that are least likely to win and towards repeating the move that wins most often. The player-dependent opponent strategy used during the training phase was a "one-ahead" strategy, where the opponent would choose an item that would have beaten the player's previous choice – similar to the *stayshift* rule, with shifts in specific directions after losses and draws and repetitions after wins. That is, both staying after wins and shifting after losses and draws are likely beneficial. Against such an opponent, a player playing according to the *stayshift* strategy would not have a similar edge as they would against a frequency-biased opponent: even if shifting after losses or draws may lead to a winning move, the player repeating their winning moves likely leads to losses. It may be that the worse performance against the player-dependent opponent in Stöttinger et al. (2014) stemmed from a reliance on the *stayshift* heuristic. One way to measure the flexibility or inflexibility of the *stayshift* heuristic, then, is to pit human players against a computer opponent that is optimized to play against such a biased player, and examine the participants' decisions in comparison with their behaviour against a random opponent.

2.2.1.3 Choosing a strategy. As mentioned above, the lose/draw-shift rule in RPS generally means *more likely shifting after a loss or a draw.* In the context of RPS, because there are three response, two types of shift behaviour are available across consecutive trials. I will call the two different forms of shifting in RPS *downgrading* and *upgrading,* respectively (after Dyson et al, 2016) - see Figure 2.1 for a schematic of the two forms of shifting. Downgrading is shifting to an item that would have lost to one's own previous choice, while upgrading is shifting to an item that would have won against one's own previous choice.

The *stayshift* tendency could be seen as a shortcut towards a myopic best reply strategy (Alós-Ferrer & Ritschel, 2018), where decisions are made with the assumption that the opponent will repeat their choice irrespective of the preceding outcome – i.e. choosing the option that would have won in the previous round. This would mean downgrading after losses, upgrading after draws, and staying after wins. The evidence for the myopic best reply strategy in three-choice games is mixed. In Dyson et al. (2016) and Forder and Dyson (2016), there is a general trend towards shifting after losses and draws, but no reliable trend towards shifting significantly more often to the directions dictated by myopic best reply. On the other hand, Alós-Ferrer & Ritschel (2018) found a significant trend towards myopic best reply shifts in a slightly modified RPS game with two different magnitudes of losses instead of the typical loss and draw. The common trend across these studies was that players followed reinforcement: shift responses in general were more likely after non-win outcomes than stay responses. To create a strategy that would be hard for participants to learn via reinforcement, in Experiment 1, I decided on a simple *self-downgrade* rule for the computer opponent. To formalize, the strategy can be explained in three rules (Dyson, 2019):

(1) IF O(n) = r THEN O(n+1) = s

(2) IF O(n) = s THEN O(n+1) = p

(3) IF O(n) = p THEN O(n+1) = r

where O = opponent's choice, n = number of trial, r = rock, p = paper, s = scissors.

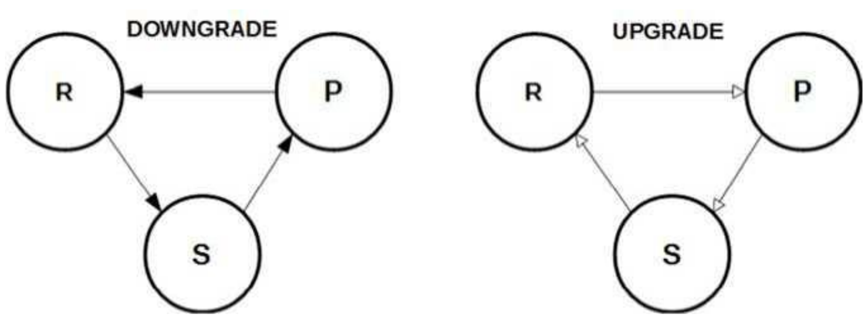

Figure 2.1. Schematic showing the cyclical nature of downgrading and upgrading in RPS. In *downgrading*, the shift is towards the item that would lose against the previous item; in *upgrading*, towards the item that would win against the previous item.

The optimal way of playing against the self-downgrade biased opponent is to downgrade after a win, upgrade after a loss, and stay after a draw. To formalize this for all item choices made by the player:

(4) IF S(n) = r AND (n) = W THEN S(n+1) = s *(downgrade)*

(5) IF S(n) = r AND (n) = L THEN S(n+1) = p *(upgrade)*

(6) IF S(n) = r AND (n) = D THEN S(n+1) = r *(stay)*

(7) IF S(n) = p AND (n) = W THEN S(n+1) = r *(downgrade)*

(8) IF S(n) = p AND (n) = L THEN S(n+1) = s *(upgrade)*

(9) IF S(n) = p AND (n) = D THEN S(n+1) = p *(stay)*

(10) IF S(n) = s AND (n) = W THEN S(n+1) = p *(downgrade)*

(11) IF S(n) = s AND (n) = L THEN S(n+1) = r *(upgrade)*

(12) IF S(n) = s AND (n) = D THEN S(n+1) = s *(stay)*

where additionally to the above, S = self i.e. the player's choice, W = Win, L = Lose, D = Draw. Note that this set of rules is in contrast to the usual *stayshift* heuristic for wins and draws, and in contrast to myopic best reply for all outcomes. This makes the strategy well suited for examining the malleability of reinforcement-based decision-making. If players are driven by reinforcement, they should make more errors after wins and draws compared to losses, as the optimal choices after wins (i.e. downgrade) and draws (i.e. stay) are misaligned with reinforcement, unlike the optimal choice after losses (i.e. upgrade).

As the self-downgrade strategy amounts to a simple three-step sequence repeated over and over, I anticipated an opponent following this strategy all of the time would be trivial to exploit. This would lead to a situation where the participant would have very few if any losses or draws during a series of trials. Due to this, the strategic opponent would have to have a degree of randomness to its play (as per Stöttinger et al., 2014). I decided that 70% of trials should follow the *exploitable* rule (similar to the training period in Stöttinger et al., 2014, which ranged from 60% to 80%). The remaining 30% of trials would consist of the computer opponent randomly picking, without replacement, from a flat distribution of the three potential choices. The choice to use sampling without replacement was made to ensure no accidental item biases on the opponents' part.

2.2.1.4 Metacognition and perception of the opponent. The fact that people have been found to engage in systematic behaviour in a random environment may also reflect a false belief of strategic advantage (see Section 1.3). That is, people may think that they are more likely to win, or that they have recognized a pattern in the opponent's behaviour even when there is none - a case of faulty metacognition, or an illusion of control (Langer, 1975). The tendency toward *stayshift* behaviour in games may also be understood as a tendency to expect and/or perceive "random" sequences to be positively autocorrelated (Scheibehenne et al., 2011; Wolford et al., 2004). Thus, it seems plausible that an explicit belief of a randomly playing opponent being *exploitable* could form. Consequently, I decided to include a confidence measure to gauge participants' certainty of a win throughout the trials. Explicit representations of confidence have been shown to play an integral role in value- based choice processing in the vmPFC (De Martino et al.,, 2013) and to predict future value-based choices and changes of mind depending on whether confidence is high or low, respectively (Folke et al., 2016). Thus, a higher confidence of winning based on faulty inferences of the behaviour of an *unexploitable* opponent could be associated with biased play behaviour.

However, the above-mentioned studies on confidence (De Martino et al., 2013; Folke et al., 2016) did not measure confidence in a competitive game context, where success and failure may affect a person's confidence in a choice: rather, the choices were made between reward items with no competition or clearly defined success involved. Confidence may not follow success in a task linearly. In Stöttinger et al. (2014), participants' confidence ratings increased through trials against frequency-biased and player-dependent RPS opponents to a similar degree, despite players making a different number of optimal plays against these opponents. However, the increase in confidence was not simply a function of time spent playing but coincided with above-chance success in exploiting the opponent: participants exposed to a shift in opponent strategy that led to suboptimal performance also decreased their confidence ratings. The similar increase in confidence for different rates of above-chance performance may then have been simply due to all above-chance performance leading to a similar increase, or due to the participants being able to recognize a pattern even for the "harder" player-dependent opponents, but being less able to play optimally.

In Experiment 1, I predicted that participants' confidence would be correlated with their success in the game leading to overall higher confidence in the *exploitable* condition given successful exploitation, and that the correlation between confidence and win-rate would be higher in the *exploitable* opponent condition, where participants would be more likely to know that they are responsible for their wins. To further examine the possible false belief of a strategic advantage against the random opponent, I included as an exploratory measure an open-ended question about the opponent's pattern of play after each block. In addition to giving more insight on how people actually view the *unexploitable*, random opponent, the answers would also shed light on how the participants understood the *exploitable* opponent's behaviour. Strictly speaking, the *exploitable* opponent does not care about the participants' actions, but participants may still believe so.

I also included three short questionnaires, after Dyson et al. (2016) and Forder and Dyson (2016), measuring game engagement (Brockmyer et al., 2009), felt co-presence (Nowak & Biocca, 2003) and perceived anthropomorphism in the opponent (Epley et al., 2008). The questionnaires were included to better understand the interpretations people make of the behaviour of *unexploitable* and *exploitable* computer opponents. As an additional exploratory measure, I also included the 60-item HEXACO personality questionnaire (Ashton & Lee, 2009) to examine whether certain traits might predict success in the game. Crucially, the questionnaire measures the Honesty/Humility trait, a construct similar to the Dark Triad traits (Lee & Ashton, 2014). These traits have been shown to correlate willingness to engage in risk-taking and impulsive behaviour (Crysel et al., 2013; De Vries et al., 2009), which could affect game outcomes.

2.2.1.5 Hypotheses. To summarize, in Experiment 1, I examined patterns of play against both *unexploitable (mixed strategy)* and *exploitable (self-downgrade)* opponents. Players played against both types of opponents under two different value conditions. I examined player confidence on ten different occasions for each game block, and recorded reaction times for all decisions. My hypotheses were:

1) Players will be more likely to repeat an item choice after a win than after a loss or a draw in the *high value* condition (similar to Forder & Dyson, 2016)

2) Players will be generally more likely to shift after a loss or a draw than after a win (similar to Dyson et al., 2016; Forder & Dyson, 2016; Wang et al., 2014)

3) Consequently, players will make more optimal choices after losses compared to draws or wins against the *exploitable* opponent

4) Confidence will be generally higher when playing against the *exploitable* opponent

5) The correlation between confidence and win-rate will be higher in the *exploitable* condition compared to the *unexploitable* condition

6) Reaction times after wins against the *unexploitable* opponent will be slower than reaction times after draws or losses (similar to Dyson et al., 2016; Forder & Dyson, 2016)

2.2.2 Method.

2.2.2.1 Participants. Forty subjects (N = 40; 31 female; M_{age} = 21.13, SD_{age} = 4.37) from the University of Sussex participant pool were recruited. Participants received course credit or £10 (their choice, unless course enrolment dictated they take the course credit) for their participation. Informed consent was obtained from all participants before testing, and the experiment was approved by the Sciences & Technology Research Ethics Committee (C-REC) at the University of Sussex (ER/JS753/1).

2.2.2.2 Materials.

2.2.2.2.1 Game trials. Static pictures of a white-gloved and a blue-gloved hand signaling Rock, Paper and Scissors poses were displayed center screen by the experiment program at approximately 6° x 6° each, with participants sat approximately 57 cm away from a 22" Diamond Plus CRT monitor (Mitsubishi, Tokyo, Japan). Stimulus presentation was controlled by Presentation 19 (build 03.31.15) and responses were recorded using a keyboard.

2.2.2.2.2 Questionnaires. I administered four short questionnaires following the completion of each RPS block to assess participants' engagement with the game, the degree of anthropomorphism assigned to the computer opponent, and co-presence felt between the player and opponent. First, I measured game engagement per block using a modified Game Engagement Questionnaire (Brockmyer et al., 2009). I changed the items from present to past tense (e.g. 'I lose track of time' became 'I lost track of time') and measured agreement on a 5-point Likert scale from 1 (“Strongly disagree”) to 5 (“Strongly agree”). I used sixteen of the original nineteen items: the items excluded from the modified version were 'I played longer than I meant to', ‘I felt like I just couldn’t stop playing’ and ‘I got really into the game’. As the number of game rounds was fixed, items relating to stopping playing did not fit; the last item was removed due to overlap with the other questionnaires. Second, I measured self-reported co-presence using items by Forder & Dyson (2016) and two modified items from scales by Nowak and Biocca (2003), using the same 5-point Likert scale as the GEQ. Third, I measured the degree of anthropomorphism attributed to the opponent on the basis of Epley et al. (2008), where five anthropomorphic states ('mind of its own', 'intentions', 'free will', 'consciousness', 'experienced emotion') and three non-anthropomorphic states ('attractive', 'efficient', 'strong') were measured on an 11-point Likert scale, from 0 (“Not at all”) to 10 (“Very much”). See Appendix 1 for the items of the modified questionnaires. Fourth, I asked participants to write on a piece of paper, in their own words, what they thought the opponent’s strategy, if any, had been in the block. Finally, to explore possible personality effects, I administered the 60-item HEXACO self-report questionnaire (Ashton & Lee, 2009) at the end of the experiment. With the exception of the writing prompt, the questionnaires were included in the experimental program.

2.2.2.3. Design and procedure. The experiment had a within-subjects 2x2 design with value *(high, low)* and opponent *(unexploitable, exploitable)* as factors. Each participant completed a block of 90 game trials for each four conditions (360 trials in total) in a semi-counterbalanced order across participants. The only constraint imposed on the counterbalancing orders were that no two consecutive blocks were allowed to be in the same opponent condition; this was to avoid ceiling effects due to learning in the *exploitable* condition.

In the *high value* condition, the participants gained 3 points for a win and lost 3 points for a loss; in the *low value* condition, they gained 1 point for a win and lost 1 point for a loss. In both conditions, a draw yielded 0. For the opponent, in the *unexploitable* condition, the computer opponent made each choice drawing randomly without replacement from an equal distribution (30 instances of R, P and S each). Note that this was slightly different from a true mixed strategy were the draws would be with replacement. I made this change in order to avoid any possibility of accidental item biases over the whole block caused by true randomness. In the *exploitable* condition, the computer opponent followed a *self-downgrade* rule for 70% of the trials (63 self-downgrade trials), making choices drawn at random, without replacement, from an equal distribution for the rest (9 instances of R, P and S each).

At the beginning of each block, the experimental program informed participants how much their wins and losses would affect the score, based on the value condition of the block. Regardless of the opponent condition of the block, participants were also informed that their opponent would play in a certain way that would be revealed to them at the end of the experiment. Participants were instructed to try and win as many rounds as possible.

For each trial, the participant was first presented with three pictures of a hand in a white glove representing the three possible choices (Rock, Paper, Scissors), presented in the same order as the response buttons used. Upon pressing a response button, the program presented the participant with a picture representing the choices made by the participant and the computer opponent. This picture consisted of the white-gloved hand presenting the response the participant had chosen, and a blue-gloved hand presenting the opponent's choice. This picture was presented for 500ms. Then, after a 500ms interval a text screen, presented for 1000ms, informed the participant if they had won, lost or drawn the trial. Finally, the scoreboard and trial counter, presented throughout the trials, were updated after another 500ms interval.

For every 9th trial of a block, after the participant had made their choice and before presenting the results, the program asked the participant to state their confidence of a win or a loss on a 5-point scale. The scale was from 1 for "extremely confident of win" through 3 for "unsure either way" to 5 for "extremely confident of loss". These items were reverse coded in the final analyses.

At the end of each block, the participants responded to the modified Game Engagement Questionnaire, the modified co-presence questionnaire, the anthropomorphism questionnaire, and the open-ended question of what they thought the opponent's strategy had been, if any. After the questionnaires, the participants were instructed to take a break before continuing with the experiment. After the final block, the participants filled out the 60-item HEXACO self-report personality inventory, after which I debriefed them and thanked them for their time.

2.2.3 Results.

2.2.3.1 Behavioural measures.

2.2.3.1.1 Item selection and outcome at trial n. I analysed proportions of item selection at trial *n* for each block using a three-way repeated measures ANOVA with opponent *(unexploitable, exploitable),* value *(high, low)* and item choice *(rock, paper, scissors)* entered as factors (see Table 2.1). In this and all subsequent analyses, I used Greenhouse-Geisser corrections to degrees of freedom whenever Mauchly's test indicated violations of sphericity. Due to using proportion data, main effects of the opponent and value conditions were meaningless (the proportions sum to 1, leading to no variance; see Dyson et al., 2016; Forder & Dyson, 2016). I made the choice to use an ANOVA, despite its assumptions being violated due to using proportion data, partially to maintain similarity with prior studies in the area using the method and partially due to practical reasons (no statistical package offering a clear optimal solution to analysing multinomial proportion data available at the time). I will cover this issue in more depth in the general discussion for Chapter 2 (see Section 2.4).

Table 2.1. Proportions of outcomes and item choices in Experiment 1

	Unexploitable opponent			*Exploitable opponent*		
	Rock	*Paper*	*Scissors*	*Rock*	*Paper*	*Scissors*
High value	36.2%	31.5%	32.3%	34.6%	32.2%	33.3%
	(1.1%)	(1.0%)	(1.2%)	(0.9%)	(0.9%)	(1.0%)
Low value	34.5%	32.0%	33.4%	34.2%	33.3%	32.6%
	(1.3%)	(1.3%)	(2.1%)	(1.1%)	(1.0%)	(1.2%)
	Win	*Lose*	*Draw*	*Win*	*Lose*	*Draw*
High value	33.4%	34.6%	32.0%	48.9%	23.7%	27.4%
	(0.7%%)	(1.0%)	(0.8%)	(2.5%)	(1.4%)	(1.6%)
Low value	33.5%	33.7%	32.8%	49.9%	23.5%	26.6%
	(0.9%)	(0.8%)	(0.8%)	(2.5%)	(1.6%)	(1.2%)

Note: standard error in parentheses.

Participants did not differ in their item choice [$F(1.61, 62.91) = 1.27$, $MSE = .03$, $p = .282$, $\eta_p^2 = .03$], nor did this differ according to value [$F(2, 78) = 0.62$, $MSE = .01$, $p = .538$, $\eta_p^2 = .02$], or opponent [$F(2, 78) = 1.04$, $MSE < .01$, $p = .358$, $\eta_p^2 = .03$]. There was no three-way interaction [$F(1.58, 61.69) = 0.64$, $MSE = .01$, $p = .494$, $\eta_p^2 = .02$].

I conducted the same analysis for outcome at trial *n (win, lose, draw;* see Table 2.1). Here, there was a significant main effect for outcome at trial *n* [$F(1.30, 50.55) = 46.86$, MSE = .03, p < .001, $\eta_p^2 = .55$], and a significant interaction with opponent [$F(1.17, 45.71) = 36.09$, MSE = .04, $p < .001$, $\eta_p^2 = .48$], but not with value [$F(2, 78) = 0.24$, MSE = .01, $p = .785$, $\eta_p^2 = .01$]. Neither was there a significant three-way interaction [$F(1.71, 66.49) = 0.30$, MSE = .01, $p = .705$, $\eta_p^2 = .01$]. Thus, the only factor driving the uneven distribution of wins, losses and draws was opponent condition. When playing against an *unexploitable* opponent, wins, losses and draws were distributed roughly uniformly (33.4%, 34.1% and 32.4% of trials, respectively), but participants won significantly more often when playing against an *exploitable* opponent (49.4%, 23.6% and 27.0%: Tukey's HSD; $p < .05$). The results suggest that on average, participants were able to apply the correct strategy against the *exploitable* opponent, but the degree of winning was not modulated by value.

2.2.3.1.2 First-order repetition effects. I analysed proportion data using the last 89 trials in each block (the first trial having no history) using a four-way repeated- measures ANOVA, with opponent *(unexploitable, exploitable),* value *(low, high),* outcome at trial *n (win, lose, draw)* and player strategy at trial *n+1 (stay, upgrade, downgrade)* as factors. See Table 2.2 for distribution of player strategies across the four conditions. Due to using proportion data, main effects of the opponent and value conditions were meaningless (the proportions sum to 1, leading to no variance; see Dyson et al., 2016; Forder & Dyson, 2016). See section 2.4 for further discussion on the methodological choice to use ANOVAs for proportion data.

Table 2.2. Proportions of choice types in Experiment 1

	Low value			*High value*		
	Unexploitable			*Unexploitable*		
	Win	*Lose*	*Draw*	*Win*	*Lose*	*Draw*
Stay	29.2%	32.6%	34.7%	28.2%	31.1%	35.9%
	(2.9%)	(2.8%)	(2.9%)	(3.0%)	(2.2%)	(2.6%)
Upgrade	33.5%	30.8%	29.3%	32.6%	32.1%	30.7%
	(2.3%)	(2.0%)	(1.9%)	(2.1%)	(1.8%)	(1.9%)
Downgrade	37.3%	36.6%	35.9%	39.2%	36.8%	33.4%
	(2.6%)	(2.4%)	(2.2%)	(2.6%)	(2.0%)	(2.3%)
	Exploitable			*Exploitable*		
	Win	*Lose*	*Draw*	*Win*	*Lose*	*Draw*
Stay	18.1%	32.9%	53.5%	15.6%	30.6%	50.3%
	(3.0%)	(2.9%)	(3.6%)	(2.4%)	(2.6%)	(3.9%)
Upgrade	20.0%	39.6%	19.1%	20.7%	39.4%	22.1%
	(2.7%)	(3.0%)	(2.0%)	(2.7%)	(2.9%)	(2.2%)
Downgrade	61.8%	27.5%	27.3%	63.6%	30.0%	27.6%
	(4.6%)	(2.4%)	(2.2%)	(4.2%)	(2.1%)	(2.4%)

Note: standard error in parentheses.

The main effect of strategy at trial *n+1* was significant [$F(1.52, 59.24) = 7.73$, $MSE = .17$, $p = .003$, $\eta_p^2 = .17$], indicating that strategy at *n+1* was not random. There was a significant interaction effect between opponent and player strategy at trial *n+1* [$F(2, 78) = 6.77$, $MSE = .03$, $p = .002$, $\eta_p^2 = .15$], as well as outcome at trial *n* and strategy at trial *n+1* [$F(1.93, 75.09) = 33.98$, $MSE = .13$, $p < .001$, $\eta_p^2 = .47$]. Further, there was a significant three-way interaction between opponent, outcome at trial *n* and strategy at trial *n+1* [$F(2.13, 83.23) = 31.12$, $MSE = .08$, $p < .001$, $\eta_p^2 = .44$]. See Figure 2.2 for distribution of player strategies across the four conditions. The interactions between value and player strategy at trial *n+1* [$F(2, 78) = 0.53$, MSE = .04, $p = .589$, $\eta_p^2 = .01$], value, opponent and player strategy at trial *n+1* [$F(1.74, 68.03) = 0.45$, MSE = .03, $p = .611$, $\eta_p^2 = .01$] and between all four factors [$F(2.95, 114.87) = 0.20$, MSE = .03, $p = .89$, $\eta_p^2 = .01$] were all non-significant, suggesting no effect of the value manipulation on behaviour, contrary to my hypotheses.

In the *unexploitable* opponent condition, there were no significant differences between the proportions of reactions to different outcomes (Tukey's HSD, $p > .05$ for all comparisons), contrary to my hypotheses and previous studies (Dyson et al., 2016; Forder & Dyson, 2016). For the *exploitable* opponent condition, the optimal *win-downgrade* responses were significantly more likely than other responses to wins, and the optimal *draw-stay* responses were significantly more likely than other responses to draws (Tukey's HSD, $p < .05$ for each comparison). However, the proportion of the optimal *lose-upgrade* responses did not differ significantly from the *stay* or *downgrade* responses to losses (Tukey's HSD, $p > .05$ for both comparisons). The proportion of optimal responses to wins was not significantly different from optimal responses to draws (Tukey's HSD, $p > .05$), but the proportion of optimal responses to losses was lower than the proportions of optimal responses to both wins and draws (Tukey's HSD, $p < 05$ for both comparisons). Taken together, the results suggest participants were on average likely to respond optimally to both wins and draws against the *exploitable* opponent, but failed to respond optimally after losses.

I further explored the *exploitable* data by categorizing participants' win-rates as successful or failed based on a one-tailed one-sample proportions test, with a proportion of 33.3% set as the null hypothesis, run separately for each participant in each block. Of the 40 participants, 8 failed to reach a win percentage significantly higher than chance level on both blocks, and 15 other participants failed on one block out of two (8 in the *low value* and 7 in the *high value* block). Thus, there were 24 successful participants in the *low value* block and 25 in the *high value* block. For the successful participants, percentages of optimal responding distributed across the three outcomes was similar to that of the entire sample (see Figure 2.3). Therefore, the reduction in optimal behaviour after loss seems to not have been driven by the unsuccessful participants. Rather, the participants who were unsuccessful at exploitation did not show strategic learning following any outcome, and instead behaved similarly to an overall MES trend in both conditions.

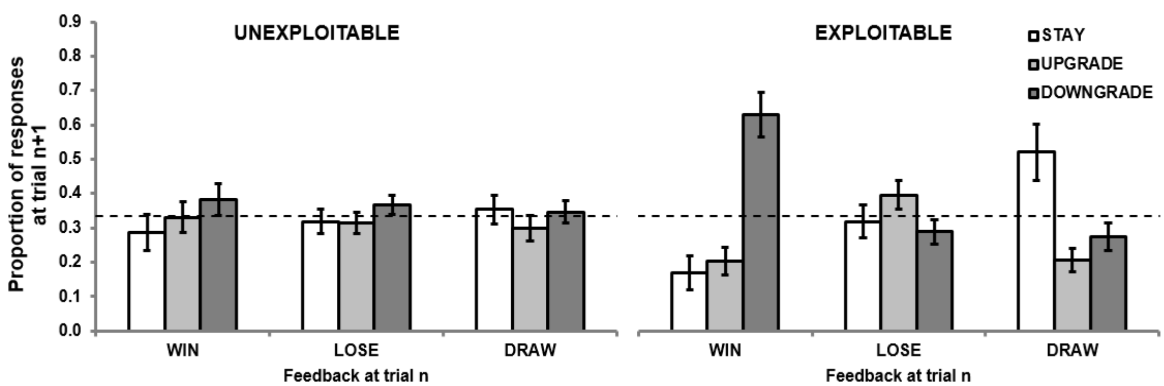

Figure 2.2. Distributions of participants' strategic choices collapsed across the value conditions in Experiment 1. Error bars represent 95% CIs. Dashed line represents chance-level responding.

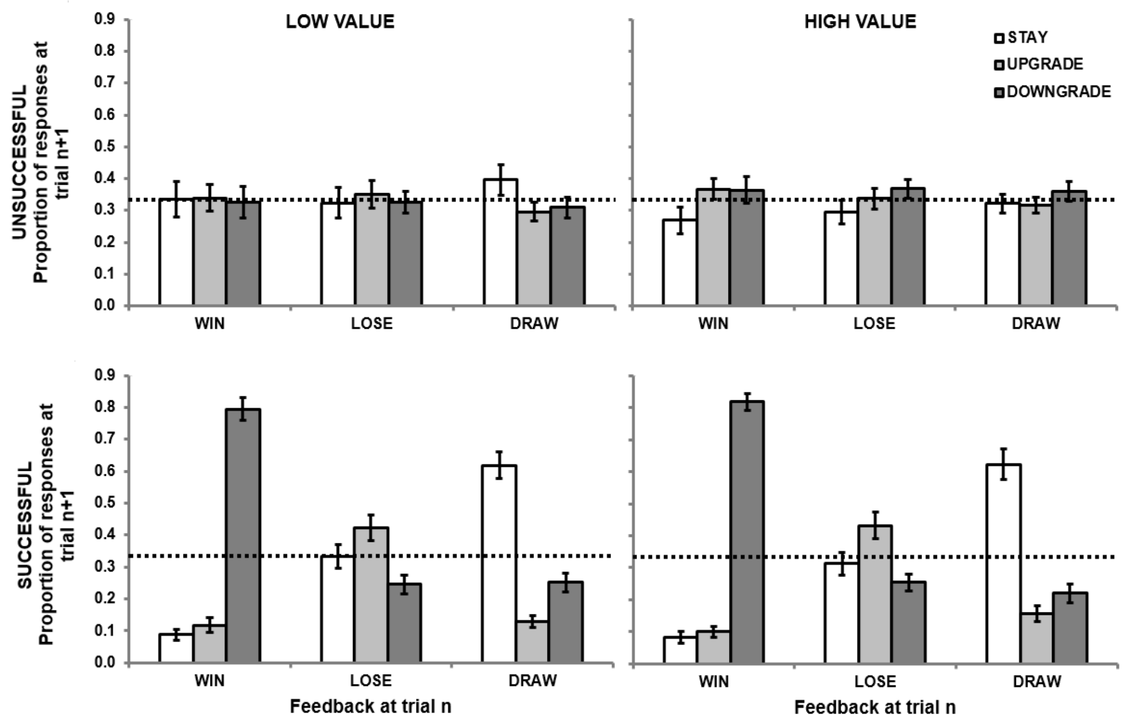

Figure 2.3. Distributions of unsuccessful and successful participants' strategic choices in the *exploitable* condition in Experiment 1. Error bars represent SEs. Dashed line represents chance-level responding.

2.2.3.1.3 Reaction time analysis. I entered reaction time data for decision at trial n+1 into a three-way repeated measures ANOVA with opponent (unexploitable, exploitable), value (high, low) and outcome on trial n (win, loss, draw) entered as factors. Ten participants were excluded from this analysis due to having at least one average median reaction time (averaged across wins, losses and draws) that was at least twice the block average median (after Forder & Dyson, 2016).

There was a significant main effect of outcome [$F(2, 58) = 10.58$, $MSE = 108292$, $p < .001$, $\eta_p^2 = .27$] and of opponent [$F(1, 29) = 5.09$, $MSE = 178580$, $p = .032$, $\eta_p^2 = .15$], but no significant main effect of value [$F(1, 29) = 1.38$, $MSE = 202493$, $p = .250$, $\eta_p^2 = .05$].

The interaction between value and opponent [$F(1, 29) = 3.14$, $MSE = 426768$, $p = .087$, $\eta_p^2 = .10$], the interaction between value and outcome [$F(2, 58) = 1.88$, $MSE = 67234$, $p = .163$, $\eta_p^2 = .06$] and the three-way interaction of opponent, value and outcome [$F(2, 58) = 0.13$, $MSE = 118297$, $p = .875$, $\eta_p^2 = .01$] were non-significant. There was, however, a significant interaction effect between outcome and opponent [$F(2, 58) = 3.88$, $MSE = 73120$, $p = .03$, $\eta_p^2 = .118$, see Figure 2.4]. Post-hoc tests revealed the reaction times for game choices made after losses were significantly slower in the *exploitable* opponent condition compared to the *unexploitable* opponent condition (1257ms and 868ms, respectively; Tukey's HSD, $p < .05$), with no other differences in reaction times between the opponent conditions. For both opponent types, reaction times for wins were slower than reaction times for draws, replicating the results of Forder & Dyson (2016), suggesting more cognitive control for decisions made after wins against the *unexploitable* opponent. The marginal interaction effect between value and opponent type seems to have stemmed from an increase in reaction times (especially after wins) against the unexploitable opponent when value was high compared to when it was low (see Figure 2.4). There was also an increase in the variability of reaction times in the high value condition, suggesting that the value manipulation may have affected different participants to differing degree.

To further explore the relationship between reaction times and performance, I plotted individual win-rates (averaged across the two *exploitable* blocks) against the reaction time difference between decisions made after losses and wins in the *exploitable* condition. A significant positive correlation ($r = .52$, $p < .005$; see Figure 2.5) indicates that reaction times after losses in relation to wins increased as an individual participant's win-rate increased.

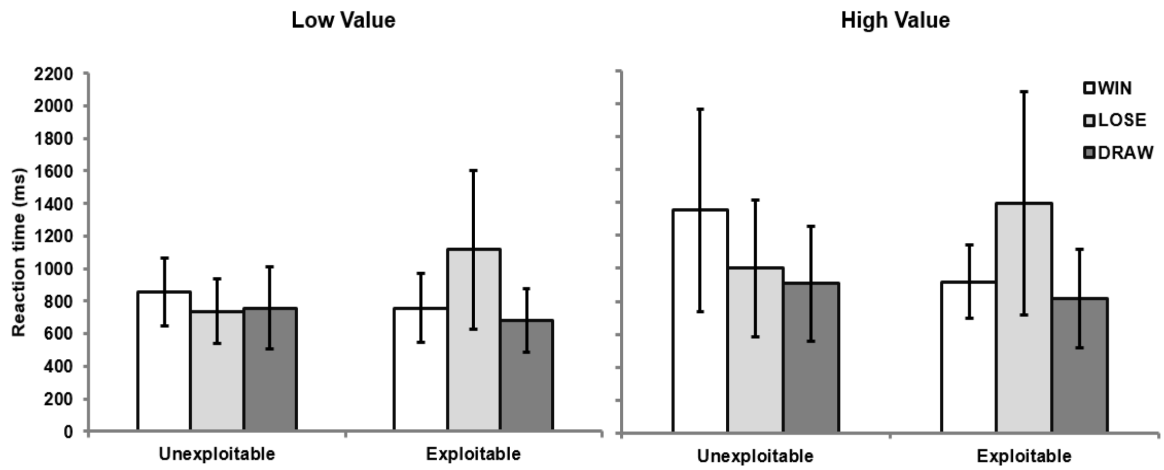

Figure 2.4. Means of average median reaction times in Experiment 1. Error bars represent 95% CIs.

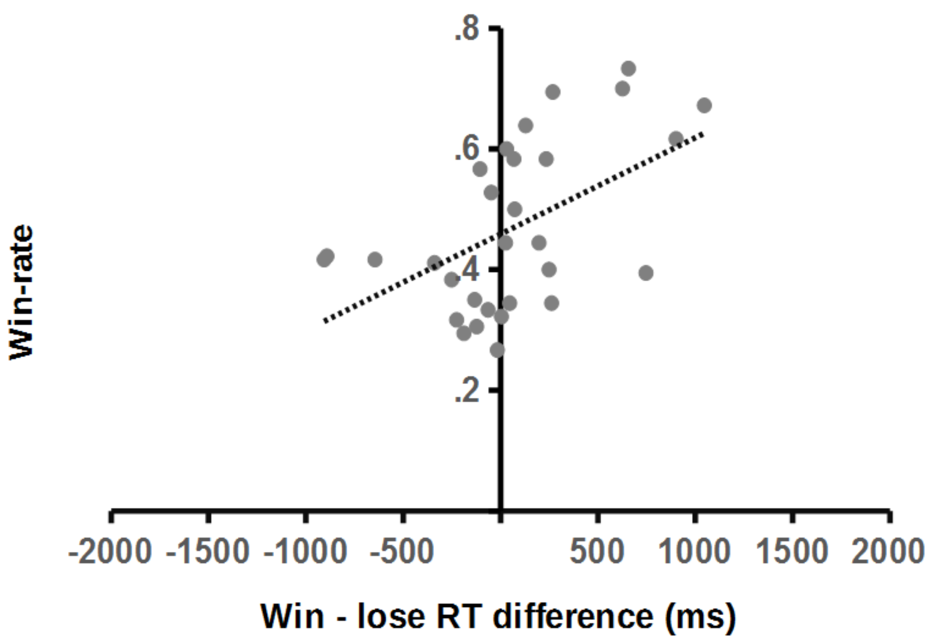

Figure 2.5. Correlation between win-rate and difference between average win and lose reaction times in the *exploitable* conditions in Experiment 1. A positive RT difference indicates slower post-error than post-success RTs.

2.2.3.1.4 Mean confidence measure. I analysed mean confidence rates using a two-way repeated measures ANOVA with opponent *(unexploitable, exploitable)* and value *(high, low)* entered as factors. There was a significant main effect of opponent [F(1, 39) = 14.93, MSE = .58, $p < .001$, $\eta_p^2 = .28$], no significant main effect of value [F(1, 39) = 0.13, MSE = .16, $p = .726$, $\eta_p^2 < .01$], and no significant interaction [F(1, 39) = 0.46, MSE = .18, $p = .501$, $\eta_p^2 = .01$]. Mean confidence rates were higher when playing against the *exploitable* opponent than when playing against the *unexploitable* opponent (see Table 2.3).

I examined the relationship between confidence ratings and individual win rates by first calculating z-transformed correlation coefficients based on the 10 data points for each player in each block. Each data point consisted of the reported confidence every 9 trials and the average win rate for the preceding 8 trials. Six participants had to be excluded due to having no variance in their reported confidence on one or more blocks. For the remaining thirty-four participants, I calculated an average correlation measure for each condition by averaging Fisher transformed individual correlation scores, with the averages inverse Fisher-transformed to produce a condition-level correlation (see Table 2.3). I entered the Fisher-transformed individual correlation coefficients into a two-way repeated measures ANOVA with opponent *(unexploitable, exploitable)* and value *(low, high)* entered as factors. There was a significant main effect of opponent [F(1,33) = 4.51, MSE = .27, $p = .041$, $\eta_p^2 = .12$], no significant main effect of value [F(1,33) = 1.91, MSE = .10, $p = .174$, $\eta_p^2 = .06$], and no significant interaction [F(1,33) < 0.001, MSE = .192, $p = .993$, $\eta_p^2 < .01$]. Participants had higher z-scores in the *exploitable* (M = .38, SE = .08) compared to the *unexploitable* (M = .20, SE = .05) conditions. The results suggest that confidence ratings, on average, more accurately tracked win-rates when playing against an *exploitable* opponent, but note that the final condition-level correlations were all relatively low (all below .4).

Table 2.3. Mean confidence measure (Likert, 1-5), n = 40, and correlation between confidence and win-rate in Experiment 1, n = 34

Mean confidence (range: 1 – 5)		
	Unexploitable opponent	*Exploitable opponent*
High value	3.1 (.1)	2.6 (.1)
Low value	3.1 (.1)	2.7 (.1)
Mean confidence measure / win-rate correlations (Fisher transformed z values)		
	Unexploitable opponent	*Exploitable opponent*
High value	.23 (.06)	.40 (.10)
	[0.22]	[0.38]
Low value	.16 (.07)	.33 (.09)
	[0.16]	[0.32]

Note: standard error in parentheses. Inverse Fisher transformed z values in brackets.

2.2.3.2 Questionnaire data. I entered the end-of-block questionnaire data into separate two-way ANOVAs, with the opponent *(unexploitable, exploitable)* and value *(low, high)* conditions entered as factors for each separate questionnaire analysis. See Table 2.4 for descriptive statistics.

2.2.3.2.1 Game engagement questionnaire. There were no significant main effects for opponent [$F(1, 39) = 0.02$, MSE = .14, $p = .897$, $\eta_p^2 = .00$] or value [$F(1, 39) = 1$, MSE = .16, $p = .323$, $\eta_p^2 = .03$], or a significant interaction effect [$F(1, 39) = 0.73$, MSE = .09, $p = .397$, $\eta_p^2 = .02$]. Game engagement was similar across all blocks.

2.2.3.2.2 Co-presence. There was a significant main effect of opponent [F(1, 39) = 4.88, MSE = .47, p = .033, η_p^2 = .11] but no significant main effect of value [F(1, 39) = 0.57, MSE = .23, p = .457, η_p^2 = .01],or a significant interaction effect [F(1, 39) = 0.10, MSE = .42, p = .756, $\eta_p^2 < .01$]. Felt co-presence was higher in the *unexploitable* condition (M = 2.8, SE = .1) than in the *exploitable* condition (M = 2.6, SE = .1).

2.2.3.2.3 Anthropomorphism. There was a significant main effect of opponent strategy [F(1, 39) = 7.69, MSE = 2.49, p = .008, η_p^2 = .17] and no significant main effect of value [F(1, 39) = 0.41 MSE = 1.36, p = .528, η_p^2 = .01] or a significant interaction [F(1, 39) = 0.05, MSE = 2.06, p = .818, $\eta_p^2 < .01$]. The perceived anthropomorphism of the opponent was higher in the *unexploitable* condition (M = 3.8, SE = .3) than in the *exploitable* condition (M = 3.1, SE = .3).

2.2.3.2.4 Free-form question. Data for majority of participants was lost due to filing error, with responses from only 16 participants remaining. Themes of responses were used as a basis for a short questionnaire to be used in Experiment 2 (see Appendix 2).

Table 2.4. Mean game engagement (range = 1 – 5), felt co-presence (range = 1 – 5) and perceived anthropomorphism (range = 1 – 11) in Experiment 1

	Game engagement	
	Unexploitable opponent	*Exploitable opponent*
High value	2.5 (0.1)	2.5 (0.1)
Low value	2.4 (0.1)	2.5 (0.1)
	Co-presence	
	Unexploitable opponent	*Exploitable opponent*
High value	2.8 (.1)	2.6 (0.1)
Low value	2.8 (.2)	2.5 (0.1)
	Anthropomorphism	
	Unexploitable opponent	*Exploitable opponent*
High value	3.7 (0.3)	3.0 (0.3)
Low value	3.8 (0.4)	3.1 (0.3)

Note: standard deviation in parentheses.

2.2.3.2.5 HEXACO. I entered the six HEXACO factor scores for Honesty/Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness and Openness as covariates into a two-way repeated measures ANCOVA, with individual win-rates as the dependent variable and the opponent *(exploitable, unexploitable)* and value *(high, low)* conditions as factors. My aim in this analysis was to explore whether personality measures affected success in the learning task. I made the methodological choice of using an ANCOVA this way due to the suggestions made by Schneider et al. (2015) on the use of ANCOVAs in within-subjects designs. Specifically, Schneider et al. (2015) recommend that in within-subjects designs, ANCOVAs should only be used to analyse the effects of the covariate itself and the covariate's interactions with the within-subjects factors. Conversely, ANCOVAs should not be used to analyse the main effects of the within-subjects factors because the within-subjects main effect estimates can be distorted by covariate variance.

Thus, in this case an ANCOVA is not used to measure the “pure” effect of the independent within-subjects variables, controlling for the covariate, but to specifically measure the effect of the covariate on the dependent variable. The ANCOVA yielded no significant main or interaction effects for any of the personality measures ($p > .05$ for each effect).

2.2.4. Discussion. In Experiment 1, my hypothesis that participants would be generally more likely to shift after a loss or a draw against *unexploitable* opponents was not supported by the behavioural data. The participants exhibited no *stayshift* bias in the form that it has been observed before against an *unexploitable* opponent in RPS (Dyson et al., 2016; Forder & Dyson, 2016), in studies using binary choice tasks (e.g. Achtziger et al., 2015; Scheibehenne et al., 2011; Wilke et al., 2014), or in animal studies (Lee et al., 2005). According to Rapoport and Budescu (1992), players may more easily produce mixed-strategy behaviour in a competitive game situation in contrast to passive production tasks (see Neuringer, 1986; Terhune & Brugger, 2011). However, the participants in Rapoport and Budescu’s (1992) experiments only expressed such randomness when they were playing against each other. As two human players playing RPS would likely try to exploit one another, a cycle of attempts at exploitation could lead to randomness in the long run. This is in contrast to Experiment 1, where the *unexploitable* opponent never attempted to exploit the participant even if they expressed biases (see also Lee et al., 2005, for an instance of monkeys starting to approximate MES after a computer opponent started exploiting their predictability). Thus, the dynamics between the two players in Experiment 1 and prior studies that showed a *stayshift* bias were likely very different than in studies pitting humans against each other (see West & Lebiere, 2001, for an exploration of a plausible model of how humans play). Nevertheless, unlike prior studies, the players in this study seem to have avoided following the *stayshift* rule to a significant degree.

The results from the *exploitable* condition were also contrary to my hypotheses. I expected that participants would be drawn to follow the *stayshift* rule and thus make less optimal choices against the opponent when the optimal strategy conflicted with this rule. Here, a majority of participants succeeded in learning to take advantage of the *exploitable* opponent as shown by a greater-than-chance average win-rate and non-random choices as function of the outcome of the previous round. However, the rate of optimal choices differed between outcomes, but not in the way that would support the notion of reinforcement biases. Players made more optimal choices after wins (62.7%) and draws (51.9%) relative to losses (39.5%). This was the case even though the optimal strategies after wins and draws were contrary to reinforcement (*win-downgrade* and *draw-stay*) whereas the optimal strategy after losses was not (*lose-upgrade*). It is worth noting here that participants in the *exploitable* condition had no significant differences in their rates of losses and draws (*low value*: 23.47% and 26.58%; *high value*, 23.69% and 27.44%, respectively; Tukey's HSD; $p < .05$ for both). Thus, it is unlikely that the difference in performance between losses and draws was due to the players having had less chances to learn the optimal choice to be made after a loss.

Further, all optimal moves in the experiment were counter to myopic best reply, which would dictate *staying* after wins, *upgrading* after draws and *downgrading* after losses. Thus, there is no confound with this strategy (see Alós-Ferrer & Ritschel, 2018). There is, however, a potential confound with decision inertia (Alós-Ferrer et al., 2016; Alós-Ferrer & Ritschel, 2018): the optimal *stay* response after draws may be easier for participants to adopt due to alignment with inertia, that is, the automatic tendency to repeat choices regardless of outcomes. Thus, the more optimal behaviour after draws compared to losses may stem from the optimal *draw-stay* rule aligning with inertia. I decided to control for this confound in Experiment 2 by introducing a different set of optimal responses. However, even if this confound would explain the differences in performance between losses and draws, the fact remains that participants made the most optimal responses after wins, where the optimal strategy (*downgrade*) was in conflict with reinforcement (*stay*), myopic best reply (*stay*) and inertia (*stay*), suggesting more strongly that specific outcomes mattered in terms of performance.

Additionally, losses against the *exploitable* opponent also led to slower reaction times than losses against the *unexploitable* opponent. The difference between lose and win reaction times against the *exploitable* opponent was positively correlated with individual success rates, meaning higher win-rates led to more deliberation after losses compared to wins. On one hand, these results indicate post-error slowing (Rabbitt & Rodgers, 1977) which should reduce the likelihood of errors (Dutilh et al., 2012). On the other hand, decisions following losses gave rise to worse rather than better performance - there was no cognitive benefit of the longer decision time. However, this may only seemingly be a contradiction. Dutilh et al. (2012, p. 463) describe post-error slowing as follows: "people adaptively change their response thresholds to a possibly non-stationary environment—by becoming more daring after each correct response and more cautious after each error, people reach an optimal state of homeostasis that is characterized by fast responses and few errors". If people become more cautious after errors (in the case of Experiment 1, losses against the *exploitable* opponent), this cautiousness is not guaranteed to lead to optimal performance. Participants may have taken losses, especially when they should have won (i.e. they recognized the specific way in which the opponent was *exploitable*, but the opponent then made a random move) as signs of something in the opponent's behaviour having changed, which may have led to more exploratory (random) behaviour (Wilson et al., 2014).

The value manipulation in Experiment 1 yielded no effects on decision-making. As changes in game choices have previously been observed with simple point manipulations with no monetary reward (Forder & Dyson, 2016), this lack of effect was not necessarily due to the lack of monetary reward. The lack of behavioural change may simply reflect the fact that the rewards and penalties were symmetrical. Forder & Dyson (2016) observed an effect of a score-based value manipulation were the penalties and rewards were different in absolute numerical value (e.g., +2 vs. -1). In Experiment 1, the rewards and penalties were of different magnitudes in the two conditions, but these magnitudes were always equal within the condition (e.g., +3 vs. -3). To more strongly incentivize the participants, in Experiment 2 I replaced the *high value* condition with a monetary incentive condition. The value and opponent conditions in Experiment 1 had a marginal interaction effect, with a trend towards longer reaction times in the unexploitable, high value condition. This trend was most pronounced for win reaction times, perhaps reflecting participants thinking longer when value is high and especially when they have won in a seemingly random game. It also seems that there was more variability in reaction times in the high value condition: this may be another sign that the value manipulation in its present form was too weak to affect all players and may have had an effect on behaviour in only some participants.

Participants in Experiment 1 were on average more confident of winning against the *exploitable* opponent, consistent with appropriately different meta-cognitive understanding about the two types of opponent. However, confidence measures did not only reflect higher win-rates in the *exploitable* condition: additionally, the correlation between win-rate and confidence was higher in the *exploitable* condition. This suggests that the participants' confidence ratings tracked win-rates more accurately against the *exploitable* opponent, further suggesting that perceived exploitability may independently affect the appraisals of the likelihood of winning participants make as the game goes on.

Finally, the participants rated the *unexploitable* opponent as more anthropomorphic and more present than the *exploitable* opponent. It seems then that quasi-randomness is considered more human-like. In their written descriptions and in debriefing, several participants expressed that the *unexploitable* opponent felt hard to play against, or that it seemed to change strategy whenever the player was on a winning streak. It seems reasonable that the *exploitable* opponent would seem less human-like, as it would not correct as a result of a participant successfully exploiting it. It may be that the *unexploitable* strategy may seem "responsive" simply due to the fact that it avoids situations where it seems the opponent is "allowing" the player to win for an extended period of time. This, however, does not seem to reflect as biased behaviour against the opponent.

2.3 Experiment 2

2.3.1 Introduction. In Experiment 1, I found no evidence of inflexible following of reinforcement principles in choices in RPS, regardless of whether the players were playing against an *unexploitable* opponent or an opponent who could be exploited. On average, losing led to the most suboptimal game choices when playing against the *exploitable* opponent even though losses were the only case in Experiment 1 where a reinforcement-aligned decision was the optimal one (i.e., *lose-shift* versus *win-shift* or *draw-stay*). The rate of optimal choices after losses against the *exploitable* opponent was significantly lower than the rate of optimal choices after wins or draws, and further did not differ significantly from the rate of either suboptimal game choice after losses. This effect did not seem to be due to the rarity of losses relative to draws. It is notable that the rate of optimal decisions made after wins was not significantly different from the rate of optimal decisions after draws, even though for a participant successfully exploiting the *exploitable* opponent, wins would be more common than both draws and losses, which had roughly equal probability. Participants should thus have had equal opportunities to learn the optimal choice to make after both draws and losses. The difference between the rate of optimal play between draws and losses would imply that losses specifically affect decision-making in a way draws don't (e.g., Dixon, MacLaren, Jarick, Fugelsang & Harrigan, 2013; Ulrich & Hewig, 2017). Additionally, the reaction times for decisions made after losses were longer in the *exploitable* condition compared to the *unexploitable* condition. Taken together, the results suggest that losses against an *exploitable* opponent led to more deliberation but also less optimal play, possibly reflecting exploratory behaviour. I predicted this trend in reaction times would continue in Experiment 2.

However, there was a potential confound in the way the optimal strategy against the *exploitable* opponent was set up in Experiment 1 that may account for the observed differences in optimal responding between losses and draws: if and when there was a draw, the optimal choice was to *stay*, and if and when there was a loss, the optimal response was to *upgrade*. While both of these responses were in conflict with reinforcement and myopic best reply, the stay response after draws was aligned with inertia (Alós-Ferrer & Ritschel, 2018). The participants may have fared better after draws not because draws have a different effect on behaviour from losses, but simply because alignment with inertia makes the decision easier. To control for this, In Experiment 2, I changed the *exploitable* opponent's strategy from a self-downgrade rule to a *self-upgrade* rule while keeping the number of trials and the likelihood of following the rule consistent with Experiment 1. The *self-upgrade* rule leads to a different set of optimal choices for the player: the optimal move after wins is *upgrading*, the optimal move after draws is *downgrading*, and most importantly the optimal move after losses is *staying*. To formulate:

(13) IF S(n) = r AND (n) = W THEN S(n+1) = p *(upgrade)*

(14) IF S(n) = r AND (n) = L THEN S(n+1) = r *(stay)*

(15) IF S(n) = r AND (n) = D THEN S(n+1) = s *(downgrade)*

(16) IF S(n) = p AND (n) = W THEN S(n+1) = s *(upgrade)*

(17) IF S(n) = p AND (n) = L THEN S(n+1) = p *(stay)*

(18) IF S(n) = p AND (n) = D THEN S(n+1) = r *(downgrade)*

(19) IF S(n) = s AND (n) = W THEN S(n+1) = r *(upgrade)*

(20) IF S(n) = s AND (n) = L THEN S(n+1) = s *(stay)*

(21) IF S(n) = s AND (n) = D THEN S(n+1) = p *(downgrade)*

where S = self, n = number of trial, W = win, L = lose, D = draw, r = rock, p = paper and s = scissors.

The optimal choices are again all in conflict with myopic best reply. Additionally, the optimal choices following wins and losses are in conflict with reinforcement, but the optimal choice after losses is aligned with inertia. If the suboptimal performance after losses but not after draws in Experiment 1 was due to the optimal *stay* response to draws being aligned with inertia and thus easier to learn, one should expect to see better performance after losses and worse performance after draws in Experiment 2. If, on the other hand, the effect on performance was due to something fundamental about the experience of loss causing the players to either fail to learn the optimal strategy or to suspect the opponent has changed their strategy and thus play in a more exploratory manner, one should expect a replication of Experiment 1. I will assess this with a cross-experiment comparison in addition to analysing the results of Experiment 2 separately, as a significant difference in one data set and a non-significant difference in another data set might not be significantly different from each other (Gelman & Stern, 2006).

Additionally, to make the difference between high and *low value* in my manipulation clearer, I changed the value manipulation from points to a monetary reward to offer a stronger incentive in Experiment 2. In economic decision tasks, a monetary incentive tends to shift behaviour closer to equilibrium strategies and optimal choices (Smith & Walker, 1993). However, this may not be the case with the present task, as monetary incentives may also make reinforcement stronger and thus reinforcement errors more common (Achtziger et al., 2015). Given that the *stayshift* bias is specifically a tendency to follow reinforcement, I expected to see more of it in the *unexploitable* condition when incentive is offered. Additionally, given that the optimal choices for rounds following wins and losses in the *exploitable* condition will be in conflict with reinforcement, I expected less optimal choices in the *exploitable* condition when the incentive is offered.

Participants were on average more confident against the *exploitable* opponent in Experiment 1. Given that the correlation between confidence ratings made throughout the block and the win-rate on the 9 trials preceding each confidence rating was higher in the *exploitable* condition, confidence ratings may reflect participants' beliefs of exploitability, in addition to their success. In Experiment 2, I included an additional measure of confidence by asking the participants to rate to what extent they thought their results in each block were dictated by luck as opposed to skill. I expected participants would associate their results in the *exploitable* condition more to skill than luck.

Finally, to have a better understanding of the reasons behind the considerable individual differences in rates of optimal play in Experiment 1 (with only 17 out of 40 participants successfully exploiting the *exploitable* opponent on both blocks; see Figure 2.3), I added a working memory (WM) task to Experiment 2 (the Operation Span task; after Turner & Engle, 1989; Unsworth et al., 2005). Whether the successful participants understand the actions of the strategic opponent as a sequence of different items or, in an arguably more complex way, as a set of contingencies between the outcome of the previous trial and the next item (see formulae 13-21), it seems plausible that the participants' performance relies to some degree on their WM capacity. I hypothesized that the participants' rate of optimal play against the *exploitable* opponent would be predicted by their WM capacity. Additionally, I predicted that a lower WM capacity would predict a better approximation of randomness against the random opponent, as temporarily reduced WM capacity predicts better performance when participants are asked to produce random sequences (Terhune & Brugger, 2011). As Terhune & Brugger (2011) also found evidence that participants with high levels of executive control also predicted better production of random sequences, I also added a flanker task (Eriksen & Eriksen, 1974) to measure executive control. I hypothesized that higher rates of executive control would correlate with fewer deviations from randomness in the *unexploitable* condition.

To summarize, in Experiment 2, I again examined patterns of play against both *unexploitable (mixed strategy)* and *exploitable (self-upgrade)* opponents. Players played against both types of opponents under two different payoff conditions (monetary incentive vs. no incentive). I examined player confidence on ten different occasions for each game block, and recorded reaction times for all decisions. Additionally, I asked participants to rate after each block how much, on a scale from 0% to 100%, they thought their results were a result of skill or luck. After the game rounds, participants' WM capacity and level of executive control were tested. The main aim of Experiment 2 was to control for a confound brought on by the optimal strategy against the *exploitable* opponent in Experiment 1

2.3.1.1 Hypotheses. My hypotheses were:

1) If the trend of suboptimal play after losses but not wins would continue in Experiment 2, this would support the notion that losses but not draws lead to suboptimal behaviour in the task. If the trend were to reverse, it would support the notion that alignment with inertia may make an optimal response easier to learn even if it is in misalignment with reinforcement.

2) More stayshift behaviour in the *high value*, *unexploitable* condition

3) Less optimal choices in the *high value, exploitable* conditions

3) In the *exploitable* condition, win-rate will correlate positively with the difference between reaction times after wins and losses

4) Reaction times after wins will be slower in the *unexploitable* condition compared to the *exploitable* condition

5) Average confidence ratings will be higher in the *exploitable* than in the *unexploitable* condition

6) Ratings of skill will be higher in the *exploitable* than in the *unexploitable* condition

7) The correlation between confidence and win-rate will be higher in the *exploitable* than in the *unexploitable* condition

8) WM capacity will be positively correlated with success against the *exploitable* opponent

9) WM capacity will be negatively correlated with approximation of randomness against the *unexploitable* opponent

10) Level of executive control will be positively correlated with approximation of randomness against the *unexploitable* opponent

2.3.2 Method.

2.3.2.1 Participants. Forty subjects (N = 40; 28 female; M_{age} = 22.95, SD_{age} = 5.53) from the University of Sussex participant pool were recruited. Participants received a flat £10 reward for their participation and an average of £2 extra for performance (see Design and Procedure, section 2.3.2.3). Informed consent was obtained from all participants before testing, and the experiment was approved by the Sciences & Technology Research Ethics Committee (C-REC) at the University of Sussex (ER/JS753/2).

2.3.2.2 Materials.

2.3.2.2.1 Game trials. The stimuli and experimental set-up were identical to Experiment 1.

2.3.2.2.2 Working memory task. I used a short, modified version of an operation span (OSPAN) task (after Turner & Engle, 1989; Unsworth et al., 2005) to asses participants' working memory. The Presentation version of the task I used was obtained from Neurobehavioral Systems online resources (Neurobehavioral Systems, n.d., b). In the task, participants had to solve equations while memorizing a string of letters, with the list length (number of letters to recall) increasing gradually, starting at 2. The letters used were similar to those used by Unsworth, Heitz & Engle (2005) in their automated OSPAN task. Each trial started with an equation such as "5 + (1*2) = 7" presented on screen. Participants had to indicate whether the equation presented was correct or incorrect using the left and right arrow keys, respectively. After the participant had given their response, a letter was immediately presented on screen for 1000ms, after which the next equation was presented; this continued until the list length at the current trial was reached, at which point the participant had to recall the string of letters in the correct order (responding using the keyboard). Participants completed three trials at a list length and proceeded to the next set of three trials at a longer list length, only if they had recalled the string correctly on at least two of the three trials. Participants were instructed to not guess the answers to the equation tasks and to try and be sure they get the answers right. Each participant completed a short training phase with a minimum of two trials at a list length of two; the training trials would repeat until the participant had at least 50% correct recall.

2.3.2.2.3 Executive control task. I used an arrow flanker task (based on Eriksen & Eriksen, 1974) to measure executive control. The Presentation version of this task I used was also obtained from Neurobehavioral Systems online resources (Neurobehavioral Systems, n.d., a). Participants used two keys on a keyboard to respond to a central arrow, pressing the left control key with their left index finger for arrows pointing left and the right control key with their right index finger for arrows pointing right. The central arrow (< or >) appeared directly at the location of a central fixation cross that was presented before each trial and was flanked by three arrows both to the left and right. The fixation cross was presented for 200ms, after which the central arrow and flankers on both sides of the central arrow were presented for 200ms. Each flanker was either a right- or left-facing arrow of the same size as the central arrow, approximately 3.4cm x 3.4cm. In congruent trials, each flanker was facing the same direction as the central arrow; in incongruent trials, some of the flankers were facing the opposite direction. After the stimulus presentation, participants had 500ms to respond before the initiation of the next trial. Participants completed a total of 96 trials, divided into two blocks of 48 trials, with an even number of left and right central arrows and congruent and incongruent flankers in both blocks.

2.3.2.2.4 Questionnaires. I administered a short questionnaire following the completion of each RPS block to assess participants' interpretation of the opponent's behaviour and their own success in the game. Participants first indicated their attribution of their game results in the block to luck or skill by clicking on a point on a slider ranging from "100% luck" to "100% skill". The middle-point of this slider was scored as 0, with answers indicating more luck scored with negative and answers indicating more skill with positive values. Participants also responded to a short questionnaire about their interpretation of the opponents' behaviour. This questionnaire was based on participants' descriptions of the opponents' strategies in Experiment 1 (see Appendix 2 for questionnaire and descriptive statistics).

2.3.2.3 Design and procedure. The general experimental design was identical to that of Experiment 1, with a 2x2 design with value *(high, low)* and opponent strategy *(unexploitable, exploitable)* as factors. In the *high value* condition, the participants gained 10p for each point in the summed final score of the two *high value* blocks. If this summed final score was negative due to a high number of losses, the participants received no extra cash, and only received the £10 that was the baseline participation reward for all participants. The *low value* condition and the *unexploitable* opponent condition were both identical to Experiment 1. The *exploitable* opponents followed a self-upgrade rule for 70% (63) of the trials, making random choices from a flat distribution without replacement for the remaining 30% (27) trials, as in Experiment 1.

At the beginning of each block, the experimental program informed participants whether they would be playing for money or points. Regardless of the strategy condition of the block, participants were also informed that their opponent would play in a certain way that would be revealed to them at the end of the experiment. Participants were instructed to try and win as many rounds as possible regardless of whether they would be playing for money or for points. The procedure of the game trials was identical to that of Experiment 1, with a confidence measure on every 9th trial. At the end of each block, participants completed the short skill / luck balance questionnaire and the short questionnaire about the opponent's behaviour. After all blocks, the participants completed the flanker and automated OSPAN tasks, after which I debriefed them and paid them their participation reward plus any money they earned through success in the game.

2.3.3 Results.

2.3.3.1 Behavioural measures.

2.3.3.1.1 Item selection and outcome at trial n. I analysed proportions of item selection at trial *n* in each block as in Experiment 1 (see Table 2.5; see section 2.4 for further discussion on the methodological choice to use ANOVAs for proportion data.). There was a significant main effect of item [$F(2, 78) = 5.94$, MSE = .01, $p = .004$, $\eta_p^2 = .13$], no two-way interaction between item choice and strategy [$F(2, 78) = 0.22$, MSE < .01, $p = .804$, $\eta_p^2 = .01$] or item choice and value [$F(2, 78) = 2.64$, MSE = .01, $p = .078$, $\eta_p^2 = .06$], and no three-way interaction [$F(2, 78) = 0.23$, MSE < .01, $p = .792$, $\eta_p^2 = .01$]. Participants were more likely to choose *rock* than *paper* or *scissors* (35.72%, 32.53% and 31.76%, respectively; Tukey's HSD, $p < .05$ for both comparisons), in line with previous results (e.g. Xu, Zhou & Wang, 2013; Wang, Xu & Zhou, 2014; Dyson et al., 2016) and also the non-significant trend in Experiment 1.

I conducted the same analysis for outcome at trial *n* (see Table 2.5). Here, there was a significant main effect of outcome [$F(1.18, 46.13) = 28.27$, MSE = .04, $p < .001$, $\eta_p^2 =$

.42], a significant interaction between opponent and outcome [$F(1.28, 49.85) = 25.58$, $MSE = .03$, $p < .001$, $\eta_p^2 = .40$], no significant interaction between value and outcome [$F(2, 78) = 0.58$, $MSE = .01$, $p = .564$, $\eta_p^2 = .02$] and no significant three-way interaction [$F(1.55, 60.52) = 0.73$, $MSE = .01$, $p = .454$, $\eta_p^2 = .02$]. As in Experiment 1, the only factor behind the differences between the distributions of wins, losses and draws was whether the opponent was *exploitable* or not. As in Experiment 1, the distribution of wins, losses and draws was roughly uniform against the *unexploitable* opponent (33.8%, 34.2% and 32.1%, respectively) and there was a significant majority of wins against the *exploitable* opponent (47.4%, 26.2% and 26.4%, respectively), indicating that the participants on average succeeded in exploiting the opponent. Overall, the item selection and outcome results replicate Experiment 1.

Table 2.5. Proportions of outcomes and item choices in Experiment 2 (estimated marginal means).

	Unexploitable opponent			*Exploitable opponent*		
	Rock	*Paper*	*Scissors*	*Rock*	*Paper*	*Scissors*
High value	37.1%	32.4%	30.6%	36.4%	32.6%	31.0%
	(1.1%)	(1.5%)	(1.0%)	(0.9%)	(0.9%)	(0.7%)
Low value	34.6%	33.0%	32.4%	34.8%	32.1%	33.1%
	(1.1%)	(1.3%)	(1.5%)	(0.9%)	(0.9%)	(0.8%)
	Win	*Lose*	*Draw*	*Win*	*Lose*	*Draw*
High value	33.2%	34.7%	32.2%	47.8%	26.8%	25.4%
	(0.9%)	(1.0%)	(0.8%)	(2.8%)	(1.6%)	(1.5%)
Low value	34.3%	33.7%	32.0%	46.9%	25.7%	27.4%
	(0.7%)	(0.8%)	(0.8%)	(2.6%)	(1.5%)	(1.7%)

Note: standard error in parentheses.

2.3.3.1.2 First-order repetition effects. I analysed proportion data as in Experiment 1 (see section 2.4 for further discussion on the methodological choice to use ANOVAs for proportion data). See Table 2.6 for descriptives.

Table 2.6. Proportions of choice types in Experiment 2 (estimated marginal means).

	Low value			*High value*		
	Unexploitable			*Unexploitable*		
	Win	*Lose*	*Draw*	*Win*	*Lose*	*Draw*
Stay	34.6% (3.0%)	33.4% (2.7)	37.0% (2.4%)	35.0% (3.5%)	32.3% (2.5%)	39.7% (2.2%)
Upgrade	34.5% (2.4)	29.6% (2.2%)	34.0% (1.9%)	35.5% (2.9%)	29.1% (2.0%)	28.8% (1.5%)
Downgrade	30.9% (2.3%)	37.0 (2.4%)	29.0% (2.3%)	29.5% (2.5%)	38.5% (2.2%)	31.5% (1.9%)
	Exploitable			*Exploitable*		
	Win	*Lose*	*Draw*	*Win*	*Lose*	*Draw*
Stay	21.3% (3.5%)	44.9% (3.4%)	23.0% (2.3%)	20.2% (3.4%)	47.4 (3.8%)	26.0% (2.9%)
Upgrade	57.7% (4.7%)	26.5% (2.1%)	28.9% (3.0%)	58.2% (4.9%)	24.5% (2.3%)	25.1% (2.4%)
Downgrade	21.0% (2.7%)	28.6% (2.4%)	48.2% (3.2%)	21.7% (2.8%)	28.0% (2.5%)	48.9% (3.6%)

Note: standard error in parentheses.

The main effect of player strategy at trial *n+1* was not significant [$F(1.48, 57.84) = 0.34$, MSE = .15, $p = .646$, $\eta_p^2 = .01$]. However, there were significant interactions between opponent and player strategy at trial *n+1* [$F(2, 78) = 9.91$, MSE = .03, $p < .001$, $\eta_p^2 = .20$]

and between outcome at trial *n* and player strategy at trial *n+1* [$F(2.34, 91.26) = 18.327$, $MSE = .14$, $p < .001$, $\eta_p^2 = .32$], indicating that player choices of *staying*, *upgrading* and *downgrading* were affected by the outcomes of previous trials as well as opponent exploitability, as expected based on Experiment 1. Further, there was a significant three-way interaction between opponent, outcome at trial n, and player strategy at trial *n+1* [$F(1.93, 75.26) = 25.14$, $MSE = .12$, $p < .001$, $\eta_p^2 = .39$], replicating Experiment 1. There was no significant interaction between value and player strategy at trial *n+1* [$F(1.67, 65.03) = 0.73$, $MSE = .04$, $p = .464$, $\eta_p^2 = .02$]. There was also no significant three-way interaction between opponent, value and player strategy at trial *n+1* [$F(2, 78) = 0.05$, $MSE = .03$, $p = .947$, $\eta_p^2 < .01$], no significant three-way interaction between value, outcome and player strategy at trial *n+1* [$F(3.34, 130.27) = 1.09$, $MSE = .02$, $p = .359$, $\eta_p^2 = .03$], and no four-way interaction [$F(2.51, 97.90) = 0.25$, $MSE = .04$, $p = .825$, $\eta_p^2 < .01$].

In the *unexploitable* condition, there were no significant differences between player strategy as a function of outcome (Tukey's HSD; all *p*'s > .05), consistent with Experiment 1. In the *exploitable* condition, the optimal choices of upgrade following a win, stay following a loss, and downgrade after a draw were all significantly more likely than either of the suboptimal choices for each outcome (Tukey's HSD; $p < .05$ for all comparisons). However, there were no significant differences in rate of optimal play between outcome types (Tukey's HSD; all *p*'s > .05). This was contrary to Experiment 1, where the rate of optimal play after losses was significantly lower than for wins or draws. See Figure 2.6 for the distribution of different strategies.

As an exploratory analysis, I categorized participants in the *exploitable* blocks as failed or successful similar to Experiment 1. Out of the 40 participants, 10 failed to reach a win-rate significantly higher than chance on both blocks, with a further 12 failing on one of the blocks (5 for the *high value* block and 7 for the *low value* block). Thus, there were 25 successful participants in the *high value* block and 23 in the *low value* block. As in Experiment 1, the trend in the data for the successful participants was similar to that of the whole sample (see Figure 2.7). The unsuccessful participants played essentially randomly, with no strategic learning observable after any outcome.

To assess whether exploiting the opponent was equally challenging in Experiments 1 and 2, I conducted a three-way mixed ANOVA on rates of optimal choices, with outcome *(win, lose, draw)* and value *(low, high)* entered as repeated measures factors and experiment entered as the grouping variable (see Table 2.7). There was a significant main effect of outcome [$F(2, 156) = 22.59$, $MSE = .06$, $p < .001$, $\eta_p^2 = .23$], no significant main effect of value [$F(1, 78) = 0.04$, $MSE = .04$, $p = .841$, $\eta_p^2 < .01$], andno significant main effect of experiment [$F(1, 78) = 0.02$, $MSE = .17$, $p = .889$, $\eta_p^2 < .001$]. The interaction between experiment and outcome was non-significant but marginal [$F(2, 156) = 2.86$, $MSE = .11$, $p = .061$, $\eta_p^2 = .04$]. The interaction between experiment and value [$F(2, 156) = 0.24$, $MSE = .04$, $p = .625$, $\eta_p^2 < .01$], and the three-way interaction between experiment, value and outcome [$F(2, 156) = 0.32$, $MSE = .02$, $p = .724$, $\eta_p^2 < .01$] were both non-significant. Participants made the most optimal choices after wins (60.3%), followed by draws (50.2%) and then losses (42.8%), with all pairwise comparisons being significant at the $p < .05$ level (Bonferroni).

However, interpreting the results is not straightforward due to the main interaction of interest, the interaction between experiment and outcome, being marginal ($p = .061$), combined with the fact that due to low per-experiment sample size, the between-subjects analysis is underpowered. At face value, the non-significant interaction would suggest that the difference in effects between the experiments was itself not significant (see Gelman &

Stern, 2012), and the rate of optimal choices for the outcomes was similar for the two experiments. However, looking at specifically the rates of optimal choices following losses, there is a clear numerical trend of higher rates of optimal choices in Experiment 2 (see Table 2.7). Moreover, looking at only the participants who had an above-chance win-rate in Experiments 1 and 2 (compare Figures 2.3 and 2.7), successful participants in Experiment 2 seem to have been slightly more successful after losses than successful participants in Experiment 1.

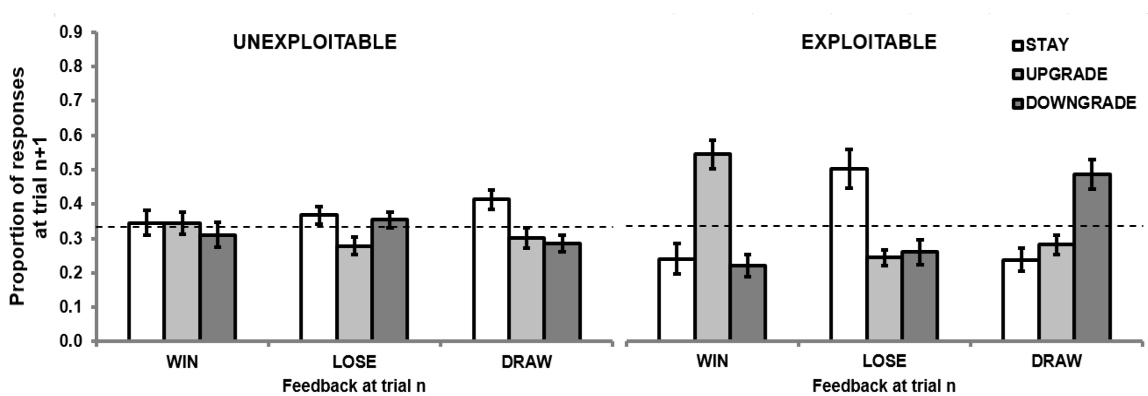

Figure 2.6. Distributions of participants' strategic choices collapsed across the value conditions in Experiment 2. Error bars represent SEs. Dashed line represents chance-level responding.

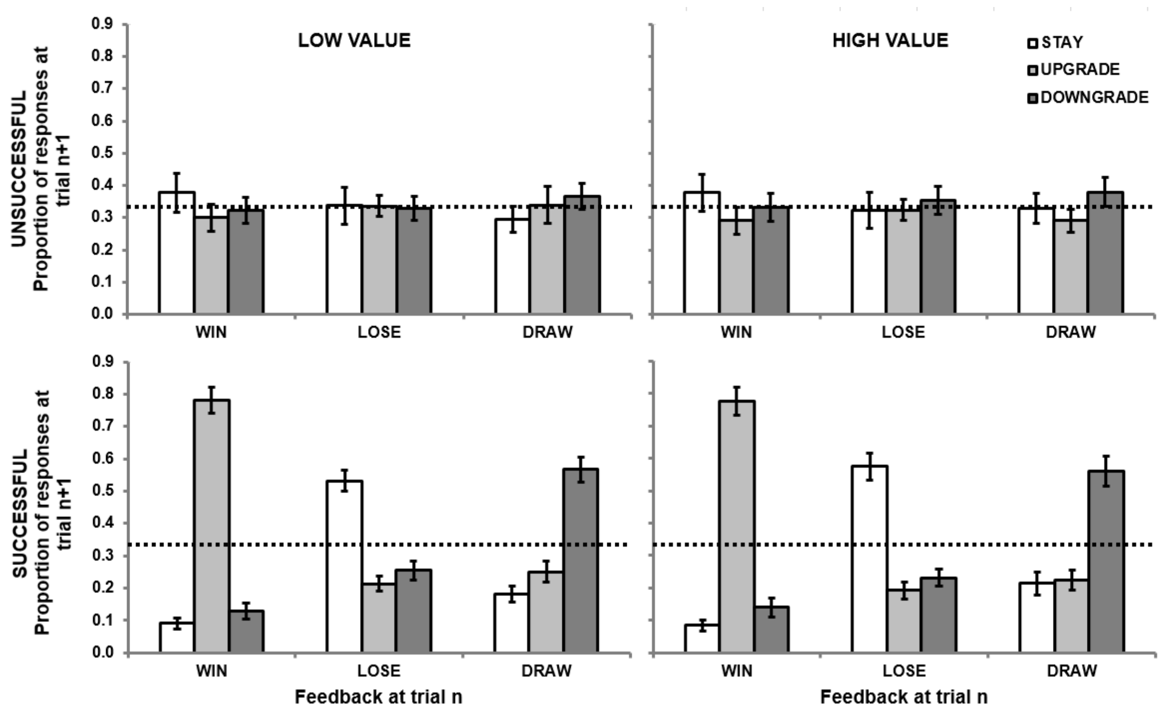

Figure 2.7. Distributions of unsuccessful and successful participants' strategic choices in the *exploitable* condition in Experiment 2. Error bars represent SEs. Dashed line represents chance-level responding.

Table 2.7. Mean proportions of optimal play after different outcomes in Experiments 1 and 2

	Experiment 1			*Experiment 2*		
	Win	*Lose*	*Draw*	*Win*	*Lose*	*Draw*
High value	63.6%	39.4%	50.3%	58.2%	47.4%	48.9%
	(4.6%)	(3.4%)	(3.7%)	(4.6%)	(3.4%)	(3.7%)
Low value	61.8%	39.6	53.5%	57.7%	44.9%	48.2%
	(4.7%)	(3.2%)	(3.4%)	(4.7%)	(3.2%)	(3.4%)

Note: standard error in parentheses.

2.3.3.1.3 Reaction time analysis. I analysed median reaction times for decisions made after wins, losses and draws as in Experiment 1, with a three-way repeated-measures ANOVA (see Table 2.8). Seven participants were excluded from this analysis due to having at least one average median reaction time (averaged across wins, losses and draws) that was at least twice the block average median (averaged across all participants in that block; after Forder & Dyson, 2016). Degrees of freedom were corrected using Greenhouse-Geisser estimates whenever Mauchly's test indicated a violation of sphericity.

Table 2.8. Average median reaction times (milliseconds) in Experiment 2, N = 33

	Unexploitable opponent			*Exploitable opponent*		
	Win	*Lose*	*Draw*	*Win*	*Lose*	*Draw*
High value	964 (116)	748 (81)	689 (65)	862 (144)	979 (126)	845 (102)
Low value	759 (73)	601 (61)	532 (46)	759 (105)	867 (101)	632 (63)

Note: standard error in parentheses.

There was a significant main effect of outcome [F(1.33, 42.49) = 3.76, MSE = 377965, p = .048, η_p^2 = .11] and of value [F(1, 32) = 6.71, MSE = 361628, p = .014, η_p^2 = .17], but no significant main effect of opponent [F(1, 32) = 4.05, MSE = 286353, p = .053, η_p^2 = .11]. The interaction between opponent and outcome was significant [F(1.31, 42.03) = 4.13, MSE = 274917, p = .038, η_p^2 = .11]. The interaction between value and opponent [F(1, 32) = .04, MSE = 469004, p = .848, $\eta_p^2 < .01$], the interaction between value and outcome [F(1.56, 49.88) = .19, MSE = 175362, p = .774, $\eta_p^2 < .01$], and the three-way interaction between opponent, value and outcome [F(1.60, 51.28) = .33, MSE = 198782, p = .673, η_p^2 = .01] were non-significant.

As in Experiment 1, the reaction times after losses in the *exploitable* condition were significantly slower than in the *unexploitable* condition (923 ms and 675 ms, respectively; Tukey's HSD; $p < .05$). Within the *unexploitable* condition, reaction times for wins were slower than reaction times for draws (862 ms and 611 ms, respectively; Tukey's HSD; $p < .05$), replicating Forder & Dyson (2016). However, this was not the case within the *exploitable* condition, where there were no significant differences between outcomes (see Figure 2.8). Contrary to Experiment 1, reaction times in the *high value* condition were overall higher than reaction times in the *low value* condition.

As in Experiment 1, to further explore the relationship between reaction times and performance, I plotted individual win-rates (averaged across the two *exploitable* blocks) against the reaction time difference between decisions made after losses and wins in the *exploitable* condition. A significant positive correlation ($r = .633$, $p < .001$; see Figure 2.9) indicates that reaction times after losses in relation to wins increased as a function of an individual participant's win-rate, replicating Experiment 1.

To address the differences in results between the two experiments, I ran a cross-experiment comparison on RT similar to the cross-experiment comparison on rates of optimal choices. I entered outcome *(win, lose, draw)*, value *(low, high)* and opponent *(unexploitable, exploitable)* as repeated measures factors and experiment as the between-groups factor in a four-way mixed ANOVA. Here, there was a significant main effect of experiment [$F(1, 61) = 5.12$, $MSE = 784392$, $p = .027$, $\eta_p^2 = .08$]. However, there were no significant interactions between experiment and any of the other factors: outcome [$F(1.53, 93.48) = 0.08$, $MSE = 239002$, $p = .880$, $\eta_p^2 < .01$]; value [$F(1, 61) = 1.68$, $MSE = 285973$, $p = .200$, $\eta_p^2 = .03$]; opponent [$F(1, 61) = 0.01$, $MSE = 235117$, $p = .913$, $\eta_p^2 = .00$]. The three-way interactions between experiment, outcome, and opponent [$F(1.45, 88.70) = 1.13$, MSE = 194929, p = .312, η_p^2 = .02]; experiment, outcome, and value [$F(1.79, 109.32)$ = 1.21, MSE = 115690, p = .298, η_p^2 = .02], and experiment, value, and opponent [$F(1, 61)$ = 1.92, MSE = 448925, p = .170, η_p^2 = .03] were all non-significant. There was also no significant four-way interaction [$F(2, 122)$ = 0.43, MSE = 139787, p = .654, η_p^2 < .01]. This suggests that there was a slight overall reaction time difference between the two experiments but that this was not driven by different RTs in any specific condition, after any specific game outcomes, or by differences in other interactions between the two experiments. Participants had overall slower mean reaction times in Experiment 2 (770 ms) compared to Experiment 1 (624 ms), but it is not clear why this was so: this may simply be an effect of slightly different participant samples.

For the within-subjects factors, there was a main effect of opponent [$F(1, 61)$ = 8.74, MSE = 235117, p = .004, η_p^2 = .13], value [$F(1, 61)$ = 7.43, MSE = 235117, p = .005, η_p^2 = .11], and outcome [$F(1.53, 93.48)$ = 11.38, MSE = 239002, p < .001, η_p^2 = .16]. There were no significant interactions between value and outcome [$F(11.79, 109.32)$ = 0.30, MSE = 115690, p = .720, η_p^2 < .01], value and opponent [$F(1, 61)$ = 1.24, MSE = 448925, p = .270, η_p^2 = .02], or a three-way interaction between the within-subjects factors [$F(2, 122)$ = 0.05, MSE = 139787, p = .952, η_p^2 < .01]. . Participants in the two experiments had overall slower mean reaction times under *high value* (750ms) compared to *low value* (644ms) conditions. The only significant interaction was between opponent and outcome [$F(1.45, 88.70)$ = 6.65, MSE = 178066, p = .005, η_p^2 = .10]. Lose reaction times were significantly slower against the *exploitable* (848 ms) compared to the *unexploitable* (617 ms) opponent (Tukey's HSD, p > .05). Win reaction times in the *unexploitable* condition were higher than both draw and lose reaction times (Tukey's HSD, p > .05 for all comparisons).

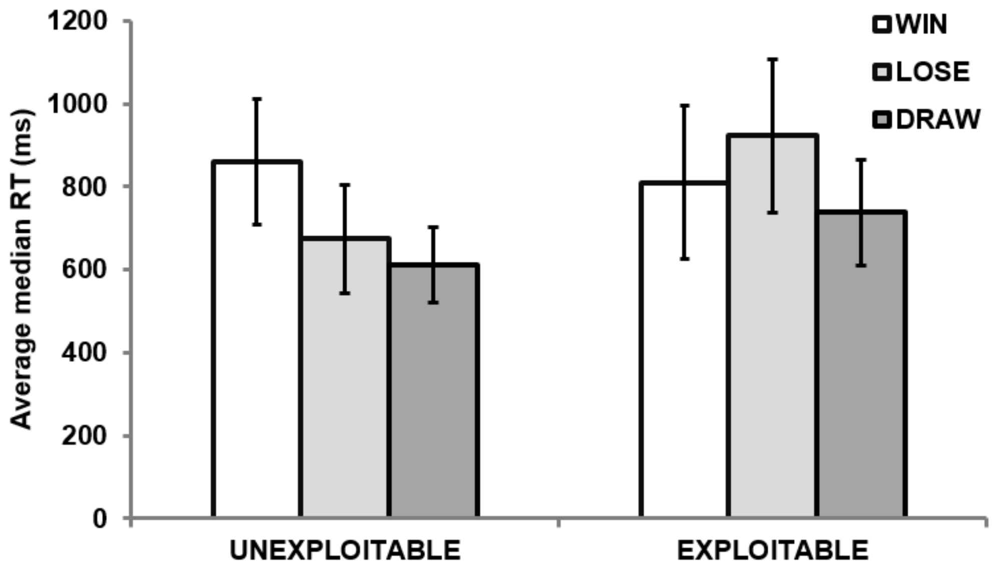

Figure 2.8. Average median reaction times collapsed across the value conditions in Experiment 2. Error bars represent 95% CIs.

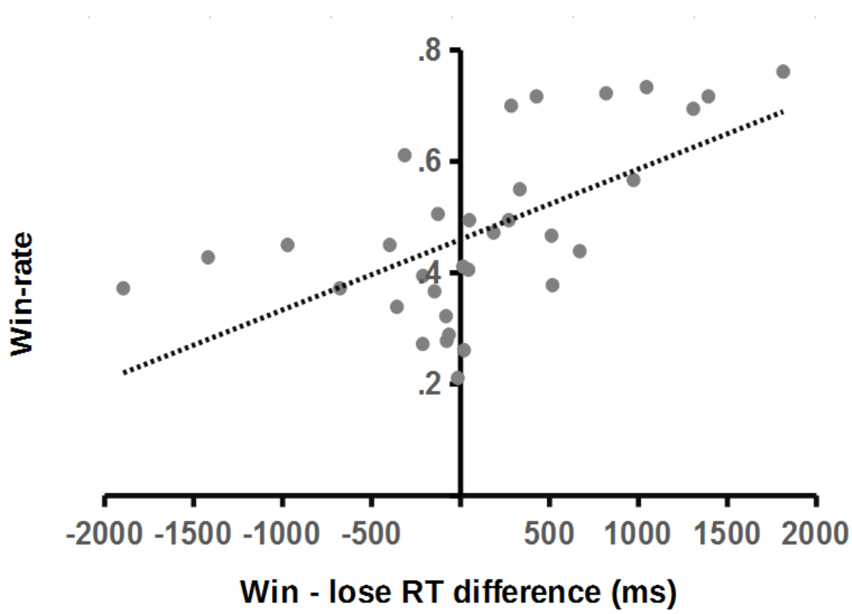

Figure 2.9. Correlation between win-rate and difference between average win and lose reaction times in the *exploitable* conditions in Experiment 2. A positive RT difference indicates slower post-error than post-success RTs.

2.3.3.1.4 Mean confidence measure. I analysed mean confidence rates were analysed using a two-way repeated measures ANOVA with opponent *(exploitable, unexploitable)* and value *(high, low)* conditions entered as factors (see Table 2.9). There was no significant main effect of strategy [$F(1, 39) = 1.68$, $MSE = .526$, $p = .202$, $\eta_p^2 = .04$], no main effect of the value condition [$F(1, 39) = 0.94$, $MSE = .16$, $p = .338$, $\eta_p^2 = .02$] and no two-way interaction [$F(1, 39) = 0.75$, $MSE = .21$, $p = .387$, $\eta_p^2 = .02$]. Unlike in Experiment 1, the participants' confidence seems to not have varied between the *unexploitable* and *exploitable* conditions, but the numerical trend was in the same direction as in Experiment 1 (see Table 2.9).

I examined the relationship between win-rate and player confidence as in Experiment 1, by calculating Fisher transformed correlation coefficients for reported confidence and the proportion of wins on the eight trials before each confidence measurement, individually for each player in each block (see Table 2.9). I then examined the variance in the strength of this association using a two-way repeated measures ANOVA with the Fisher transformed coefficients as the dependent variable, and opponent *(unexploitable, exploitable)* and value *(high, low)* entered as factors. Eight participants had to be excluded due to having no variance in their reported confidence on one or more blocks, making it impossible to calculate a correlation. There was no significant main effect of opponent [$F(1, 31) = 0.21$, $MSE = .20$, $p = .650$, $\eta_p^2 < .01$], no significant main effect of value [$F(1, 31) = 0.15$, $MSE = .09$, $p = .706$, $\eta_p^2 < .01$], and no significant interaction [$F(1, 31) = 0.89$, $MSE = .17$, $p = .353$, $\eta_p^2 = .03$]. The results thus differ from Experiment 1, where there was a significant effect of opponent on the correlation between confidence and win-rate, with participants' confidence being more highly correlated with win-rates against the exploitable opponent. Note also that the correlations between confidence and win-rate in Experiment 2 were essentailly zero, unlike in Experiment 1 (see Tables 2.3 and 2.9).

Table 2.9. Mean confidence measure (Likert, 1-5), N = 40, and correlation between confidence and win-rate in Experiment 2, N = 33

Mean confidence (range: 1 – 5)		
	Unexploitable opponent	*Exploitable opponent*
High value	2.83 (.10)	3.04 (.13)
Low value	2.96 (.07)	3.04 (.13)
Mean confidence / win-rate correlation (Fisher transformed z-scores)		
	Unexploitable opponent	*Exploitable opponent*
High value	-.00 (.07)	-.11 (.09)
	[-.00]	[-.006]
Low value	-.05 (.06)	-.02 (.06)
	[-.05]	[-.02]

Note: standard error in parentheses. Inverse Fisher transformed z values in brackets.

2.3.3.2 Other measures.

2.3.3.1.4 Perception of luck vs. skill. I analysed participants' ratings of their attribution of game results to luck or skill using a two-way repeated measures ANOVA with opponent *(exploitable, unexploitable)* and value *(high, low)* entered as factors. There was a significant main effect of opponent [$F(1, 39) = 28.97$, $MSE = 2695.83$, $p < .001$, $\eta_p^2 = .43$]. There was no significant main effect of value [$F(1, 39) = 0.42$, $MSE = 1626.51$, $p = .520$, $\eta_p^2 = .01$], and no significant an interaction effect [$F(1, 39) = 0.10$, $MSE = 1679.89$, $p = .757$, $\eta_p^2 < .01$]. Players gave responses indicating significantly more skill than luck driving the results when playing against the *exploitable* opponents (M = 26.7, SE = 7.4) than when playing against *unexploitable* opponents (M = -17.5, SE = 6.6).

The correlations between the luck/skill measure and average confidence were r = .08, p = .618 *(unexploitable, low value)*, r = .13, p = .420 *(unexploitable, high value)*, r = .49, p < .001 *(exploitable, low value)* and r = .26, p = .109 *(exploitable, high value)*, suggesting that the average confidence measure and the luck/skill measure mostly reflect different things.

2.3.3.1.5 Working memory and optimal choices. I analysed the effect of working memory span on the rate of optimal choices against the *exploitable* opponent at trial n+1. The span variable used was the last list length where the participant had correct recall on at least two of the three trials for that span length. Three participants were excluded due to an overall accuracy less than 85% on the distractor task in the OSPAN (as per Unsworth et al., 2005). A further three participants were excluded due to having failed the memory span task on the first trial. For the remaining thirty-four participants, I entered the rate of optimal choices on trial *n+1* into a two-way repeated measures ANCOVA with value *(high, low)* and outcome at trial *n (win, lose, draw)* entered as factors and working memory span entered as the covariate (mean-centered). This analysis again followed the suggestion of Schneider et al. (2015) to use ANCOVAs in within-subjects designs to only examine the main and interaction effects of the covariate and not to correct for the covariate (see also Section 2.2.3.2.5). There was no significant main effect of memory span [$F(1, 32) = 0.88$, MSE = .18, $p = .355$, $\eta_p^2 = .03$], no significant interaction between memory span and value [$F(1, 32) = 1.61$, MSE = .05, $p = .214$, $\eta_p^2 = .05$], no significant interaction between memory span and outcome [$F(2, 64) = 0.85$, MSE = .05, $p = .431$, $\eta_p^2 = .03$] and no significant three-way interaction [$F(2, 64) = 0.14$, MSE = .02, $p = .868$, $\eta_p^2 < .01$]. See

Figure 2.10 for the relationship between memory span and optimal choices collapsed across outcome type.

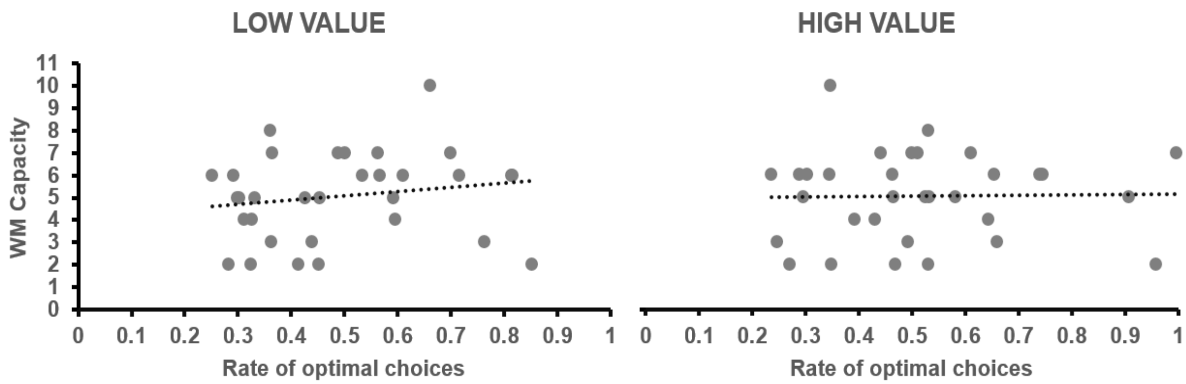

Figure 2.10. Relationship between WM capacity and optimal choices for the *low and high value exploitable* blocks in Experiment 2.

2.3.3.1.6 Deviation from randomness. I analysed the effect of executive control on deviation from randomness when playing against the *unexploitable* opponent. The executive control variable was the difference between the median reaction times for the congruent and incongruent flanker trials. I calculated deviation from randomness by deducting the rate of each move type (stay, upgrade, downgrade) from 33.3%. I entered this variable as the dependent variable into a two-way repeated measures ANCOVA with value *(high, low)* and outcome (win, lose, draw) entered as factors and the executive control variable entered as the covariate (mean-centered). This ANCOVA was again conducted per the recommendations of Schneider et al. (2015), to examine only the main effect and interactions of the covariate. Executive control had no significant main effect [$F(1, 38) = 2.41$, $MSE = .01$, $p = .129$, $\eta_p^2 = .06$], no significant interaction with value [F(1,

38) = 1.39, MSE < .01, p = .245, η_p^2 = .04], no significant interaction with outcome [F(1, 38) = 0.83, MSE < .01, p = .439, η_p^2 = .02], and no significant three-way interaction [F(2, 76) = 0.28, MSE < .01, p = .756, η_p^2 < .01].

I then ran the same ANCOVA using working memory span as the covariate (mean-centered). Six participants were excluded from the analysis (see above). There was no significant main effect of memory span [F(1, 32) = 0.08, MSE = .01, p = .782, η_p^2 < .01], no significant interaction between memory span and value [F(1, 32) = 0.13, MSE < .01, p = .717, η_p^2 < .01], no significant interaction between memory span and outcome [F(2, 64) = 0.29, MSE < .01, p = .749, η_p^2 = .02] and no three-way interaction effect [F(2, 64) = 0.70, MSE < .01, p = .500, η_p^2 = .02].

2.4 General Discussion

In Experiments 1 and 2, I failed to replicate the *win-stay, lose-shift* effect observed in previous studies of people playing RPS against an *unexploitable* opponent (Dyson et al., 2016; Forder & Dyson, 2016) and in previous studies using different binary choice paradigms (Scheibehenne et al., 2011; Wilke et al., 2014; Achtziger et al., 2015). Remarkably, participants managed to do this even without any attempts at exploitation from the computer opponent's side. That is, they would not experience more losses if they played predictably. Rapoport and Budescu (1992) found that in a competitive binary choice game between two human players, players can achieve a mixed strategy but this likely has to do with the dynamics between two players who would most likely attempt to exploit each other (see also West & Lebiere, 2001). In my experiments, it seems that no such mutual attempts at exploitation were needed to rid participants of reinforcement errors.

Carryover effects are a possible candidate for explaining the non-replication. As the optimal strategy against the *exploitable* opponent in both experiments involved choices in conflict with reinforcement, it may be that this helped participants play according to mixed strategy against the *unexploitable* opponent and avoid reinforcement errors. However, this explanation leaves open the bigger question of how participants were able to learn to avoid reinforcement errors at all, as they seem quite robust. For example, Scheibehenne et al. (2011) showed that when participants were able to examine the patterns produced by two slot machines before making their choices, they still made *stayshift* decisions even when they could have achieved above-chance win-rates by following a *shiftstay* rule.

Experiment 1 also showed that performance was non-optimal following loss against *exploitable* opponents. The motivation for changing the *exploitable* opponent's strategy in Experiment 2 was to assess whether the suboptimal performance against the *exploitable* opponent on trials immediately following a loss in Experiment 1 was due to losses themselves, or due to the specific optimal response in Experiment 1 being unintuitive or hard to learn. The finding that participants in Experiment 1 made more optimal choices after both wins and draws compared to losses was surprising given that the losses were the only case where the optimal choice aligned with reinforcement. Additionally, for a successful participant, wins were more likely than draws or losses, but draws and losses were both equally likely. This should theoretically have allowed the participants to learn what to do after draws and losses equally well unless the optimal choice for one of these outcomes was in one way or another more difficult to make, or unless the outcome of loss had some specific effect on decision making. However, the optimal *stay* response after draws in Experiment 1, while in conflict with reinforcement, was aligned with inertia, which may make choosing the option more likely (Alós-Ferrer et al., 2016; Alós-Ferrer & Ritschel, 2018). To disentangle the effects of the non-win outcomes themselves from the effects of inertia and reinforcement, in Experiment 2, the *stay* response was the optimal response to be made after a loss against the *exploitable* opponent. Consequently, the *exploitable* opponent in Experiment 2 played according to a *self-upgrade* rule (contrast to the *self-downgrade* rule in Experiment 1; see Figure 1), leading to a similar cyclical pattern as in Experiment 1.

Analysing the choice behaviour in Experiment 2 on its own, there were no statistically significant differences in the rates of optimal play after different outcomes. although the trend was similar to that in Experiment 1, with most optimal responses made after wins, followed by draws and then losses. At first blush, this would appear different from the pattern of data in Experiment 1 that suggests that both the nature of the outcome and the type of the subsequent strategy contributed to performance. While a cross-experiment comparison strictly speaking yielded no significant differences in rates of optimal play as a function of outcome between the two experiments, the interaction between experiment and outcome was marginal ($F > 2.8$). Moreover, looking only at the participants with above-chance win-rates in both experiments (see Figures 2.3 and 2.7), it seems that successful participants in Experiment 2 were better able to make optimal decisions after losses than participants in Experiment 1. This effect may not have reached significance due to low power compounded by the fact that participants could be clearly split into two sub-groups of successful and unsuccessful learners in both experiments.

Despite problems in interpreting the results, it seems that reinforcement did not affect performance against the exploitable opponents in expected ways, and this is regardless of whether the difference between experiments in optimal choices following losses is a true effect or not. The participants made optimal choices most likely after wins in Experiment 1, where the optimal choice after wins did not align with reinforcement. In Experiment 2, there was no overall difference in the rate of optimal choices following different outcomes, i.e. success on average was not dependent on whether the optimal strategy aligned with reinforcement., Further, assuming the marginal difference between the experiments in rates of optimal behavioural after losses is a true effect, it would seem that participants' choices were *more optimal* when the optimal choice after a loss was contrary to reinforcement and in alignment with inertia than when it aligned with reinforcement. If this effect is assumed to be null, it would similarly follow that whether the optimal choice aligned with reinforcement did not matter in terms of learning.

The trend of more successful performance after losses in Experiment 2 suggests that alignment with inertia may have helped participants apply the optimal decision rule (similarly to e.g. Alós-Ferrer & Ritschel, 2018; Scheibehenne et al., 2011). Experiment 1 finding more optimal decisions following draws than losses, when the optimal choice after draws aligned with inertia and the optimal choice after losses did not, also lends support to this interpretation. This interpretation would also suggest inertia being a stronger driver of decisions than reinforcement, as the lower rates of success in both of the aforementioned comparisons were in situations where the optimal choice aligned with reinforcement. However, the results of Experiment 2 in and of themselves do not fully support this interpretation, as there were no overall differences in rates of optimal choices after any outcome type. Moreover, looking only at the successful participants in Experiment 2, the rate of optimal choices after draws and losses is similar despite differences in alignment with inertia and reinforcement. Thus, it may be that the type of outcome still matters in addition to whether the optimal choice aligns with inertia or reinforcement, with draws leading to less of a difference than losses.

In general, nothing in the behavioural results suggests a stayshift bias against either the unexploitable or the exploitable opponent; even the participants who were unssuccessful in exploiting the exploitable opponent were on average playing randomly rather than mistakenly following reinforcement and thus losing more. This result is puzzling: how and why did participants not make the kinds of reinforcement-biased *stayshift* choices I would have expected? One could argue that the task structure of the experiments may have made reinforcement errors in general unlikely. As the *exploitable* opponent played according to a set of rules that would lead it to play according to a repeating, cyclical pattern 70% of the time, participants may have noticed the pattern and started acting in accordance to a higher-order model of the game. That is, becoming aware of the pattern may have allowed participants to avoid reinforcement errors due to them no longer focusing as much on single trials but a longer frame of several trials. Losses and draws may then have led to less optimal choices due to them being a signal to the participants that something about the pattern might have changed. However, this explanation does not hold for the *unexploitable* condition, where I saw no evidence of a *stayshift* bias either: taken at face value, the results would seem to imply that people avoid the *stayshift* bias in RPS..

A limitation inherent in the main behavioural analyses of Experiments 1 and 2 is the use of an ANOVA to analyse proportional response data. I conducted these analyses in line with analyses in the published literature, e.g. Dyson et al. (2016) and Forder and Dyson (2016), both of whom had similar designs with similar variables of interest. Further, Stöttinger et al. (2014) also analysed proportion data of RPS choices using ANOVAs, and Wilke and Barrett (2009), Wilke et al. (2014), and Scheibehenne et al. (2011) seem to have used an ANOVA or linear regression on at least some of their proportion data analyses. However, while this statistical choice may show up in peer-reviewed articles, it is not without issues. The use of ANOVAs on proportional data is problematic, as the dependent variable is bounded and due to the properties of proportional data, the variances are unlikely to be equal between two conditions if the proportions of those conditions are different, thus violating the assumptions of ANOVA (Jaeger, 2009). The choice I made to use an ANOVA was, primarily, a practical one. The ideal analysis for the kinds of comparisons I wished to make would have been a multinomial logistic generalized linear mixed model for repeated measures (see Jaeger, 2008). However, a usable multinomial model approach was not readily available for either SPSS or R after extensive search. For example, the lme4 package of R allowed for creating such a model but not directly comparing the proportions of three different categorical decision types, which would have been crucial for this chapter. As a beta regression would demand excluding participants who had a probability of 0 or 1 on any of the possible decision types in any condition, it would not have been optimal for the purposes of these studies, given the already low sample size.

Given these statistical limitations, comparisons of high probabilities against low probabilities in the analyses in Chapter 2 are subject to increased Type I error rates (Jaeger, 2008). A little over one third of all the different mean proportions of choice types in Experiments 1 and 2 (12 out of 36 means in Experiment 1, 15 out of 36 means in Experiment 2) fell below .3, considered the lower range of acceptable mean proportions in terms of maintaining homogeneity of variance (Jaeger, 2008). No mean probabilities in these analyses were above .7, considered the upper range (Jaeger, 2008). The majority of the proportions below .3 were in the exploitable conditions – understandably, as this is where the participants could learn an optimal rule and follow it. However, in the cross-experiment comparison of optimal choices as a function of the outcome of the previous trial, none of the sample proportions for any condition were outside of this range (see

Table 2.7). This is also true if comparing the rates of optimal choices after different outcome types in the two experiments separately. Therefore, I argue that the issue of heterogenous variances was less of a problem to the interpretation of these, the most crucial analyses. Looking at the standard errors of the variables, it also seems that at least in the aforementioned cases the analysis did not produce estimates of “impossible values” (i.e. a 95% CI that would go above 100% or below 0%). Thus, while the approach was not ideal, the important results are not uninterpretable. Moreover, whatever the issues caused by this approach, the results are at least comparable to the results of other published work using the same non-ideal approach to percentage data (see above).

In both experiments alone and in the cross-experiment analysis, manipulating the value of wins and losses in RPS failed to produce effects on choice behaviour or overall performance, whether value was manipulated as a difference in scoring (Experiment 1) or through a monetary incentive (Experiment 2). The cross-experiment comparison revealed a general effect of the value manipulation on reaction times, with overall slower reaction times in *high value* conditions, regardless of whether the value manipulation was score-based or monetary. It is unclear why participants would spend more time on their decisions in the *high value* conditions, but whatever the reason, this does not seem to have reflected on their behaviour otherwise. This is in contrast to Forder & Dyson (2016), who found changes in game choice behaviour in different non-monetary value conditions, but no effect of the value manipulation on reaction times. However, in Forder & Dyson (2016), the value manipulation emphasized either wins or losses: a greater point penalty for a loss, or a greater point reward for a win. *Win-stay* behaviour specifically was increased when the reward for a win was greater than the penalty for a loss (or vice versa). In Experiment 1, on the other hand, rewards and penalties were always of the same magnitude.

However, this difference between the design of Experiment 1 and Forder & Dyson (2016) fails to explain the non-effect on performance in Experiment 2, where value was manipulated with a financial incentive. Specifically, the participants could only gain but not lose, in the sense that a score below zero (indicating more than a third of trials lost) did not lead to a financial penalty, but a score above zero led to reward. A financial incentive should have led to an increased number of reinforcement errors (Achtziger et al., 2015). There are known issues with a blocked design such as the one employed in Experiments 1 and 2 when it comes to financial incentive, namely, a block with financial incentive can lead to a loss of motivation in a subsequent block without such incentive (Ma et al., 2014). However, in Experiment 2, one would assume this would have led to a more pronounced difference in performance between the value conditions when playing against the *exploitable* opponent, not a dilution of the effect. The counterbalancing of the conditions was so that no two *exploitable* or *unexploitable* blocks would be played immediately following each other, but there were no other constraints. The order of the value conditions was fully counterbalanced, and there was only one block of each of the four types *((high value, low value) x (exploitable, unexploitable))*. Thus, a dilution of the effect due to order effects would be possible only if, in addition to the effect highlighted by Ma et al. (2014), the opposite to that effect were *also* the case - that is, only if a lack of financial incentive in a block could lead to a loss of motivation in a subsequent block with financial incentive. Otherwise, a decrease in performance in the *low value exploitable* block after having already completed the *high value exploitable* block would show as an overall difference in performance between the value conditions in the predicted direction. A more likely explanation, then, is that given the task structure (see above), reinforcement errors in general unlikely, and thus the incentive could not increase them.

The working memory capacity measure used in Experiment 2 did not predict performance, contrary to my expectations. It may be that the learning occurring in the task follows a reinforcement learning rule without a need for working memory as such (see Collins & Frank, 2013, for an investigation of the relationship between working memory and reinforcement learning). Another option is that the participants learned the *exploitable* opponent's pattern in the simplest terms as a three-step sequence (Rock-Paper-Scissors), necessitating such a low working memory capacity that practically all participants were able to learn it if they were motivated enough or paid attention to the correct details in the game. The reason for individual differences in rates of optimal play against the *exploitable* opponent are still somewhat an open question. Moreover, neither working memory nor executive control predicted deviations from randomness in the *unexploitable* condition to a significant degree (contra Terhune & Brugger, 2011). Given the result that the participants were, on a group level, not significantly biased towards any type of move following any type of outcome, it may simply be that there wasn't much deviation for the covariates to predict.

Analysing the reaction time data for the experiments together revealed two overall significant findings. First, when playing against the *unexploitable* opponent, players reacted more slowly after wins compared to draws and losses. Second, reaction times after losses were slower when playing against the *exploitable* opponent than when playing against the *unexploitable* opponent. Lose reaction times also increased in relation to win reaction times as a function of individual win-rates against the *exploitable* opponent in both experiments, further indicating post-error slowing as a function of performance or perceived exploitability of the opponent. The lack of an overall significant post-error slowing trend in the *exploitable* condition is likely explained by the wide range of rates of successful exploitation between individuals. Together, these results suggest that post-error slowing (see Dutilh et al., 2012) or speeding (see Verbruggen et al., 2017) depends on success rate and/or perceived exploitability. When the win-rate is around chance level and the opponent cannot be exploited, participants were slower to make decisions after wins; when the win-rate increases and/or the opponent can be exploited, reaction times after losses slow down. Experiments 1 and 2 could not control for the confound between exploitability of opponent and frequency of losses, and thus the post-error slowing may be caused wholly or in part by losses being a rare event (see Dutilh et al., 2012) rather than perceived exploitation per se (I will directly explore this option in Chapter 3 using fixed rather than variable win-rates). Note, however, that reaction times after draws did not differ between the *exploitable* and *unexploitable* conditions, even though the aforementioned confound also applies to draws, suggesting that frequency alone is not enough to explain slowing in response to a non-win outcome. Participants in Experiment 2 also indicated, on average, believing that their outcomes against the *unexploitable* opponent were more due to luck, and that their outcomes against the *exploitable* opponent were more due to skill, implying a different understanding of the task structure between the conditions that could play a role in post-error slowing. Note that Verbruggen et al. (2017) found post-error speeding in an explicit gambling task, whereas much of the research finding post-error slowing (see e.g. Danielmeier & Ullsperger, 2011; Dutilh et al., 2012; Notebaert et al., 2009) have been conducted using tasks where the participants could reasonably believe are skill-dependent, such as different versions of a flanker task.

However, the increased reaction times after losses does not seem to have translated into more optimal choices after losses. Thus, longer decision times may be a necessary but not sufficient condition regarding the initiation of cognitive control and the revision of performance. The results may reflect surprise at a loss in the *exploitable* condition; losses in the *exploitable* condition are going to be a rarer event for the participants overall, and increasingly so as their success increases. Moreover, participants with a completely accurate model of the *exploitable* opponent's pattern would only experience losses in the rare (30%) number of cases where the opponent strays from their pattern and makes a random choice, leading to an even more unexpected loss - the participant may feel they "should have won" and start questioning their model of the opponent. The fact that participants performed quite poorly after losses despite increased reaction times in the *exploitable* condition suggests that they may have engaged in more exploratory behaviour (Wilson et al., 2014), potentially assuming the opponent has changed the pattern of play.

2.5 Conclusion

In two experiments, contrary to prior studies, participants exhibited no win-stay, lose-shift bias when playing RPS against an *unexploitable* opponent. Neither working memory nor executive control predicted individual level deviations from randomness against the *unexploitable* opponent, but given the overall result of no significant bias away from random play, it may simply be that there was no deviation to predict. In both experiments, participants were, on average, able to learn to exploit an opponent that played according to a simple cyclical pattern for 70% of the time, making random choices 30% of the time. Whether the optimal choices against the *exploitable* opponent were in conflict with reinforcement or not did not seem to affect the participants’ performance, contrary to prior literature. There were clear individual differences in rates of optimal play, although it is unclear why some participants failed to learn to exploit the *exploitable* opponent. Working memory did not predict exploitation, although this might simply be because the task itself was not very demanding to memory as such; the individual differences may stem from motivational factors. In both experiments, losses against the *exploitable* opponent led to slower reaction times than losses against the *unexploitable* opponent. Whether this is due simply to losses being a rare event against the *exploitable* opponent when the participant is exploiting successfully, or a combination of losses being rare and participants acknowledging that the opponent is *exploitable*, will be examined in Experiments 3 and 4.

CHAPTER 3: Reinforcement Biases, Confidence and Success

3.1 General Introduction

Contrary to the experimental hypotheses and previous research, Experiments 1 and 2 found no evidence of reinforcement biases in RPS, regardless of whether participants were playing against *unexploitable* (pseudorandom) or *exploitable* opponents. Likewise, there was no evidence of changes in biased choice behaviour as a function of outcome value, expressed either as points or money. Reaction time analyses revealed an interaction between opponent exploitability and the trends of post-error slowing and speeding, but interpretation of this trend was hindered due to the confound between the variable rate of learning between individuals and, hence, the frequency of outcomes. Experiments 3 and 4 looked more closely into reaction times and *stayshift* behaviour as a function of the order of outcomes, while controlling for individual variations in outcome frequency. Additionally, the experiments included a more thorough look into confidence as a function of fluctuating win-rates. The game task used was also changed from Rock, Paper, Scissors (RPS) to Matching Pennies (MP) to control for complexities in interpreting results from a game with an ambiguous outcome option (draw).

3.2 Experiment 3

3.2.1 Introduction. In Experiments 1 and 2 and in the context of RPS, participants had significantly slower reaction times for decisions following losses against the exploitable opponent compared to the unexploitable opponent, suggesting more cognitive control after losses in the condition where the players could actually affect their rate of wins. Further, post-error slowing when playing against exploitable opponents was predicted by the rate of successful exploitation, as measured by the win-rate of individual participants. Consequently, the participants with the highest rates of post-error slowing were also the participants with high win-rates (and thus a smaller number of losses). Contrary to this, playing against unexploitable opponents led to post-success slowing, with significantly higher reaction times for decisions made after wins compared to losses at the group level. As should be expected from the experimental design, participants also had lower rates of wins in the *unexploitable* condition. Even the highest individual win-rates achieved by chance in the *unexploitable* conditions (Experiment 1: 46.67%, Experiment 2: 46.67%) were much lower than the highest win-rates achieved in the *exploitable* conditions (Experiment 1: 78.89%, Experiment 2: 83.33%). Thus, Experiments 1 and 2 cannot dissociate the potential effects of infrequent outcomes leading to post-error or post-success slowing (the orienting account of post-error slowing; see Danielmeier & Ullsperger, 2011; Notebaert et al., 2009) from the effects of the exploitability itself. That is, when participants lost a round, they could have slowed down because they had learned a strategy that they expected would lead to a win, or they could have slowed down simply as an automatic orienting reaction to infrequent losses. To complicate matters, due to individual differences in learning, the range of win-rates in the *exploitable* condition (ranging from 21.11% to 78.89% in Experiment 1 and from 21.11% to 83.33% in Experiment 2) was much wider than in the *unexploitable* condition where wins were driven by chance (ranging from 22.22% to 46.67% in Experiment 1 and from 14.44% to 46.67% in Experiment 2). This added noise in the *exploitable* conditions makes a true group-level comparison between the conditions difficult. To address these issues, Experiment 3 shifted to a design with fixed win-rates in different success slopes (see Ejova et al., 2013; Thompson et al., 1998). That is, participants in Experiment 3 were exposed to different conditions of specified win-rate trajectories that would be similar for each participant regardless of how they played, making it possible to create conditions where win-rate and player skill were not co-dependent, and where within-subjects variation in win-rates was controlled.

Experiment 3 also moved away from using RPS as the experimental paradigm in favour of the slightly simpler Matching Pennies (MP) game. MP is a dichotomous choice task where two players both choose a side of a coin, with one player winning if the choices match and their opponent winning if they mismatch. This change was chosen due to several reasons. First, much of previous research showing consistent RL biases has been conducted using different variations of binary choice tasks (e.g. Scheibehenne et al., 2011; Wilke et al., 2011; Achtziger et al., 2015), however, these tasks have often included a financial incentive. Experiments 1 and 2 stemmed from the notion that reinforcement biases in RPS could be observed even without financial incentive (specifically *lose-shift*; Dyson et al., 2016; Forder & Dyson, 2016), but this result was not replicated. Furthermore, there were no RL biases with financial incentives (Experiment 2). As there are differences between the two game tasks (outlined below), replicating RL biases in MP is a first step towards understanding why Experiments 1 and 2 did not replicate previous findings. Second, due to the three-option nature of RPS, shifting in the game may be a more complex task for the decision-maker than staying. That is, in order to stay, one only needs to know what happened on the previous round and repeat that; in order to shift, one not only needs to know the specifics of the previous round but also choose between two directions of shifting. This was a key idea in Chapter 2. where I examined whether the ease of optimal responding following *win* was not a result of the positive outcome per se, but a result of the *stay* computation. Thus, interpretations of the relative frequency of shifting versus staying related to situations where shifts are more likely are not unequivocal. Since MP allows for only one type of shift, this issue is removed. Third, related to the former, adopting a true mixed strategy in RPS would lead to roughly equal proportions of both types of shifts, meaning a total proportion of shifts twice as large as the proportion of stays. In Experiments 1 and 2, the results were interpreted from the viewpoint that the differences between types of shifts are meaningful – the myopic best response, for example, is never simply any shift but a shift in a specific direction (see Dyson, 2019). As before, since MP allows for only one kind of shift, such assumptions are not needed. Fourth, RPS allows for ambiguous outcomes (draws) whereas MP does not. While the data regarding draws can shed light on whether behaviour is affected differently by ambiguous outcomes relative to other more transparently negative outcomes (i.e. losses), it comes at the cost of less power for analysing separate outcome types when the number of trials is limited. It also adds ambiguity in defining what the experienced win-rate of the participant is from the first-person perspective. That is, it is possible that draws are not treated as merely neutral outcomes in their own category, but e.g.as another type of loss (see Holroyd et al., 2006) or a type of "near-miss" (i.e. a win that just fell short; Dixon et al., 2013). This could lead to an experienced loss-rate significantly over chance level in a supposedly random situation of 1/3 wins, losses and draws each (i.e., 33.3% losses + 33.3% draws = 66.6% 'negative' outcomes). The results of Experiments 1 and 2 do not clearly support any single interpretation of draws and hint that the effect may be situational, similar to the effects of wins and losses. In Experiments 1 and 2, participants were generally more often able to apply the optimal strategy after draws than after losses, but the rate was still significantly lower from the rate of optimal choices after wins. Moreover, the reaction times for decisions following draws against the unexploitable opponent were similar to reaction times after losses against this opponent. However, the reaction times for losses differed between the two *exploitability* conditions, whereas the reaction times for draws did not.

These results cannot confirm any single interpretation of draws. In short, removing the ambiguous outcomes allows for better inferences about the effects of unambiguously positive and negative outcomes.

To address the specific issue of whether post-error slowing in a competitive game context simply follows from relatively infrequent losses in the case of successful exploitation, I included two conditions with overall above and below chance wins (hence *continuous success* and *continuous failure*, respectively). If the notion of post-error slowing being simply driven by the frequency of errors typically being low is correct (Danielmeier & Ullsperger, 2011), one would expect to see post-error slowing in conditions with a fixed above-chance win-rate (as in the *exploitable* conditions in Experiments 1 and 2) and no post-error slowing or even post-success slowing in conditions with a fixed below-chance win-rate (see Notebaert et al., 2009, for a manipulation of success rates in a visual decision task). Additionally, I included two conditions both with a chance-level (50%) overall win rate but with differently shifting success slopes: either a high win-rate followed by a low win-rate or vice versa (hence *descending* and *ascending*, respectively; see Figure 3.1 in section 3.2.2). These conditions were intended to roughly mimic either a situation where the player is at first exploiting an opponent who then “learns” to anticipate the player’s moves (*descending*), and a situation where the player is initially faring poorly but slowly ”learns” to exploit an opponent (*ascending*). The addition of these conditions allowed for testing not only for potential differences in post-error slowing due to outcome sequence in the absence of different outcome frequencies, but also for so-called illusion of control (Langer, 1974). The illusion of control refers to the general tendency of people to assume they have control over a random process. For the purpose of the present study, the most relevant kind of illusion of control is the increased confidence in future success participants express as a function of previous success in a guessing task (Burger, 1986; Coventry & Norman, 1998; Langer & Roth, 1975). For this, I included three different measures of confidence in winning: 1) a per-trial prediction of a win or a loss, 2) a post-block prediction of future wins over 50 rounds and 3) a period at the end of each block where a participant could choose to play up to 24 extra rounds (self-terminating play; see Ejova et al., 2013).

The success slope method can only dissociate win-rates from player skill, but cannot dissociate win-rates from *perceived* exploitability. This is especially so if the participants themselves are playing in any systematic manner. For a participant expressing a bias, in a game condition where they are set up to experience a high win-rate, the deduction that the opponent is exploitable would be completely rational: the participant is, after all, playing according to some rule or pattern and achieving results. The interpretation of performance speed as a function of success slope is as follows. On one hand, finding no difference in post-error slowing between two different win-rates (high and low) in the present experiment would be evidence against the notion that the frequency of outcomes is the main factor in post-error slowing. On the other hand, finding a significant difference in post-error slowing would suffer from ambiguity as to whether it supports the importance of frequency, of perceived exploitability, or both.

A secondary reason for manipulating the participants' win-rates and their trajectories was to examine the effects of different success slopes on *stayshift* behaviour. In Experiments 1 and 2, the random choices selected by the *unexploitable* opponent may at times have led to situations where participants experienced very high rates of wins or losses across several consecutive trials, which may have affected reliance on RL heuristics. Due to randomization, these effects will likely have been averaged out in any group level analyses. The two conditions in Experiment 3 where a player goes from mostly losing to mostly winning or vice versa (*ascending* and *descending*) make it possible to examine how success slope change affects *stayshift* behaviour or other choice biases. That is, assuming that participants start off with a *stayshift* bias, what happens when they initially do well in the game but then start losing, or vice versa? In the former case, the participant could perceive this as them learning to exploit the opponent and thus increase the bias; in the latter, the participant could perceive this as the opponent learning to exploit them and thus decrease the bias. However, this effect on the *stayshift* bias could be balanced or even countered by the local win-rate in the first half of the block. The ascending and descending conditions, then, work as a test of the importance of early vs. late outcome trends in adopting or discarding the *stayshift* bias. Similarly, the two conditions with an overall above or below chance level of wins (*continuous success* and *continuous failure*) allow for inferences about the frequency of *stayshift* behaviour as a function of success. If participants start off with a bias towards *stayshift*, a condition that forces a majority of the trials to be wins for the participant could be perceived as exploitable and increase the *stayshift* tendency. Likewise, a condition with a majority of losses could be perceived as the opponent exploiting the player, would function as negative feedback for the *stayshift* approach, and could reduce the bias. In sum, this analysis allows for examining what kind of success slope leads to the most perceived exploitation/exploitability by the opponent by using the players' rate of *stayshift* behaviour as a proxy.

The confidence measure used in Experiment 3 was also changed from the one used in Experiments 1 and 2. In Experiments 1 and 2, confidence was measured using a Likert-scale response every 9th trial of each block. In Experiment 3, the response was changed into a dichotomous response where participants simply predicted whether they thought they would win or lose after making their game choice. The measure was included on all trials to avoid an intermittent measure causing participants to erroneously think there was something special about the trials on which the measure was taken. As win-rates were controlled for each participant, the focus of the confidence analysis would be on situations where each participant had gone through a set number of wins and losses around halfway of each block and towards the end of the block (Play Points 1 and 2, respectively; see Figure 3.1 in section 3.2.2). Due to the win-rates being controlled, the rate at which participants' predictions tracked previously experienced win-rates and the rate at which these predictions changed in response to changing win-rates could be more easily measured on the group level than in Experiments 1 and 2.

Two new additional exploratory measures of confidence were included in Experiment 3: an off-line prediction of hypothetical future wins at the end of each block, and a period of optional trials (see Ejova et al., 2013, for a similar design). The former simply allowed for measuring confidence at the end of the block rather than rely on the predictions made during the block. This helps give an overall picture of how participants feel about their chances of winning after experiencing the whole block: whether they will weigh later successes or failures more. The latter allowed for measuring the association between experienced win-rates and confidence in an arguably more ecologically valid manner, by the number of extra rounds a participant was willing to play.

As a way to assess individual differences in the tendency to assume agency even in random situations, which may explain variation in the confidence measures not explained by the experimental manipulations, I used the Locus of Control (LoC; Rotter, 1966), A study by Lange and Tiggeman (1981) showed that the LoC scale has a relatively good test-retest reliability of .61, but also that it has a two-factor structure, with items loading onto a general control factor and a political control factor. A similar factor structure was also found by Parkes (1985). For the purposes of the present study, the political control factor is theoretically not relevant. I chose to use the whole questionnaire (considering the items for the political control factor as filler items) and sought to replicate the factor structure in Lange and Tiggeman (1981) and Parkes (1985) before excluding any items from the analysis.

Finally, Experiment 3 did not include a financial incentive. The proximal reason for this was that the incentive in Experiment 2 failed to produce any effects on *stayshift* behaviour and simply made reaction times in general slower. More generally, however, if *stayshift* is assumed to be a kind of default heuristic decision rule that is simply easier for people to follow than other decision rules (Alós-Ferrer & Ritschel, 2018; Scheibehenne et al., 2011; Wilke et al., 2014), there is no theoretical reason to assume that the absence of monetary incentives would make people fully capable of avoiding the use of this rule. Decisions need to come from somewhere, and “hard-wired” RL biases are a plausible mechanism for an initial preference for one decision over another. Monetary incentives may have separate effects on reliance on RL heuristics (increasing reliance; Achtziger et al., 2015) and motivation on non-incentivized trials (reducing motivation; Ma et al., 2014), making inferences about differences between incentivized and non-incentivized conditions more difficult (see Read, 2005, for a general discussion of issues relating to interpreting the effects of incentives). As one of the main aims of this series of experiments is examining how flexible or inflexible these supposedly default decision rules are, the experiments need to look at baseline performance without additional influences such as financial incentive.

3.2.1.1 Hypotheses. My hypotheses for Experiment 3 were:

1) A replication of *stayshift* bias in Matching Pennies (as per e.g. Scheibehenne et al., 2011, Wilke & Scheibehenne, 2011, Achtziger et al., 2015)

2) More *stayshift* behaviour in higher win-rate conditions

3) Post-error slowing when win-rate is above chance, post-error speeding when win-rate is at or below chance

4) More voluntary rounds played, higher confidence of winning during Play Point 2 and higher off-line predictions of wins in the ascending win-rate condition than in the descending win-rate condition

5) Overall more voluntary rounds played, higher confidence of winning during Play Point 2 and higher off-line predictions of wins in conditions with a higher overall win-rate

3.2.2 Method.

3.2.2.1 Participants. Fifty-two (44 female; M_{age} = 18.53; SD_{age} = 1.05) were recruited from the University of Sussex participant pool. Six participants were excluded due to indicating having correctly guessed what the experimental manipulation was in debriefing, leading to a final sample of forty-six. Informed consent was obtained from all participants before testing, and the experiment was approved by the Sciences & Technology Research Ethics Committee (C-REC) at the University of Sussex (ER/BJD21/21). Participants received course credit as reward for participation.

3.2.2.2 Materials.

3.2.2.2.1 Game trials. During game trials, static pictures of one penny coins (depicting the choices made by the opponent and participant) were presented on screen by the experiment program at 12º × 6º, with participants sat approximately 57 cm away from a 22" Diamond Plus CRT monitor (Mitsubishi, Tokyo, Japan). Stimulus presentation was controlled by Presentation 19 (build 03.31.15) and responses were recorded using a keyboard.

3.2.2.2.2 Questionnaire. I used Rotter's Locus of Control (LoC; Rotter, 1966) questionnaire to provide an assessment of the participants' tendency to associate the outcomes of their actions to either themselves (internal LoC) or outside factors (external LoC). The questionnaire consists of 29 dichotomous choice questions (see Appendix 3).

3.2.2.3 Design. The study was a 2x2 within-subjects design with early *(success, failure)* and late *(success, failure)* game outcomes as factors (see Figure 3.1). The order of the conditions was counterbalanced between participants. Each condition consisted of 84 mandatory rounds of the game, followed by up to 24 optional rounds, leading to a maximum of 108 rounds per block. Win-rates ranged from 1/6 to 5/6 and were defined for consecutive bins of 6 trials separately; that is, for each bin, there was at least one and at most five wins. The order of wins and losses within each bin was randomized.

Each condition had an initial win-rate of 3/6 for the first bin. During the first three bins, the win-rate either increased up to 5/6 (*descending* and *continuous success*) or decreased down to 1/6 (*ascending* and *continuous failure*), staying at this point for two bins. After this, the win-rates once again either increased (*ascending* and *continuous failure*) or decreased (*descending* and *continuous success*) for the next two bins, going back to 3/6 and staying at this point for two bins (Play Point 1; see Figure 3.1). Thus, the *continuous success* and *descending* conditions can be characterized by an initial upward trajectory that then descends back to chance level, and the *continuous failure* and *ascending* conditions can be characterized by an opposite downward trajectory that then rises back to chance level. After this return to baseline, the win-rates of the *continuous success* and *ascending* conditions increased up to 5/6 (*late success)* and the win-rates of the *continuous failure* and *descending* conditions decreased down to 1/6 (*late failure*) after two bins, staying at these points for the last three bins (Play Point 2; see Figure 3.1). For each condition, this final win-rate would also continue for any extra rounds the participant opted to play (Play Point 3: see Figure 3.1).

Thus, in the *continuous success* condition, the participant experienced an initial upward trajectory, followed by a descent to baseline, followed again by an upward trajectory *(early success, late success)*; in the *descending* condition, the participant experienced an initial upward trajectory followed by a long downward trajectory through baseline to below-chance level *(early success, late failure)*; in the *ascending* condition, the participant experienced an initial downward trajectory followed by a long upward trajectory to above-baseline *(early failure, late success)*; and in the *continuous failure* condition, the participant experienced an initial downward trajectory, followed by a return to baseline, followed again by a downward trajectory *(early failure, late failure)*.

3.2.2.4 Procedure. At the beginning of the experiment, I explained the rules of the game and the structure of the experiment to each participant, allowing them to ask any questions about how to proceed. I instructed participants to try and maximize their score for each of the four blocks and told that the opponents in different blocks may play the game in different ways. After giving the general instructions, I left the room. The experiment program started each block with a reminder of the rules of the game, the response keys, the goal the participant should aim for, and the option of playing up to 24 extra rounds.

Each round of the game began with the participant making a choice of heads or tails, with a screen prompt reminding them of the response keys (“Please choose HEADS [k] or TAILS [p]”). After this, the participant made a prediction of whether they were going to win or lose this round, again with a screen prompt reminding them of the response keys (“Please predict WIN [w] or LOSS [d]”). Participants were instructed to make the game choice response using their right hand and the prediction response using their left hand. After the game choice and prediction, the program presented the choices made by the participant and the opponent depicted by pictures of either the heads or tails side of a penny coin for 2000ms. After this, the text WIN or LOSE was presented on screen for 1000ms, and at the end of the 1000ms period the participants score updated and the next round started immediately. The score was increased by one point for each win and reduced by one point for each loss.

After 84 rounds, the program informed the participant that they could keep on playing for up to 24 rounds or quit at any point by pressing the space bar. The block ended when the participant either pressed space or played through the maximum number of optional rounds. At the end, the participant was asked to indicate how many rounds they thought they would win if they were to keep on playing for another 50 rounds. After recording the response, the program instructed the participant to take a short break before the next section. After four blocks, the participant responded to the locus of control questionnaire. After the questionnaire, the program instructed the participant to inform the experimenter that they had finished, after which I debriefed the participant and thanked them for their time.

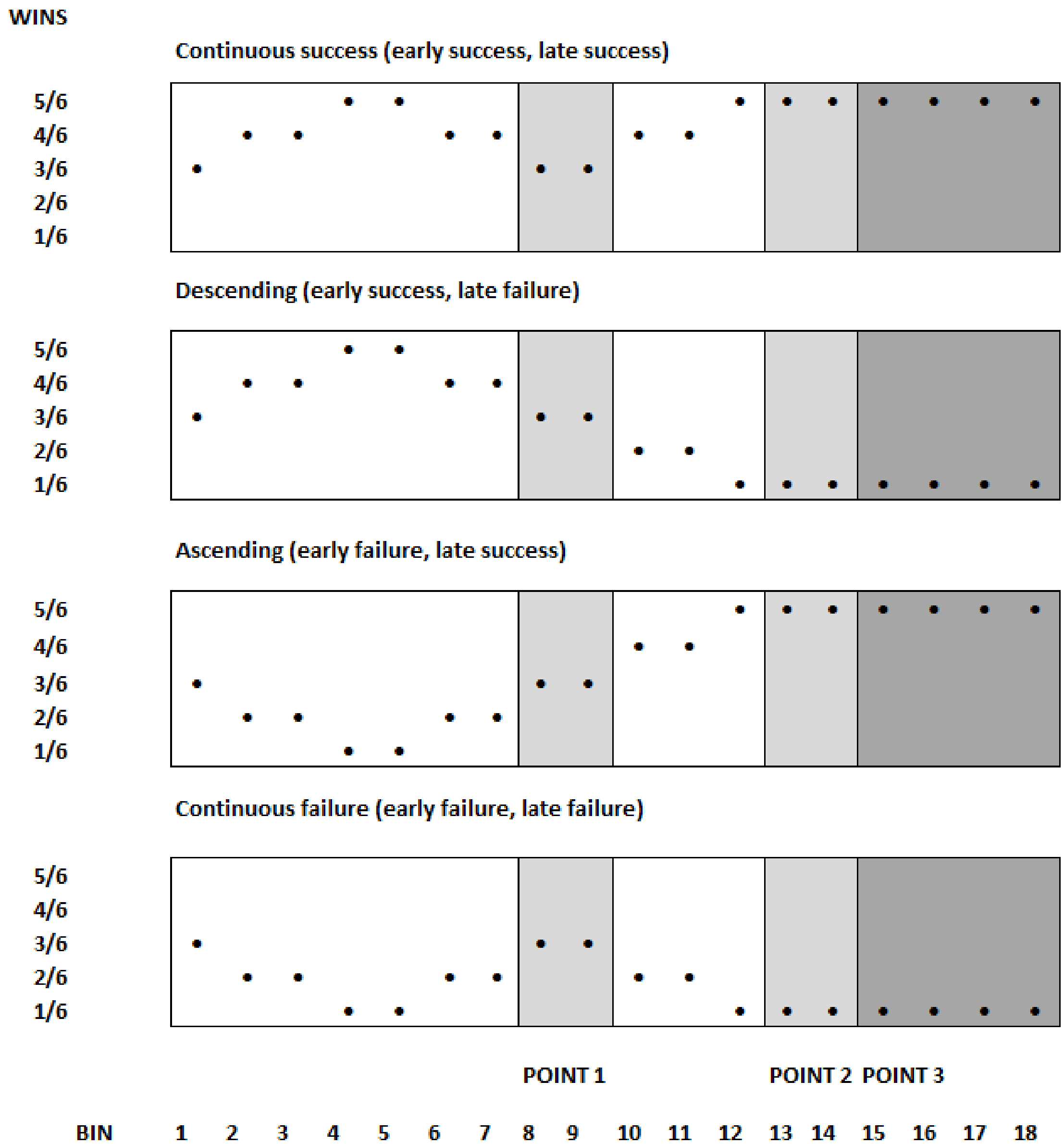

Figure 3.1. The win-rate trajectories of the four experimental conditions in Experiment 3. Each dot represents a bin of 6 game rounds.

3.2.3 Results.

3.2.3.1 Reinforcement biases. I analysed rates of *win-stay* and *lose-shift* responses wtwo ways: first, by comparing the overall rates of the specific decision type between conditions; second, by comparing the rates of individual participants who expressed a statistically significant bias towards *win-stay* or *lose-shift* (denoted by a binary variable) between conditions. I ran each analysis separately using a generalized linear mixed model (GLMM) with a logit link (as both the response variable per each trial and the variable indicating a significant bias were binary variables). I used the lme4 and emmeans packages in R, version 3.5.1. See Table 3.1 for descriptives.

The reason for running two types of analyses was as follows. A GLMM analysis of the overall rates of a specific response type can test for differences in response rates between conditions on the group level. It can also provide an estimate of the group-level likelihood of a response type in a given condition, which allows for comparing it against a baseline, thus allowing for testing if there is a group-level bias. However, the results of this analysis hides information about how common a biased response pattern is among *individuals*. Because of this, the analysis does not clearly tell whether potential differences in win-stay responding are due to individuals increasing or decreasing their biased behaviour in some conditions, or from simply more or less individuals exhibiting a bias in those conditions. A GLMM on a binary variable denoting whether an individual participant behaved in a biased way allows for examining these questions.

Table 3.1. Win-stay and lose-shift choices and likelihoods of individual participants having a win-stay or lose-shift bias (back-transformed estimated marginal means) in Experiment 3

Success slope	*WS overall*	*WS individual*	*LS overall*	*LS individual*
Cont. success	63.0% (2.7%) *	50.1% (9.8%)	58.7% (2.5%) *	34.1% (10.1%)
Descending	58.8% (2.9%) *	32.7% (8.9%)	55.9% (2.3%) *	17.8% (7.3%)
Ascending	60.6% (2.9%) *	50.1% (9.8%)	59.3% (2.3%) *	20.2% (7.9%)
Cont. failure	63.8% (2.9%) *	41.2% (9.6%)	59.7% (2.2%) *	31.1% (9.7%)

Note: standard error in parentheses. Asterisks indicate that the expected percentage of decisions (50%) falls below the 95% CI of the estimated marginal mean of observed decisions. Likelihoods of individual biases based on results obtained using binary variables based on the z-test.

3.2.3.1.1 Win-stay. I analysed the rate of *win-stay* behaviour by entering the proportion of *stay* decisions following wins into a GLMM with a logit link, with early *(success, failure)* and late *(success, failure)* outcome conditions and their interaction entered as fixed effects, and participants as random intercepts. The *success* conditions within early and late outcome conditions were fixed as the reference for all fixed effects. See Table 3.1 for back-transformed probabilities of win-stay behaviour by condition.

There was no significant main effect of early outcomes (β = -0.10, SE(β) = 0.07, z = -1.56, *p* = .118), a significant main effect of late outcomes (β = -0.18, SE(β) = 0.06, z = -2.77, *p* = .006) and a significant interaction effect (β = 0.32, SE(β) = 0.10, z = 3.07, *p* = .002) . A Tukey test of pairwise comparisons indicated that individuals' probability of staying after wins in the *descending* condition (early success, late failure; M = 58.8%, SE = 2.9%) was significantly lower than the *continuous failure* condition (early failure, late failure; M = 63.8%, SE = 2.9%) and the *continuous success* condition (early success, late success; M = 63.0%, SE = 2.7%) with no other significant differences. For each condition, the lower end of the 95% confidence interval was higher than .5, indicating that participants overall exhibited a *win-stay* bias in each condition.

To conduct an analysis of individual-level biases in different conditions, I first calculated a binary variable for each participant in each condition to denote whether they had a significant *win-stay* bias. I used a two-tailed one sample z-test of proportion on the rate of *win-stay* choices for each participant. I set P0 (the null hypothesis probability) at 50% i.e. the rate of *win-stay* behaviour one would expect if a player were playing randomly. Specifically, this z-test used P0 in calculating the standard error for the probability (as opposed to the observed probability of each participant). I assigned this variable a value of 1 if the participant had a rate of *win-stay* choices significantly higher than 50% of all eligible trials, and a value of 0 if the participants' rate of *win-stay* behaviour was significantly lower than 50% or did not differ from 50%. I then entered the binary variable into the same GLMM as above. The model tested for both main effects and the interaction between the predictors. There were no significant main effects of early (β = -0.00, SE(β) = 0.48, z = 0.00, p = 1) or late outcomes (β = -0.73, SE(β) = 0.49, z = -1.47, p = .141) and no significant interaction (β = 0.37, SE(β) = 0.69, z = 0.54, p = .592), indicating that participants were not more likely to exhibit a *win-stay* bias in any condition compared to other conditions.

Since the z-test is a normal distribution approximation for comparing percentages, it may not work perfectly when the number of trials or the number of "hits" or the occurrences of interest (in this case, stay decisions) is low (< 10). Since it could not be guaranteed that a participant would necessarily stay on any trial following a loss, I tested the robustness of the result using an alternative variable. I calculated a binary variable using the more conservative exact binomial test instead of the z-test for the rates of stay decisions after wins for each participant separately, and entered this variable into the GLMM above. The results of this analysis did not differ meaningfully from the analysis that used the binary variable based on z-test results as the dependent variable.

3.2.3.1.2 Lose-shift. I conducted the analyses for the rate of *lose-shift* choices and individual *lose-shift* bias using the same models as for *win-stay* (see Table 3.1). For the rate of individual *lose-shift* choices, there were no significant main effects of early (β = 0.03, SE(β) = 0.08, z = 0.32, p = .749) or late outcomes (β = -0.12, SE(β) = 0.08, z = -1.51, p = .131), and no significant interaction (β = 0.13, SE(β) = 0.10, z = 1.32, p = .188) . For each condition, the lower end of the 95% confidence interval was higher than .5, indicating that participants on the whole exhibited a lose-shift bias in each condition. Similarly, for the binary lose-shift bias variable, there were no significant main effects of early (β = -0.74, SE(β) = 0.55, z = -1.33, p = .181) or late outcomes (β = -0.58, SE(β) = 0.54, z = -1.07, p = .287). There was a marginal interaction effect (β = 1.45, SE(β) = 0.78, z = 1.86, p = .064), but pairwise comparisons revealed no significant differences. The marginal interaction likely stemmed from the *continuous success* and *continuous failure* conditions having numerically higher likelihoods of individual bias than the *ascending* or *descending* conditions. In sum, the results suggest that on the individual level, participants were not significantly more likely to express a bias in any condition over others. The results of this analysis did not meaningfully change when using a binary variable calculated based on the more conservative exact binomial test, other than that the interaction effect was no longer marginal (p = .433).

3.2.3.2 Reaction time. I entered median reaction times (in milliseconds) for decisions at trial *n+1* into a two-way repeated measures ANOVA with game outcome at trial *n (win, lose)* and experimental condition (*continuous success, descending, ascending, continuous failure*; see Figure 3.1) entered as factors. I made the choice to treat the conditions as a single factor instead of a 2x2 (early x late) due to the early/late distinction causing a confound with overall block win-rate. That is, the *descending* (early success, late failure) and *ascending* (early failure, late success) conditions both had an overall win-rate of 50% for the whole block, whereas the *continuous success* condition (early success, late success) had an overall win-rate of roughly 70% and the *continuous failure* (early failure, late failure) had an overall win-rate of roughly 30%. Due to this, a main effect of either the late or the early factor would be the same in terms of mean win-rate, rendering the analysis difficult to interpret. Eight participants with an average median reaction time more than two times the group average median in any block were excluded from the analysis (as in previous experiments; see Chapter 1), yielding a final sample of 38. I ran the analyses using the ez and emmeans package in R, version 3.5.1. Mauchly's test indicated violations of sphericity, so I used the Greenhouse-Geisser correction for effects with violations. There were no main effects of experimental condition [$F(2.24, 82.77) = 0.55$, MSE = 87091, $p = .598$, $\eta_p^2 = .02$], or outcome at trial *n* [$F(1, 37) = 0.30$, MSE = 17126, $p = .585$, $\eta_p^2 < .01$], but there was a significant interaction [$F(2.37, 87.56) = 3.24$, MSE = 12651, $p = .036$, $\eta_p^2 = .08$]. However, Tukey's test indicated no significant differences for any comparisons (every $p > .05$). The results suggest no post-error slowing as a function of win-rate trajectories or overall win-rate. See Table 3.2 for descriptive statistics.

Table 3.2. Average median reaction times (milliseconds) in Experiment 3, N = 38

Success slope	*Win*	*Lose*
Cont. success	448 (26)	463 (30)
Descending	440 (26)	442 (26)
Ascending	445 (27)	503 (46)
Cont. failure	512 (61)	470 (39)

Note: standard error in parentheses.

3.2.3.3 On-line confidence measures. I analysed on-line confidence ratings collected from Play Points 1 and 2 similarly to *win-stay* and *lose-shift* biases: by first running a GLMM on the overall rate of win predictions and then running the same model on a binary variable denoting whether a participant had made significantly overconfident predictions in relation to the actual win-rate they had experienced prior to that Play Point. I used the lme4 and emmeans packages in R, version 3.5.1. See Table 3.3 for descriptives.

Table 3.3. Win prediction rates and likelihoods of individual participants having an overconfidence bias (back-transformed estimated marginal means) in Experiment 3

Success slope	*PP1 overall*	*PP1 biased*	*PP2 overall*	*PP2 biased*
Cont. success	82.0% (3.2%) * [69.1%]	21.0% (7.8%)	83.7% (2.7%) * [66.7%]	31.0% (8.4%)
Descending	84.8% (2.8%) * [69.1%]	23.6% (8.3%)	78.9% (3.2%) * [55.6%]	31.0% (8.4%)
Ascending	73.0% (4.1%) * [31.0%]	77.1% (8.2%)	89.1% (2.0%) * [44.4%]	85.4% (5.9%)
Cont. failure	67.5% (4.6%) * [31.0%]	63.1% (10.1%)	56.0% (4.6%) * [33.3%]	33.6% (8.7%)

Note: standard error in parentheses. Expected win prediction rate in square brackets; expected win-rate is equal to the win-rate of all trials prior to that Play Point. Asterisks indicate that expected percentage of win predictions falls outside the 95% CI of the estimated marginal mean of observed win predictions. Likelihoods of individual biases based on results obtained using binary variables based on the z-test.

3.2.3.3.1 Play Point 1. To examine differences in rates of prediction, I entered win/lose predictions for the 12 game trials in Play Point 1 into a GLMM with a logit link function with win-rate prior to Play Point 1 *(high, low)* entered as a fixed effects and participants and a dummy variable denoting the first and second measures in both win-rate conditions entered as random intercepts. I collapsed the data across *consistent success* and *descending* conditions for *high*, and acrosss *consistent failure* and *ascending* condition for *low*. The dependent variable had a value of 1 for win predictions and 0 for loss predictions. See Table 3.3 for back-transformed probabilities of predicting win in each condition.

The high win-rate condition was fixed as the reference. There was a significant main effect of win-rate prior to Play Point 1 ($\beta = -0.76$, $SE(\beta) = 0.11$, $z = -7.15$, $p < .001$). There was an overall higher probability of an individual predicting wins in the high win-rate (M = 83.4%, SE = 2.8%) compared to the low win-rate (M = 70.3%, SE = 4.1%) condition. In each condition, the hypothesized rate of win predictions (the rate of wins experienced in the condition prior to Play Point 1) fell below the 95% confidence interval of the estimated marginal mean of win predictions in the condition, suggesting overall over-confidence. However, the difference between the prior experienced win-rate and average predicted win-rate was greater in the low win-rate condition (see Table 3.3).

In order to test for individual level differences in expressing overconfident predictions, I calculated a binary variable for each participant, separately for each condition, to denote whether their rate of win predictions differed from the prior experienced win-rate. I calculated this variable similarly to the binary bias variables for reinforcement biases: by running a two-tailed z-test of proportion on the rate of win predictions the participant made, with the prior experienced win-rate (approximately 30% in the low win-rate condition and 70% in the high win-rate condition) entered as the null. I assigned the variable a value of 1 if the participant predicted significantly more wins than they had experienced and a value of 0 if they predicted significantly less wins than experienced or if there was no significant difference. I entered this variable into the same GLMM as before, with the high win-rate condition fixed as the reference. There was a significant main effect of win-rate ($\beta = 2.09$, $SE(\beta) = 0.44$, $z = 4.74$, $p < .001$). Individuals were overall more likely to have a bias in the low win-rate condition (M = 70.2%, SE =

7.5%) compared to the high win-rate condition (M = 22.5%, SE = 6.5%). To test for the robustness of this result, I conducted the same analysis using a binary variable calculated using the more conservative exact binomial test instead of the z-test for the win prediction rates of each participant separately. The results using this variable were similar to those obtained using the binary variable created based on the results of z-tests.

In sum, all conditions had an overall significantly higher rate of win predictions than the prior experienced win-rate, with a greater difference between prior and predicted win-rates in the low win-rate conditions. Participants in the high win-rate condition predicted overall more wins, suggesting predictions tracked actual experienced win-rate to a degree. In terms of individual overconfidence, there were more individuals with a significant overconfidence bias in the low win-rate condition, while only a minority in the high win-rate condition exhibited a significant bias.

3.2.3.3.2 Play Point 2. To examine differences in rates of prediction, I entered the rate of win predictions during the 12 trials in Play Point 2 into a GLMM with a logit link function with early win-rate *(high, low)* and late win-rate *(high, low)* and their interaction entered as fixed effects, and participants as random intercepts. See Table 3.3 for back-transformed probabilities of predicting wins in each condition. The high win-rate condition within both early and late outcome conditions was fixed as the reference for each fixed effect. There was a significant main effect of early ($\beta = 0.47$, $SE(\beta) = 0.17$, $z = 2.77$, $p < .01$) as well as late ($\beta = -0.32$, $SE(\beta) = 0.15$, $z = -2.08$, $p < .05$) outcome types and a significant interaction ($\beta = -1.55$, $SE(\beta) = 0.22$, $z = -7.00$, $p < .001$). A Tukey test of pairwise comparisons indicated that individuals' probability of predicting wins differed significantly between all conditions ($p < .05$) except for the comparison between the *continuous success* (early success, late success) and *descending* (early success, late failure) conditions. The *ascending* condition (early failure, late success) had a higher win prediction rate than any other condition (M = 89.1%, SD = 2.0%), while the *continuous failure* condition (early failure, late failure) had a lower win prediction rate than any other condition (M = 56.0%, SD = 4.6%). Each condition had a win prediction rate significantly higher than the prior experienced win-rate in that condition (see Table 3.3).

To repeat the binary overconfidence bias variable analysis from Play Point 1, I calculated a similar variable for each participant for Play Point 2. The null hypothesis was the win-rate experienced prior to Play Point 2: 66.7% in the *continuous success* condition, 55.6% in the *descending* condition, 45.6% in the *ascending* condition, and 33.3% in the *continuous failure* condition., I then entered the binary overconfidence bias variable into the GLMM from the prior analysis of Play Point 2. There was a significant main effect of early outcomes ($\beta = 2.57$, $SE(\beta) = 0.59$, $z = 4.33$, $p < .001$), no significant main effect of late outcomes ($\beta = 0.00$, $SE(\beta) = 0.49$, $z = 0.00$, $p = .999$) and a significant interaction effect ($\beta = -2.45$, $SE(\beta) = 0.76$, $z = -3.20$, $p < .01$). A Tukey test of pairwise comparisons revealed that the *ascending* condition (early failure, late success) had a significantly higher likelihood of an individual being overconfident (M = 85.4%, SE 5.9%) than any other condition ($p < .05$ for all comparisons with the *ascending* condition), with no other significant differences between conditions ($p > .05$ for all other comparisons). Running the same analysis on a binary variable based on the more conservative exact binomial test, the results differed in that there was no significant interaction effect. However, the numerical trend of a notable difference between the ascending condition (M = 57.1%, SE = 7.9%) and the descending (M = 9.7%, SE = 4.6%), continuous success (M = 22.3%, SE = 6.7%) and continuous failure (M = 13.8%, SE = 5.4%) conditions remained. Pairwise comparisons (Tukey's HSD) also indicated a difference between the ascending condition and all other conditions ($p < .05$ for all comparisons), with no other significant differences, when using this alternative calculation of bias. Thus, the *ascending* condition had both the overall highest win prediction rate in absolute terms during Play Point 2, and also the highest number of individual people predicting significantly more wins than they had experienced prior to Play Point 2.

3.2.3.4 Locus of control. To form a baseline assessment of the participants' tendency to associate outcomes to themselves or outside influences, I calculated an internal vs. external locus of control score for each participant based on the items loading onto the general control factor to be used as a covariate for Play Point 3 and the off-line prediction measure (see below). Based on studies by Lange and Tiggeman (1981) and Parkes (1985), the locus of control questionnaire measures two distinct factors: a general control factor relating to an individual's tendency to associate outcomes in their life to either their own agency, hard work, ability etc. rather than luck, the actions of others, or fate; and a political control factor, relating to an individual's beliefs about control over political institutions and world events. For the purposes of the present study, only the former is theoretically relevant. For this factor, Parkes (1985) lists items 2, 4, 5, 9, 10, 11, 15, 16, 18, 23 and 25 of the questionnaire, whereas Lange and Tiggeman (1981) list items 5, 9, 11, 13, 15, 16, 18, 25, and 28 (see Appendix 3.1 for items). To calculate a sum score of the relevant items, I first ran a single-factor maximum-likelihood exploratory factor analysis (EFA) using all of the items listed by both Parkes (1985) and Lange and Tiggeman (1981). I excluded items with loadings below .40 and ran the EFA again with the remaining items until all items had loadings at or above the threshold. Items 5, 10, 15, 18, 25 and 28 remained for the final sum variable. These items had an acceptable reliability: Cronbach's $\alpha = .68$.

3.2.3.5 Additional confidence measures. I analysed the number of extra rounds played in Play Point 3 (self-terminating play) and the number of extra rounds the participants predicted they would win at the end of the block as additional confidence measures. I analysed these measures first on their own in an ANOVA, then with the added covariate of the LoC measure in an ANCOVA. I used the ANCOVA only to interpret the effects of the covariate; I interpreted main effects only from the ANOVA without the added covariate, in accordance with the recommendations of Schneider et al. (2015) on the use of ANCOVAs in repeated-measures designs. I used the ez and emmeans packages in R, version 3.5.1. See Table 3.4 for descriptives.

Table 3.4. Extra rounds and off-line predictions of future wins in Experiment 3

Success slope	*Extra rounds*	*Off-line prediction*
Cont. success	10.9 (1.6)	27.9 (1.7)
Descending	8.4 (1.6)	15.7 (1.0)
Ascending	12.4 (1.6)	21.8 (1.7)
Cont. failure	8.2 (1.6)	9.3 (1.0)

Note: standard error in parentheses.

3.2.3.5.1 Play Point 3. I measured the amount of self-terminating play as the total number of extra rounds played (ranging from 0 to 24). I entered this variable into a two-way repeated measures ANOVA with early *(success, failure)* and late *(success, failure)* overall outcomes as the within-subjects factors. See Table 3.4 for descriptive statistics.

There was no significant main effect of early outcomes [$F(1,45) = 0.38$, $MSE = 47.98$, $p = .540$, $\eta_p^2 < .01$], a significant main effect of late outcomes [$F(1,45) = 4.53$, $MSE = 110.78$, $p = .039$, $\eta_p^2 = .09$], and no significant interaction effect [$F(1,45) = 0.37$, $MSE = 85.74$, $p = .548$, $\eta_p^2 < .01$]. In *late success* conditions, participants played significantly more extra rounds ($M = 11.6$, $SE = 1.3$) than they did in *late failure* conditions ($M = 8.3$, $SE = 1.4$). In the ANCOVA with the locus of control measure as the covariate, there was no significant main effect of the locus of control measure, and no interactions with it and the within-subject factors (every $p > .05$).

3.2.3.5.2 Off-line prediction. I analysed the off-line prediction measure regarding how many rounds participants thought they would win if they were to keep on playing for another 50 rounds in the same manner as the number of extra rounds played. See Table 3.4 for descriptives. In the ANOVA, there was a significant main effect of early outcomes [$F(1,45) = 35.97$, $MSE = 50.13$, $p < .001$, $\eta_p^2 = .44$], a significant main effect of late outcomes [$F(1,45) = 76.04$, $MSE = 91.91$, $p < .001$, $\eta_p^2 = .63$] and no interaction effect between early and late outcomes ($F < 1$). Participants had higher future win predictions in the *early success* conditions ($M = 21.8$, $SE = 1.1$) than in the *early failure* conditions ($M = 15.5$, $SE = 0.9$). Likewise, participants had higher future win predictions in the *late success* conditions ($M = 24.826$, $SE = 1.307$) than in the *late failure* conditions ($M = 12.5$, $SE = 0.9$). In the ANCOVA with the locus of control measure as covariate, there was no significant main effect of the locus of control measure [$F(1,44) = 1.35$, $MSE = 130.89$, $p = .252$, $\eta_p^2 = .03$] and no significant interaction effects for the locus of control measure (every $p > .05$).

3.2.4. Discussion. In Experiment 3, participants played Matching Pennies (MP) against a computer opponent and experienced four different predetermined success slopes (with differing overall win-rates) in four separate blocks of games. I examined reinforcement biases, reaction times, and different measures of confidence as a function of success slopes and win-rates.

There was a general trend of both *win-stay* and *lose-shift* biases regardless of the success slope condition. Thus, the results replicate the general trend from earlier research using binary choice zero-sum games (see e.g. Achtziger et al., 2015; Scheibehenne et al., 2011; Wilke & Scheibehenne, 2011), but not the results of Experiments 1 and 2. The results did not support my initial hypothesis that *stayshift* responding would increase with win-rate. My hypothesis here was based on the idea that if players start with an initial bias but keep losing at above-chance levels, this could be interpreted as negative feedback for the response pattern. Crucially, *stayshift* responding in the *continuous success* condition (early success, late success), with an overall win-rate of roughly 70%, did not differ significantly from *stayshift* responding in the *continuous failure* condition (early failure, late failure), with an overall win-rate of roughly 30%. The condition where the participant is winning the least could reasonably be considered the one condition that participants would experience as the most exploiting, and should thus cause the most reduction in *stayshift*, but this does not seem to be the case. The *continuous success* condition did not have a higher rate of *stayshift* responding than any other condition, and the *continuous failure* condition did not have a lower rate of *stayshift* responding than any other condition. *Win-stay* responding in the *descending* condition (early success, late failure; M = 58.77%, SE = 2.90%) was significantly lower than in the *continuous failure* condition (early failure, late failure; M = 63.84%, SE = 2.92%). It is unclear why only these two conditions would differ from each other, and why the effect would be in this direction, as the *continuous failure* condition had a lower win-rate. One could argue that a success slope with actual large shifts in win-rate could be considered more realistic scenarios of exploiting an opponent or being exploited by them than a success slope that stays above or below chance-level most of the time. However, this would imply that the *ascending* condition (early failure, late success) should cause an increase in *stayshift* responding, but the ascending condition did not differ significantly from any other condition. In sum, the one small but statistically significant difference observed here does not fit neatly with any explanation when all of the results are taken into account. Additionally, the likelihood of an individual expressing a *win-stay* or *lose-shift* bias was similar between all conditions. Thus, the results suggest that the effect of manipulated win-rates or win-rate trajectories on reinforcement biases is likely negligible.

Why did participants not increase their rate of *stayshift* behaviour when that biased pattern of play should have been rewarded by conditions where it seemingly led to high win-rates, or vice versa? This would seem to imply that win-rates significantly higher or lower than chance are not strong enough feedback to affect reinforcement biases. Yet it is obvious that people can learn from such feedback - people increase their rate of *stayshift* behaviour when the game task actually is exploitable by that strategy i.e. when it leads to wins (Scheibehenne et al., 2011). The question, then, is more specifically: why did the feedback in the current experiment not allow for players to “learn”? In hindsight, this may be caused by the way the conditions were set up, leading to feedback that was too noisy for the participants. Imagine a player who starts playing, and initially has a *stayshift* bias. The player observes that they are winning/losing more than would be expected by chance. However, the player is also making choices that do not follow the *stayshift* rule (a bias does not equal to rote following of a rule on every round). On the rounds that the player does not follow the rule, they are still equally likely to win/lose. Over several rounds, this makes the feedback ambiguous: everything the player does yields positive/negative feedback. In contrast, a hypothetical opponent that simply behaved in a way that would be specifically weak to exploitation by a participant with a *stayshift* bias would yield the player increased wins only when the player made *stayshift* decisions and not in other situations. Thus, in the present experiment, both increased positive and increased negative feedback may have been pulling players into several directions, leading them to eventually play the game relatively similarly in all situations, i.e. defaulting to *stayshift* when nothing can be learned (see Lee et al., 2004; Lee et al., 2005).

There was no post-error slowing in any condition, and no other reaction time effects. If the "infrequent outcomes" notion of post-error reaction time changes (see Danielmeier & Ullsperger 2011) were correct for the present game task, one should expect post-error slowing in the *continuous success* condition and post-success slowing in the *continuous failure* condition. The numerical trend pointed in this direction, but the difference was over all very small: given the differences in design between Experiment 3 and Experiments 1 and 2, the effect should have been larger in Experiment 3 due to controlling the win-rates of each participant. It thus seems that something else than or in addition to high or low win-rates is needed to account for these reaction time differences in games. One possibility is that these differences occur when participants are winning at above-chance rates an opponent they *perceive* as exploitable with a clear pattern, which the participants may not have been able to do even in the high win-rate condition (see above). However, the additional win prediction measures taken between game responses and feedback may also explain the lack of post-error slowing, as introducing delays between response and feedback may mask the slowing down or decrease it by preventing impulsivity. Introducing delays can also reduce *lose-shift* but not *win-stay* responding (Gruber & Thapa, 2016), which may explain why the *win-stay* bias was stronger in Experiment 3 than in Experiments 1 and 2.

To measure how confidence updated as the game goes on, I measured the participants' predictions of the round outcome on each round of the game and compared the percentage of "win" predictions to the actual experienced win-rate during two separate 12-round periods: one after 42 rounds out of 84 (Play Point 1) and the second after 72 rounds out of 84 (Play Point 2). During Play Point 1, participants in both high and low win-rate conditions were predicting overall more wins than they had experienced prior to that point. The rate of win predictions was higher in the high win-rate conditions, suggesting that the predictions tracked actual prior win-rates to some degree. The likelihood of a participant exhibiting a statistically significant bias was much higher in the low win-rate condition, suggesting that the average overconfidence in the high win-rate condition was caused by a smaller number of people with over-confident predictions.

Why did a majority of players in the low win-rate conditions (*ascending* and *continuous failure*) of Play Point 1 seem to predict wins above their prior win-rate? The bias was not only common in terms of rates of individuals being overconfident, but relatively high: the participants predicted on average over twice as many wins as they had experienced previously. Only a minority of participants had an overconfidence bias in the high win-rate conditions (*descending* and *continuous success*). Due to the way individual biases were defined, this result could be argued to simply stem from participants predicting high wins in both conditions, which would only show up as bias in the low win-rate condition. However, since the overall rate of win predictions was higher in the high win-rate conditions, it seems that win-rate did have an effect on predictions, and this was not a case of participants simply predicting above-chance wins at the same rate in both conditions. It may be that a low win-rate in a game the participants assume is random may more likely induce a Gambler's Fallacy, whereas the high win-rate could more likely induce a Hot Hand Fallacy (see Section 1.3). That is, the participants could be making overconfident predictions in the two conditions for different reasons. The hot hand fallacy, characterized by an expectation of a continuing streak is observed when people believe the cause of the streak to be a non-random process, whereas the gambler's fallacy, characterized by the opposite expectation of a streak ending is observed when the streak is believed to be caused by a random process (see e.g. Ayton & Fischer, 2004; Burns & Corpus, 2004; Gronchi & Sloman, 2008; Tyszka et al., 2008). Due to the way the conditions are structured, conditions of early success may be more likely to lead to a perceived covariation between behaviour and outcome. On the other hand, the early failure conditions would cause the player to lose no matter what they do, making it more likely for them to conclude that the process generating the outcomes is random, and that thus they should start winning at some point.

After Play Point 1, the trajectories diverged, leading to four different success slopes. In the *continuous success* (early success, late success) and *continuous failure* (early failure, late failure) conditions, the prior win-rate trend continued. In the *ascending* (early failure, late success) and *descending* (early success, late failure) conditions it reversed. Based on the prediction rates in Play Point 2, participants were still overconfident in general (i.e. predicting more wins than they had previously experienced). However, the *ascending* condition had both the overall highest win prediction rate, and the highest likelihood of individual participants being overconfident. This is in line with the hypothesis that the *ascending* condition would induce more illusion of control than the descending condition, replicating Ejova et al. (2013) and Matute (1995). The result does not align with some earlier studies of the illusion of control (see Burger, 1986; Langer, 1975; Langer & Roth, 1975; see also Thompson et al., 1998, for a review), where the most common observation was that participants had an illusion of control in descending success slope conditions. As Ejova et al. (2013) note, this is most likely due to differences in measurement and certain methodological issues. For example, Langer & Roth (1975) used participants' estimates of how many times they had succeeded in a guessing task as a measure of inferred control, whereas the present experiment relied on predictions about future successes.

Another issue Ejova et al. (2013) point out is earlier uses of measures where participants are directly asked to what degree they believe they can anticipate future events (e.g. Burger, 1986). A question like this could mask the fact that participants may believe in being able to anticipate events not due to their skills but e.g. them being "lucky". Measuring participants' predictions of whether they will succeed or fail is a less direct method, but it avoids the issue of memory effects. This method cannot dissociate whether participants believed in skill or luck, but seems to produce results similar to more contemporary studies on illusion of control (Ejova et al., 2013; Matute, 1995). The fact that participants did not vary much in their rate of predictable *stayshift* behaviour between conditions also fits better with the notion that participants may have simply felt lucky (participants who indicated having believed the game was rigged were excluded from the analyses).

Note that the results of Play Point 2 cannot be fully explained by either the overall block win-rates or the local win-rates during the 12 trials in Play Point 2. The overall win prediction rate in the *ascending* condition was higher than in the *continuous success* condition (and all other conditions). This was despite the overall win-rate prior to Play Point 2 of the *ascending* condition being lower than that of the *continuous success* condition (44.44% vs. 66.66%, respectively). The local win-rate during Play Point 2, however, was identical between these conditions (83.33% in both conditions), suggesting that the win prediction rates during Play Point 2 are not simply short-sighted responses to the win-rate the participant has been experiencing within the last few trials. The number of individuals with a bias was roughly 33.3% in conditions other than the *ascending*, where roughly 80% of the participants expressed a bias (see Table 3.3). The *ascending* condition may be an exception among the conditions as its trajectory most closely resembles that of a player actually learning to exploit an opponent, i.e. "becoming better at the game". Ejova et al. (2013) also suggested that participants in an *ascending* win-rate condition may think they are becoming better at the game, while noting that it is also possible that the overconfident prediction of wins is the norm, and that only in a *descending* condition do participants discard this overconfidence. The results of Experiment 3 actually support both the notion of general overconfidence and a special effect of the *ascending* condition. It is unclear why the rate of win predictions in Play Point 2 was lower in the *continuous success* condition than in the *ascending* condition, since the former never fell below chance-level wins locally. Once again, it may be that the nature of the *continuous success* condition, where anything is reinforced, paradoxically makes the interpretations participants make of the opponent more uncertain. This is in contrast to the *ascending* condition, which reinforce participants' play patterns more slowly after an initial state of failure, potentially giving a stronger signal of an opponent that can be exploited in a specific way. However, if this is the case, it is not reflected as increased *stayshift* behaviour – it may thus simply be that participants are more likely to consider the ascending condition as the most "lucky".

Finally, the results from the additional confidence measures (self-terminating play and the off-line prediction measure) indicated an effect of late success on self-terminating play, and an effect of both early and late success on off-line predictions of success. Participants were more likely to play extra rounds in late success conditions, and predicted higher wins in both late and early success conditions. Note that the latter analysis suffers from some ambiguity, as averaging over late or early success conditions in effect means creating an average prediction score for the average between a 50% win-rate condition *(ascending* or *descending)* and an above-chance win-rate condition that ends with high wins *(continuous success)*. This means that both main effects on the off-line prediction score could still be mainly driven by late outcomes. The fact that the ascending condition did not stand out in these analyses undermines the notion of its specialty: there is a disconnect between participants' confidence in Play Point 2 and their confidence of future wins. Nevertheless, it is worth noting that despite having no extra financial incentive, participants did continue playing a very simple game in predictable ways, validating the self-terminating play measure.

In sum, the results of Experiment 3 do replicate several earlier studies, but are inconclusive as to why participants were overconfident in different situations, and why the rate of *stayshift* behaviour varied the way it did. The difference in win-rates between conditions both globally and locally (during Play Point 2) is a potential confound to some of the analyses. Would participants still have made overconfident predictions in the ascending condition if Play Point 2 saw a return to a local 50% win-rate? How much of the effect is caused by whether the rate of wins prior to any given Play Point was increasing or decreasing? I attempted to address the win-rate confound in Experiment 4, along with replicating the results of Experiment 3.

3.3 Experiment 4

3.3.1 Introduction. Experiment 4 was an attempt to replicate the results of Experiment 3, while correcting for some of the potential confounds in Experiment 3. In Experiment 4, I still used different win-rate trajectories, but this time each trajectory had an overall win-rate of 50%, ensuring that any potential differences in behaviour would be solely due to the order of wins and losses and not their rate. Additionally, Play Points 1 and 2 both had a local win-rate of 50%, thus maintaining the effect of local win-rates constant for each analysis.

The success slope conditions used in Experiment 4 consisted of a *descending, ascending* and a *baseline* (or flat) success slope condition. The *descending* and *ascending* conditions resembled the similarly named conditions from Experiment 3, with slightly modified slopes (see Figure 3.2). Most importantly, the success slopes now end, in Play Point 2, with a local 50% win-rate in each condition to address the issue of local high and low win-rates in Experiment 3. Alterations were made to the *ascending* and *descending* slopes to ensure that the slopes prior to and after Play Point 1 were mirror images of each other, and that the overall win-rate both prior to and after Play Point 2 was 50% in both conditions. This was done while ensuring that win-rate prior to Play Point 1 was above-chance in the *descending* condition and below-chance in the *ascending* condition, and vice versa for the win-rate of the trials between Play Points 1 and 2. The flat condition was included as a control for Play Points 1 and 2, with the win-rates prior to each Play Point being 50% in this condition. Additionally, I included a condition with no fixed Play Points, an overall win-rate of 50%, and a fully randomized order of wins and losses to function as an additional control condition for the effect of success slopes. Outside of these amendments, the design of Experiment 4 and the measures included were identical to that of Experiment 3 (see section 3.2).

3.3.1.1 Hypotheses. My hypotheses for Experiment 4 were:

1) A replication of a *win-stay* and *lose-shift* bias in all conditions (as per Experiment 3)

2) No post-error slowing in any condition (as per Experiment 3)

3) Highest rates of win predictions in the ascending condition (as per Experiment 3)

4) Initial overconfidence during Play Point 1 in the ascending condition will be higher than in the descending condition (as per Experiment 3)

3.3.2 Method.

3.3.2.1 Participants. Fifty-five participants (N = 55; 46 female; M_{age} = 19.60; SD_{age} = 1.33) were recruited from the University of Sussex participant pool. Recruitment was continued until there were six viable participants for each eight counterbalancing orders of the experimental conditions (see section 3.3.2.3): seven of the fifty-five participants were excluded due to indicating having correctly guessed what the experimental manipulation was in debriefing. Informed consent was obtained from all participants before testing, and the experiment was approved by the School of Psychology at the University of Sussex (ER/BJD21/26). Participants received course credit as reward for participation.

3.3.2.2 Materials.

3.3.2.2.1 Game trials. Static pictures of one penny coins (depicting the choices made by the participant and the computer opponent) were presented on screen at 12º × 6º, with participants sat approximately 57 cm away from a 22" Diamond Plus CRT monitor (Mitsubishi, Tokyo, Japan). The experiment was coded in Matlab (The MathWorks Inc., Natick, MA) using the Psychophysics Toolbox extension (version 3; Brainard, 1997; Kleiner, Brainard & Pelli, 2007; Pelli, 1997), and responses were recorded using a keyboard.

3.3.2.2.2 Questionnaires. At the end of the experiment, participants filled out the LoC measure previously used in Experiment 3 (Rotter, 1966). The 10-item Big Five Inventory (Rammstedt & John, 2007) and the Ego Resiliency scale (Block & Kremen, 1996) were included as additional covariate measures along with other studies run in the lab at the time, but were not considered important for the purposes of the present experiment. For the LoC, I used items 5, 10, 15, 18, 25 and 28 to form a sum score, per the results of the EFA I ran for the LoC data from Experiment 3. The LoC was used as in Experiment 3, as a covariate for the Play Point 3 and off-line prediction measure analyses.

3.3.2.3. Design. The study was a within-subjects design with four different success slope conditions with a 50% win-rate each: three conditions where the win-rates were manipulated to follow a certain trajectory (see Figure 3.2), and one condition where wins were random but fixed at 50% globally across the block. Therefore, at the end of each condition, all participants would have experienced an overall win rate of 50%. The order of the conditions was counterbalanced between participants so that there were eight counterbalancing orders in total. In four of these, the changing trajectory conditions (*ascending* and *descending*) followed each other and the control conditions (*baseline* and *random*) followed each other, starting with either the changing trajectory conditions or the control conditions. For the other four counterbalancing orders, a changing trajectory condition followed a control condition, followed then by a control condition and a changing trajectory condition, or vice versa. Each block consisted of 84 mandatory rounds of the game, followed by up to 24 optional rounds, leading to a maximum of 108 rounds per block, as in Experiment 3. Win-rates in the trajectory manipulation conditions ranged from 1/6 to 5/6 and were defined for consecutive bins of 6 trials separately; that is, for each bin, there was at least one and at most five wins. The order of wins and losses within each bin was randomized for each participant.

The *descending* trajectory started with an initial win-rate of 4/6 in the first bin, followed by two bins at 5/6, then one bin at 4/6, and one bin at 3/6, leading to an average win-rate of 70% across these five bins. After this, the win-rate would remain at 3/6 for the following two bins (Play Point 1). Following this, the win-rate would then begin to descend, with the first bin after Play Point 1 at 2/6, then two bins at 1/6, then one bin at 2/6, and one bin at 3/6, leading to an average win-rate of 30% across these five bins. After this, the win-rate would again remain at 3/6 for two bins (Play Point 2). For the *ascending* trajectory, the five bins before and after Play Point 1 were simply swapped. For both the descending and the ascending trajectory, the average win-rate before Play Point 2 and the overall average win-rate were thus all 50%.

For the *baseline* trajectory, the win-rate fluctuated around the baseline of 50%, never going above 4/6 or below 2/6, with Play Points 1 and 2 being identical to the other two trajectory conditions. During Point 3 win-rates were held at 50% for the 24 optional rounds (3 wins per 6 trials). The *random* condition had no trajectory. The overall win-rate of the block up until Play Point 3 was fixed at 50%, with win and lose trials fully randomized for the 84 trials. To maintain similarity between the other conditions, trials at Play Point 3 were fixed to follow the trend in the trajectory conditions, with bins of 6 trials each having 3 randomly allocated wins. This condition was included as a control for reaction time, reinforcement bias, self-terminating play and off-line confidence measures, having the same overall win-rate but no trend to the occurrence of wins and losses. This allowed me to better differentiate the potential effects of patterns in wins and losses from the frequency of wins and losses.

3.3.2.4. Procedure. The experimental procedure for the game trials and instructions given to participants were identical to Experiment 3. After finishing all their game rounds, participants filled out the three questionnaires used, after which I debriefed them and thanked them for their time.

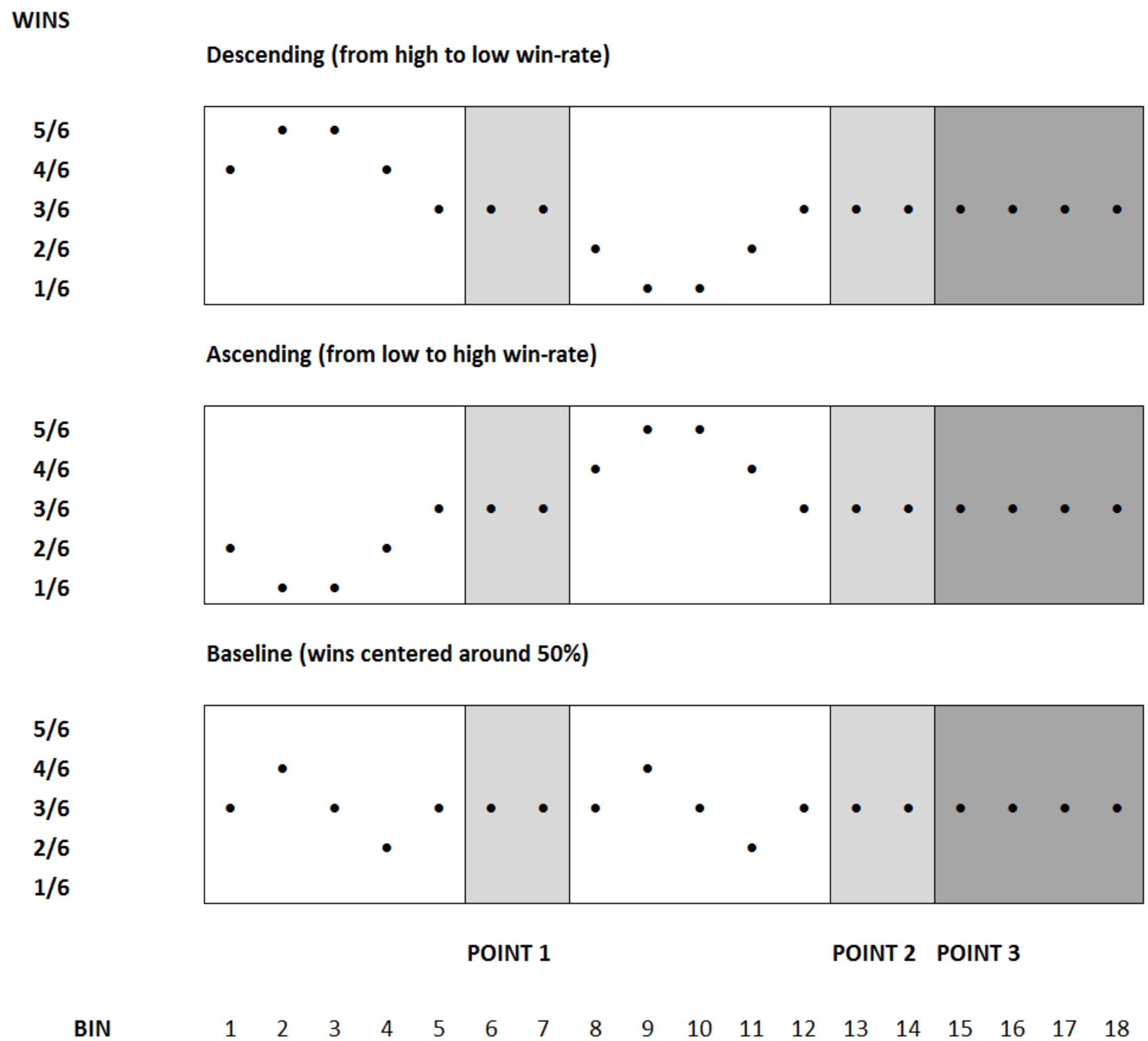

Figure 3.2. The win-rate trajectories of the experimental conditions in Experiment 4. Each dot represents a bin of 6 game rounds.

3.3.3 Results.

3.3.3.1 Reinforcement biases. I analysed *win-stay* and *lose-shift* biases in two ways, following Experiment 3: first, by comparing overall percentages of *win-stay* and *lose-shift* behaviour in each condition, and second, by using independent samples z-tests of proportion to calculate a binary variable for each participant in each condition denoting whether they had a significant bias in any direction. I analysed these measures with a GLMM with a logit link, with the experimental condition *(descending, ascending baseline, random)* entered as the within-subjects factor and participants entered as a random effect. The *random* condition was fixed as reference for both analyses. I predicted an overall trend of *win-stay* and *lose-shift*, as in Experiment 3. I used the lme4 and emmeans packages in R, version 3.5.1. See Table 3.5 for descriptives.

Table 3.5. Win-stay and lose-shift decisions and likelihoods of individual participants having a win-stay or lose-shift bias (back-transformed estimated marginal means) in Experiment 4

Success slope	*WS overall*	*WS individual*	*LS overall*	*LS individual*
Descending	63.2 (2.3) *	35.69% (9.37%)	55.79 (2.39) *	21.89% (7.89%)
Ascending	63.5 (2.3) *	44.43% (9.93%)	55.02 (2.39) *	19.56% (7.43%)
Baseline	58.0 (2.4) *	27.46% (8.41%)	55.24 (2.39) *	11.42% (5.38%)
Random	62.7 (2.3) *	44.43% (9.93%)	54.41 (2.40)	21.88% (7.89%)

Note: standard error in parentheses. Asterisks indicate that the expected percentage of decisions (50%) falls below the 95% CI of the estimated marginal mean of observed decisions. Likelihoods of individual biases based on results obtained using binary variables based on the z-test.

3.3.3.1.1. Win-stay. For the overall rate of *win-stay* decisions, the model indicated no significant differences between the *random* condition and the *descending* condition (β = -0.02, SE(β) = 0.07, z = 0.30, *p* = .764) or the *ascending* condition (β = 0.03, SE(β) = 0.07, z = 0.47, *p* = .637), but there was a significant difference between the *random* and the *baseline* conditions (β = -0.20, SE(β) = 0.07, z = -2.97, *p* = .003). Tukey-corrected pairwise comparisons yielded similar results, and also indicated a significant difference between the *descending* and *baseline* conditions ($p < .05$) and between the *ascending* and *baseline* conditions ($p < .01$). All other comparisons were non-significant ($p > .05$ for all). The baseline condition had a significantly lower rate of *win-stay* behaviour (M = 58.0%, SE = 2.4%) than the other conditions (see Table 3.5). The lower confidence level of the 95% confidence interval for win-stay was above 50% in each condition as in Experiment 3.

For the binary individual bias variable, the model indicated no significant differences between the *random* condition and the *descending* condition ($\beta = -0.36$, $SE(\beta) = 0.48$, $z = -0.56$, $p = .578$), the *ascending* condition ($\beta = -0.17$, $SE(\beta) = 0.48$, $z = 0.00$, $p = 1$) or the *baseline* condition ($\beta = -0.75$, $SE(\beta) = 0.50$, $z = -1.51$, $p = .131$). Tukey-corrected pairwise comparisons yielded similar results and indicated no other significant differences ($p > .05$ for all comparisons; see Table 3.5 for back-transformed likelihoods of individual-level bias). The results of this analysis did not change if I replaced the binary variable based on the z-test with a binary variable based on the more conservative exact binomial test. The results replicate the results of Experiment 3 in that there was a significant bias in each condition, with no difference between condition in the likelihood of an individual expressing a significant bias. The result of the baseline condition having the lowest rate of *win-stay* behaviour does not align with the results of Experiment 3, as there the lowest rate of *win-stay* behaviour was in the *descending* condition.

3.3.3.1.2. Lose-shift. For the rate of lose-shift decisions, the model indicated no significant differences between the *random* condition and the *descending* condition ($\beta = 0.06$, $SE(\beta) = 0.07$, $z = 0.85$, $p = .398$), the *ascending* condition ($\beta = 0.03$, $SE(\beta) = 0.07$, $z = 0.37$, $p = .711$) or the *baseline* condition ($\beta = 0.03$, $SE(\beta) = 0.07$, $z = 0.51$, $p = .612$). Tukey corrected pairwise comparisons yield similar results and found no other pairwise differences ($p > .05$ for all). The lack of significant difference in lose-shift across the four conditions was also consistent with Experiment 3. The lower confidence level of the 95% confidence interval for lose-shift was above 50% in each condition except the random condition (see Table 3.5). This is in slight contrast to Experiment 3, where an overall *lose-shift* bias was observed in each condition. The numerical trend is similar for Experiment 3 and the present experiment, with lower rates of *lose-shift* than *win-stay* responding, and *lose-shift* responding thus being closer to chance.

For the binary individual bias variable, the model again indicated no significant differences between the *random* condition and the *descending* condition ($\beta = 0.00$, $SE(\beta) = 0.53$, $z = 0.00$, $p = 1$), the *ascending* condition ($\beta = -0.14$, $SE(\beta) = 0.53$, $z = -0.27$, $p = .790$) or the *baseline* condition ($\beta = -0.78$, $SE(\beta) = 0.57$, $z = -1.37$, $p = .172$). Tukey corrected pairwise comparisons yield similar results and found no other pairwise differences ($p > .05$ for all). The results of this analysis did not change if I replaced the binary variable based on the z-test with a binary variable based on the more conservative exact binomial test. The results replicate Experiment 3, with no differences in the rates of individuals expressing a *lose-shift* bias between conditions, and the likelihood of an individual *lose-shift* bias being numerically lower than that of an individual *win-stay* bias (see Table 3.5).

3.3.3.2 Reaction times. I entered median reaction times for decisions at trial *n+1* in each condition into a two-way repeated measures ANOVA with the experimental condition (*ascending, descending, baseline, random*) and outcome at trial *n* (*win, lose*) entered as factors. Five participants were excluded from the analysis due to a block median reaction time more than two times the block average median reaction time in one block, leading to a final sample size of 43 for the reaction time analysis. I used the ez and emmeans packages in R, version 3.5.1. Mauchly's test indicated violations of sphericity, and I used the

Greenhousse-Geisser correction for effects with violations. There was no significant main effect of condition [$F(3,126) = 0.26$, $MSE = 81324$, $p = .857$, $\eta_p^2 < .01$] and no significant main effect of outcome type [$F(1,42) = 1.15$, $MSE = 12245$, $p = .289$, $\eta_p^2 = .03$], but there was a significant interaction effect [$F(2.49, 104.74) = 3.71$, $MSE = 10394$, $p = .020$, $\eta_p^2 = .08$]. However, Tukey corrected pairwise comparisons found no significant pairwise differences ($p > .05$ for all comparisons), suggesting no post-error slowing and no overall reaction time differences between conditions (see Table 3.6 for descriptive statistics) as in Experiment 3. The numerical trend was towards post-error slowing in the ascending condition and post-error speeding in the descending and baseline conditions.

Table 3.6. Average median reaction times (milliseconds) in Experiment 4, N = 43

Success slope	*Win*	*Lose*
Descending	478 (34)	448 (39)
Ascending	463 (32)	497 (40)
Baseline	519 (36)	464 (29)
Random	457 (34)	457 (34)

Note: standard error in parentheses.

3.3.3.3 On-line confidence measures. I analysed on-line confidence measures as in Experiment 3, with separate analyses for overall rates of win predictions and a binary overconfidence bias variable. See Table 3.7 for descriptives.

Table 3.7. Win prediction rates and likelihoods of individual participants having an overconfidence bias (back-transformed estimated marginal means) in Experiment 4

Success slope	*PP1 overall*	*PP1 individual*	*PP2 overall*	*PP2 individual*
Descending	76.2% (4.2%) [70.0%]	17.5% (7.7%)	69.0% (5.0%) * [50.0%]	26.4% (8.9%)
Ascending	60.0% (5.5%) * [30.0%]	73.1% (9.7%)	71.7% (4.8%) * [50.0%]	37.5% (10.1%)
Baseline	74.7% (4.4%) * [50.0%]	31.0% (10.2%)	65.4% (5.3%) * [50.0%]	29.0% (9.3%)

Note: standard error in parentheses. Expected win prediction rate in square brackets; expected win-rate is equal to the win-rate of all trials prior to that Play Point. Asterisks indicate that expected percentage of win predictions falls outside the 95% CI of the estimated marginal mean of observed win predictions. Likelihoods of individual biases based on results obtained using binary variables based on the z-test.

3.3.3.3.1 Play Point 1. I entered win prediction rates from the 12 trials in Play Point 1 into a GLMM with a logit link function, with the experimental condition *(ascending, descending, baseline)* entered as the predictor and with participants entered as a random effect. I excluded the *random* success slope condition from this analysis as it had no such subsections with controlled win-rates as the other conditions. The model fixed the *baseline* condition as reference. The model indicated no significant difference between the *baseline* condition and the *descending* condition ($\beta = 0.08$, $SE(\beta) = 0.23$, $z = 0.58$, $p = .559$) and a significant difference between the *baseline* condition and the *ascending* condition ($\beta = -0.68$, $SE(\beta) = 0.14$, $z = -4.85$, $p < .001$). Tukey corrected pairwise comparisons for back-transformed estimated marginal means yielded similar results, and also indicated a significant pairwise difference between the *descending* and *ascending* conditions ($p < .001$). The *ascending* condition had a lower win prediction rate (M = 60.0%, SE = 5.5%) than the *baseline* condition (M = 74.7%, SE = 4.4%) or the *descending* condition (M = 76.2%, SE = 4.2%), consistent with the *ascending* condition having the lowest actual win-rate of the three conditions. I assessed overall bias as in Experiment 3, by checking if the participants' average predictions in Play Point 1 matched with the win-rate they had experienced prior to Play Point 1. The 95% confidence interval for the prediction rates contained the prior experienced win-rate in the *descending* condition but not in the *ascending* or *baseline* conditions. In both the *ascending* and *baseline* conditions, the lower confidence level of the observed win prediction rate was above the prior expected win-rate. This result indicates that participants were overall over-predicting wins in these conditions (see Table 3.7), but probability-matching in the *descending* condition, the condition with the highest win-rate prior to Play Point 1. This is in contrast to Experiment 3, where participants over-predicted wins during Play Point 1 in the *descending* condition as well. Overall, the data point to a rough approximation of probability matching when win-rates are high, but general over-confidence when win-rates are low.

I calculated the individual level overconfidence bias variable as in Experiment 3, with a two-sided independent samples z-test of proportion: participants with win prediction rates significantly higher than predicted were classified as having an overconfidence bias. The null hypothesis probabilities were 30% for the *ascending* condition, 70% for the *descending* condition and 50% for the *baseline* condition. In the GLMM on individual level bias, the *descending* condition did not differ significantly from the *baseline* ($\beta = -0.75$, $SE(\beta) = 0.56$, $z = -1.34$, $p = .181$), but the *ascending* condition did ($\beta = 1.80$, $SE(\beta) =$

0.60, z = 2.99, *p* = .003). Tukey corrected pairwise comparisons yielded similar results, and additionally found a significant difference between the *ascending* and *descending* conditions (p < .001). The *ascending* condition had a significantly higher likelihood for an individual to make overconfident predictions (M = 73.1%, SE = 9.7%) than the *descending* condition (M = 17.5%, SE 7.7%) or the *baseline* condition (M = 31.0%, SE = 10.2%). I conducted the same analysis using a binary variable calculated using the more conservative exact binomial test instead of the z-test for the win prediction rates of each participant separately. Here, the difference between the *ascending* and *baseline* conditions was non-significant. There was still a numerical trend of the highest likelihood of individual overconfidence bias in the *ascending* condition (M = 46.8%, SE = 11.9%) compared to the *descending* (M = 16.3%, SE = 7.7%) and *baseline* (M = 31.0%, SE = 10.2%) conditions. Together, the results indicate that the *ascending* condition had both the highest general win prediction rate and the highest number of participants making over-confident predictions, in line with the results of Experiment 3.

3.3.3.3.2 Play Point 2. I analysed the rate of win predictions during Play Point 2 similarly to Play Point 1. The model fixed the baseline condition as reference, and indicated no significant differences between the baseline and descending conditions (β = 0.16, SE(β) = 0.14, z = 1.14, *p* = .256) but a significant difference between the baseline and ascending conditions (β = 0.29, SE(β) = 0.14, z = 2.07, *p* = .038), with the ascending condition having a higher win prediction rate (M = 71.7%, SE = 4.8%) than the baseline condition (M = 65.4%, SE = 5.3%). However, Tukey corrected pairwise comparisons for back-transformed estimated marginal means found no significant pairwise differences for any of the comparisons (*p* > .05 for all). In each of the conditions, the lower confidence level of the observed win prediction rate was above the prior experienced win-rate (see

Table 3.7). The finding of overall overconfidence matches Experiment 3, but the finding that the ascending condition did not differ significantly from the descending condition is contrary to the hypotheses, and to the results of Experiment 3.

I calculated and analysed the binary overconfidence bias variable similarly to Play Point 1 (the null hypothesis of the z-test was 50% for each condition). The model fixed the baseline condition as reference, and found no difference between the baseline and the descending condition ($\beta = -0.13$, $SE(\beta) = 0.52$, $z = -0.26$, $p = .797$) or the baseline and the ascending condition ($\beta = 0.38$, $SE(\beta) = 0.51$, $z = 0.76$, $p = .451$). Tukey-corrected pairwise comparisons yielded similar results and indicated no other significant differences ($p > .05$ for all comparisons). I conducted the same analysis using a binary variable calculated using the more conservative exact binomial test instead of the z-test for the win prediction rates of each participant separately. The z-test and binomial test categorized participants identically, so the results of the analysis were also identical. Again, this result differs from the results of Experiment 3, where participants were much more likely to have a bias in the condition most closely matching the ascending condition of Experiment 4. This is likely due to differences in the construction of win-rate trajectories between Experiments 3 and 4, since blocks with high local win-rate in Experiment 4 descended prior to Play Point. The numerical trend was still in the direction of the *ascending* condition having the highest likelihood of individual biases, but the difference between this condition and the others is noticeably smaller than it was in Experiment 3. See Table 3.7 for back-transformed likelihoods of individual participants making overconfident predictions.

3.3.3.4 Additional confidence measures. I analysed the number of extra rounds played in Play Point 3 (self-terminating play) and the number of extra rounds the participants predicted they would win at the end of the block as in Experiment 3; first on their own, then with the LoC measure as covariate, interpreting only the effects of the covariate from the ANCOVA (see Schneider et al., 2015).

3.3.3.4.1 Play Point 3. I entered the number of extra rounds played into a one-way repeated measures ANOVA with the experimental condition *(descending, ascending, baseline, random)* entered as the factor; see Table 3.8 for descriptive statistics. I expected to replicate the results of Experiment 3, with more voluntary rounds played for the ascending than the descending win-rate condition. Contrary to the expectation, there was no main effect of the experimental condition [$F(3,141) = .32$, $MSE = 51.36$, $p = .809$, $\eta_p^2 < .01$], suggesting that participants played similar numbers of extra rounds in each condition (grand mean = 11.7). Further, the result suggests that the differences in self-terminating play between the conditions in Experiment 3 were driven primarily by the trend in win-rate towards the end of the block, as blocks ending in a high local win-rate had a higher average number of extra rounds played. The ANCOVA with the LoC measure added as the covariate found no significant main effect for the covariate or interactions (every $p > .05$).

3.3.3.4.2 Off-line prediction. I analysed the number of trials participants predicted they would win if they were to keep on playing for another 50 rounds similarly to self-termination of play (see above). See Table 3.8 for descriptive statistics. I predicted that participants in the ascending condition would predict more wins (as per the main effect of late outcomes in Experiment 3). Mauchly's test indicated violations of sphericity, and I used the Greenhouse-Geisser correction for effects with violations. Contrary to my expectations, there was no significant main effect of experimental condition [$F(2.42, 113.81) = 0.20$, $MSE = 31.53$, $p = .862$, $\eta_p^2 < .01$], suggesting no differences in off-line win predictions between conditions (grand mean = 19.2). Further, the result suggests that the effects of both early and late outcomes on off-line win predictions in Experiment 3 were mostly due to differences in overall win-rates (absent in Experiment 4, as each block had a win-rate of 50%) or the local trend during Play Point 2. The ANCOVA with the LoC measure added as the covariate found no significant main effect for the covariate or interactions (every $p > .05$).

Table 3.8. Extra rounds and off-line predictions of future wins in Experiment 4

Trajectory	*Extra rounds*	*Off-line prediction*
Descending	11.9 (1.5)	18.9 (1.1)
Ascending	10.8 (1.5)	19.0 (1.2)
Baseline	12.0 (1.3)	19.4 (1.1)
Random	12.1 (1.4)	19.5 (1.1)

Note: standard error in parentheses.

3.4. General Discussion

In two experiments, I examined reinforcement biases, reaction times and different measures of confidence in different series of Matching Pennies games with manipulated success slopes. In general, the results suggest that participants do not increase their *stayshift* responding even if they achieve a high win-rate with this initial bias, nor do they increase it in success slope conditions that mimic learning (the *ascending* condition; see Matute, 1995; Ejova et al., 2013). In Experiment 3, the *descending* condition had significantly lower rates of *win-stay* behaviour than the *continuous success* condition (and numerically lower than any other condition), and in Experiment 4, the *baseline* condition had significantly lower rates of *win-stay* behaviour than all other conditions. Taken at face value, these results would fit with the notion that the *win-stay* response is modulated more easily than the *lose-shift* response (see Forder & Dyson, 2016; Scheibehenne et al., 2011, pp. 330-331). However, it is unclear why the modulation would have happened the way it did here. The reduction in *win-stay* (but not *lose-shift*) behaviour in neither experiment seems to be because of negative feedback, as in Experiment 3, *win-stay* was higher in the condition with more negative feedback, and in Experiment 4, total negative feedback was the same in each condition. Moreover, it does not seem to be clearly caused by a specific type of success slope either, as the *descending* condition of Experiment 3 is quite different from the *baseline* condition of Experiment 4, and the *descending* condition of Experiment 3 showed no such reduction in *win-stay* behaviour. In sum, there is no obvious explanation that would cover both of these observed reductions in *win-stay*.

Likewise, the single case of overall chance-level *lose-shift* responding in the *random* condition of Experiment 4 is not easily explained by the properties of the condition. It is specifically in conditions with random opponent behaviour and thus random outcomes that previous studies have observed *lose-shift* responding (Dyson et al., 2016; Forder & Dyson, 2016; Lee et al., 2004; Lee et al., 2005). Note, however, that *lose-shift* in all conditions across the two experiments was closer to chance than *win-stay*, and the one condition that did not reach significance was not significantly different from the others. In other words, this result may not need explaining other than statistical noise. Moreover, in both experiments, the likelihood of an individual participant expressing a *win-stay* or *lose-shift* bias was not significantly affected by the experimental conditions. It thus seems that the bias is generally quite stable in the people who actually express it, at least in the binary choice task, even without financial incentives. Note also that participants were more likely to express a *win-stay* than a *lose-shift* bias, and that in no condition did the likelihood of individuals expressing either bias significantly exceed 50% (the highest back-transformed mean of individuals with a *win-stay* bias: M = 50.07%, SE = 9.81, in Experiment 3). This suggests that the overall bias observed in the experiments was driven at most by roughly half the sample. This also means that the number of participants expressing both types of biases will necessarily be as low or lower than the number of people with a *lose-shift* bias, suggesting that a "true" *stayshift* bias may be quite rare – though the opposite *shiftstay* bias is clearly not as common or strong, as this would have shown in the group average data. This raises some questions about the suggested evolutionary origins of the *stayshift* bias and its status as a default decision rule (Scheibehenne et al., 2011; Wilke et al., 2014; Wilke & Barrett, 2009), as a sizable portion of the samples in Experiments 3 and 4 had no such bias.

Neither fixed win-rates nor success slopes seemed to affect median reaction times as a function of wins and losses. Participants exhibited no post-error slowing in any of the conditions in Experiments 3 and 4. The numerical trend in Experiment 3 was in the predicted direction, with higher lose than win reaction times in the continuous success condition, and the opposite trend in the continuous failure condition. The trend in Experiment 4 suggested post-error speeding in the baseline and descending conditions, but post-error slowing in the ascending condition. There is a potential methodological issue here masking a larger effect that may reach significance, as the delay introduced by the win prediction measures in these experiments may have masked slowing (see e.g. Ting et al., 2019, for a demonstration of delays in choice eliminating reaction time differences). The lack of any large reaction time differences could also be caused by a task switching effect: the participants have to constantly switch between game decisions and predictions of the outcomes of those decisions. Task switching can lead to notable slowing for the trials following a switch, as well as more errors on those trials (see Monsell, 2003, for a review). Since essentially every game trial in the current experiment save the first trial in each block followed a switch, the slowing caused by this may mask any slowing that would have been caused by other factors. While no decision in Experiments 3 and 4 could be considered an "error", the fact that task switching can increase errors suggests it may also have affected e.g. *stayshift* behaviour in these experiments. This issue will be addressed in Chapter 4.

The on-line confidence results from Experiment 4 did not fully align with the results of Experiment 3. The difference in win prediction rates during Play Point 1 was partially replicated, with participants overall predicting higher win-rates when the prior win-rate was high *(descending)* than when it was low *(ascending).* However, unlike in Experiment 3, the predicted win-rate in the high early win-rate condition *(descending)* was not significantly higher than the actual prior experienced win-rate in that condition. As in Experiment 3, the likelihood of individuals having an overconfidence bias in Play Point 1 was higher in the *ascending* than in the *descending* condition (the latter of which did not differ from *baseline*). However, while participants in Experiment 3 had the highest incidence of overconfidence bias and the highest overall rate of win predictions in the *ascending* condition during Play Point 2, this result did not replicate in Experiment 4. In Experiment 4, there was simply a general trend of overconfident win predictions in each condition during Play Point 2, and no significant differences in rates of biased participants between conditions. Although the numerical trend favoured the ascending condition, the likelihood of individual participants having an overconfidence bias was over twice as high in Experiment 3 (see Tables 3.3 and 3.7). Thus, while the general trend of group-level overconfidence replicated, the effect of success slopes on confidence in Experiment 3 did not find support in Experiment 4. This suggests that the local win-rate during Play Point 2 in Experiment 3 played a large role in win predictions and eliminating this confound in Experiment 4 thus eliminated the effect.

An additional possibility is that participants were sensitive not only to prior experienced win-rate but also the trend in win-rate prior to Play Point 2. That is, in Experiment 3, in the ascending condition, participants were experiencing an upward trend in win-rate prior to Play Point 2 (contrast Figures 3.1 and 3.2). However, the design of Experiment 4 was slightly different: here, participants had experienced wins above chance after Play Point 1, but the win-rate was on a downward trend towards Play Point 2 (where it stabilised into 50%). Unlike the notion that local win-rate was the only crucial factor, the notion that local trend explains the results of Play Point 2 would also fit with the results of Play Point 1 in both experiments. That is, in the conditions where Play Point 1 followed a low win-rate, the trend was upward towards Play Point 1 (where it stabilised into 50% in both experiments), and this resulted in a higher rate of win predictions and higher rates of individual-level overconfidence. Likewise, participants may have been over-predicting wins prior to Play Point 2 when it followed a high win-rate, but reduced this as the trend turned downwards before Play Point 2.

Similarly to the results of Play Point 2, the results of the additional confidence analyses in Experiment 4 indicated no differences between conditions for self-terminating play or the off-line prediction measure. This suggests that the differences between conditions observed in Experiment 3 were due to differences in overall win-rate between the conditions and/or the local high win-rate in the *late success* conditions. Specifically, in Experiment 3, self-terminating play was predicted only by late outcomes (with more extra rounds played in *late success* conditions), whereas the off-line prediction of wins was affected by both early and late outcomes. However, as noted earlier, this dual main effect on the off-line predictions is confounded by the fact that the main effects of early or late outcomes in Experiment 3 necessarily compared the average of a 50% win-rate condition and an above-chance win-rate condition to the average of a 50% win-rate condition and a below-chance win-rate condition. Taken together with the results of the on-line confidence measures, these results suggest that both trial-to-trial predictions and decisions to keep on playing are highly dependent on local trends, whereas off-line predictions seem to be more sensitive to longer trends.

3.5 Conclusion

In sum, Experiments 3 and 4 replicated a general trend towards *stayshift* on the group level, while indicating that the number of individuals who actually have both a *win-stay* and a *lose-shift* bias was quite small. Additionally, neither increased win-rates nor success slopes mimicking learning seem to increase the rate of the bias or lead to post-error slowing. When it comes to the illusion of control, the clearest trend within Experiments 3 and 4 is that people may be generally slightly overconfident given their prior experienced win-rates, but also that the overconfidence in the aggregate data may be at times driven by a minority of participants. The results also suggest that participants react to local trends in win-rates in a logical way, such that at times their overconfidence seems to be a rational reaction to an upward trend in win-rate. Due to the potential interruption effect of the trial-to-trial confidence measures on reaction times and potentially stayshift behaviour, Experiment 5 will remove the confidence measures and attempt to replicate the results of the previous four experiments. The questions still stand: why did participants in Experiments 1 and 2 not express a *stayshift* bias, given its ubiquity and replication in Experiments 3 and 4, and why did increased win-rates not induce post-error slowing?

CHAPTER 4: Comparing RPS and MP

4.1 General Introduction

In Experiments 1-2 (Chapter 2), I observed no *stayshift* bias in RPS (a three-option choice task). In Experiments 3-4 (Chapter 3), I observed biased *stayshift* behaviour in MP (a binary choice task), with more *win-stay* than *lose-shift* behaviour. Thus, the results of Experiments 3-4 align with prior studies of reinforcement biases in binary-choice tasks (see e.g. Achtziger et al., 2015; Scheibehenne et al., 2011). However, the results from Experiments 1 and 2 do not align with prior studies on reinforcement biases in RPS (Dyson et al., 2016; Forder & Dyson, 2016) reporting a *lose-shift* but not a *win-stay* bias. It is currently unclear why this result was not replicated in Experiments 1 and 2.

More broadly, the experiments reported in the previous chapters and prior studies converge in arguing for a difference in the rates of *win-stay* and *lose-shift* behaviour between two simple game types (MP and RPS). In Experiments 5 and 6, I attempted to firstly replicate the results of the previous chapters, and secondly to examine further the differences between the two game types in terms of the flexibility of *win-stay* and *lose-shift* biases. In Experiment 5, participants played both RPS and MP with manipulated win-rates (as per Chapter 3), winning either 33% of the time or 50% of the time, with the outcomes distributed randomly across blocks. I focused on manipulated win-rates in order to test if the lack of post-error slowing in Experiments 3-4 when win-rates were above-chance was simply caused by a disruption by the additional confidence task, and if the baseline win-rate of each game type may have an effect on *stayshift* behaviour. In Experiment 6, participants played both game types against both *unexploitable* and *exploitable* computer opponents (as per Chapter 2). Specifically, the *exploitable* opponents in Experiment 6 were opponents with above or below chance autocorrelation in their choices, but with no specific pattern of shifts in the case of the negatively autocorrelated patterns, unlike the exploitable opponents in Experiments 1-2. Increased or decreased autocorrelation in an opponent's choice patterns can be exploited by either increasing or decreasing *stayshift* behaviour, respectively. Given the differences in rates of *win-stay* and *lose-shift* between the game types, this design allowed for testing whether participants would also learn at different rates, or find it easier to increase a specific response type in one game over another.

In sum, the aim of Chapter 4 was to compare behaviour in RPS and MP within the same samples, in order to address the differences in design between Chapters 2 and 3, and help identify some of the factors behind the differences between the two game types. Specifically, Experiment 5 attempted to replicate the results observed in MP with fixed win-rates (Chapter 3) as well as extend this paradigm to RPS. Experiment 6 attempted to replicate in the results observed in RPS with *exploitable* and *unexploitable* opponents (Chapter 2) as well as extend this paradigm to MP. Comparing the two game types in the same paradigm allowed me to eliminate any potential effects any differences in e.g. the chosen method of randomization of outcomes or opponent choices (randomization of a flat distribution of opponent choices in Experiments 1-2 and randomization of a flat distribution of outcome types in Experiment 4) in the experimental designs of the previous chapters may have had on results.

4.2 Experiment 5

4.2.1 Introduction. Taken together, the results of Experiments 1-4 seem to support a difference in the rates of reinforcement biases between RPS and MP. The result of a *win-stay* bias in MP but no such bias in RPS has some support from earlier RPS studies (Dyson et al., 2016; Forder & Dyson, 2016) that found no reliable *win-stay* trend, and earlier studies using different binary choice tasks (similar to MP) finding a general *stayshift* bias (Achtziger et al., 2015; Scheibehenne et al., 2011; Wilke et al., 2014). However, there are also studies of RPS (Lee et al., 2005, in monkeys) or RPS-like games (Alós-Ferrer & Ritschel, 2018, in human players) that have found a *win-stay* bias as well. Thus, there is some precedence of binary choice games and three-choice games differing in terms of reinforcement biases, but it is not fully clear why, and the literature is conflicting. In Experiment 5, I attempted to replicate this finding and test whether the *baseline win-rate* of binary or three-choice zero sum games might affect the expression of reinforcement biases.

Given an equal number of rounds and assuming roughly flat distributions of each outcome type, a participant playing RPS would experience fewer wins than a participant playing MP or a similar two-response game (33.33% vs. 50.00%). Thus, repeating prior winning moves in RPS simply "leads" to fewer wins than in MP purely due to the structure of the game. Further, while the distribution of outcomes in both games is flat, this may not be how RPS is perceived, based on event-related potential (ERP) literature and the neural interpretation of draws as measured by feedback-related negativity (FRN). Specifically, the FRN response to neutral outcomes (draws) seems to be similar to the response to losses (Holroyd et al., 2006). This would imply that from the reinforcement learning system's perspective, the typical situation in RPS does not actually contain an equal number of wins and losses, but rather 1/3 positive outcomes (wins) and 2/3 negative outcomes (losses and draws). Therefore, a possible explanation for the lack of a reliable *win-stay* trend in Experiments 1-2 and in earlier RPS experiments where players experienced roughly equal distributions of outcomes (Dyson et al., 2016; Forder & Dyson, 2016) could be that the *win-stay* response, being more flexible, is sensitive to the frequencies of outcomes. In the typical situation of flatly distributed outcomes in MP, a participant using a *win-stay* rule would be rewarded for the use of this rule more often than a participant using a similar rule in RPS. It may be that *win-stay* is a heuristic people initially hold but that is more likely discarded in RPS than in MP due to less frequent wins using the rule – in other words, the reinforcement rule of *win-stay* is simply not itself reinforced as much. The results of Experiment 3 yielded no support for this hypothesis in MP, but the results of those experiments may be confounded by a task-switching effect due to the confidence measures used on each round (see Monsell, 2003).

The aim of the present study was to examine the effects of game type (RPS or MP) and outcome frequencies on the win-stay heuristic. To replicate previous results from Experiments 1 and 2 (RPS) and Experiments 3 and 4 (MP), I included blocks of games where participants played RPS or MP with the game outcomes fixed so that each participant would experience a flat distribution of all outcome types, randomized across the block, leading to 1/3 of each outcome type in RPS and 1/2 of each outcome type in MP. This would serve as a test of the baseline biases of both games against each other. It is unclear why Experiments 1-2 failed to find any reinforcement biases in the RPS conditions where opponent responding was random and outcomes were on average at chance-level, unlike in prior studies of biases in RPS (Dyson et al., 2016; Forder & Dyson, 2016; Lee et al., 2005). Given different randomization methods and general designs between the RPS (Experiments 1-2) and MP (Experiments 3-4) experiments reported here, Experiment 5 served to eliminate these as potential explanations for differences in bias.

To test the hypothesis that the *win-stay* heuristic is sensitive to the frequency of wins with unexploitable opponents, I included two other conditions where the expected win frequencies of RPS and MP were switched, leading to a block of MP where a participant would experience 1/3 win trials and 2/3 lose trials (as in standard RPS), and a block of RPS where a participant would experience 1/2 win trials and 1/2 of lose and draw trials in total (as in standard MP; see Table 4.1). A main effect of win frequency in absence of other effects would indicate that the *win-stay* heuristic is sensitive to wins; a main effect of game type in absence of other effects would similarly indicate that the difference in the use of reinforcement rules between the games is due to some other difference between the games, such as the presence of a third option in RPS. While the results of Chapter 3 do not give much reason to assume an effect of manipulated win-rates at least in MP, a final replication of the results of Chapter 3 in both game types without task interruptions and using identical methods is relevant. Additionally, this allows for checking whether manipulated win-rates that are randomized rather than presented in fixed success slopes (as they were in Chapter 3) produce different kinds of behaviour in participants.

Table 4.1. Distribution of outcomes in the experimental conditions in Experiment 5

	Win	Lose	Draw	Total
RPS				
50.00% Win-rate	45	23	22	90
33.33% Win-rate	30	30	30	90
MP				
50.00% Win-rate	45	45	NA	90
33.33% Win-rate	30	60	NA	90

The present study was run without financial incentive for two primary reasons. First, in order to maintain similarity between previous experiments in order to be able to compare the results; and second, because an incentive could compromise the interpretation of results. Achtziger et al. (2015) showed that in a Bayesian updating binary choice task, a higher monetary incentive produced a correlation between feedback-related negativity (FRN) amplitudes and reinforcement error rates – that is, the rate of *stayshift* decisions when the opposite decision rule (i.e. *shiftstay*) would yield better results in the task. Monetary incentives may also affect FRN amplitudes on trials with no financial incentives if both trial types are carried out during the same experiment (Ma et al., 2013). This would make comparing an incentivized and a non-incentivized condition in a within-subjects design difficult. Given these issues and that my primary aim was to examine how reliance on reinforcement rules may differ between the two game types (MP vs. RPS) or between situations with different win frequencies, leaving financial incentives out of the equation seemed like the best option. Given that reinforcement biases were observed in Experiments 3-4 and in Dyson et al. (2016) and Forder & Dyson (2016), all without financial incentive, it seems that money is not necessary for reinforcement-based deviations from randomness in zero-sum games. Additionally, as the *win-stay* heuristic specifically seems to be sensitive to changes even in non-monetary outcome value in contrast to the *lose-shift* heuristic (Forder & Dyson, 2016), it may be that at least some of the instances of *win-stay* bias observed in the literature may be due to incentives/value rather than the heuristic being a default decision rule.

In an attempt to resolve inconsistencies between the observations of post-error speeding/slowing across Chapters 2 and 3, I also examined reaction times as a function of outcome type. Post-error slowing was observed in Experiments 1 and 2 as a function of win-rate in the *exploitable* condition. This was in contrast to the *unexploitable* conditions in Experiments 1 and 2, where win-rates were on average at chance-level, showing post-error speeding. The modulation of reaction times as a function of outcome failed to replicate in Experiments 3 and 4, with no post-error slowing or speeding in any condition, despite participants in Experiments 3-4 experiencing both chance-level and above-chance win-rates (as in Experiments 1-2). In Experiment 5, I wanted to re-test post-error reaction time effects without the intervening confidence measurement that was included in Experiments 3-4 after each game choice, which introduced a delay and may have thus eliminated reaction time effects. A delay between response and outcome feedback may reduce post-error slowing compared to tasks with no such delay (Dudschig & Jentzsch, 2009; Jentzsch & Dudschig, 2009; see also Danielmeier & Ullsperger, 2011). If response interruptions in Experiments 3 and 4 (relative to Experiments 1 and 2) were the cause of eliminating RT differences as a function of outcome, then using a fixed and no-delay design in Experiment 5, I expected to see post-error slowing in the conditions with a higher win-rate in the present experiment, and post-error speeding in the lower win-rate conditions.

4.2.1.1 Hypotheses. My hypotheses for Experiment 5 were:

1) Win-stay bias more common than lose-shift bias in MP (as per Experiments 3-4)

2) An overall lose-shift bias in RPS (as per Dyson et al., 2016; Forder & Dyson, 2016)

3) An overall win-stay bias in MP (as per Experiments 3-4)

4) Post-error speeding in conditions with chance or below-chance level of wins

5) Post-error slowing in the one condition with above-chance level of wins (RPS with 50% win-rate)

4.2.2 Method.

4.2.2.1. Participants. Forty-eight (48) participants (38 female; M_{age} = 19.69, SD_{age} = 1.94) were recruited from the University of Sussex participant pool. Informed consent was obtained from all participants before testing, and the experiment was approved by the School of Psychology at the University of Sussex (ER/JS/753/9). Participants received course credit as reward for participation. None of the participants had taken part in Experiments 1-4.

4.2.2.2 Materials. Static photographs of gloved hands making the Rock, Paper and Scissors gestures and static pictures of one penny coins were presented on screen as per Experiments 1-2 and Experiments 3-4, respectively. Participants sat approximately 57 cm away from a 22" Diamond Plus CRT monitor (Mitsubishi, Tokyo, Japan). Stimulus presentation was controlled by MATLAB 2016 (The MathWorks, Inc.) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997), and responses were recorded using a keyboard.

4.2.2.3 Design. The study followed a 2x2 within-subjects design, with game type (RPS, MP) and win-rate (33.33%, 50%) as factors. The order of the four different conditions was fully counterbalanced among participants leading to 24 unique orders of the four conditions. I continued data collection until each of the 24 counterbalancing orders had two participants. Each condition consisted of 90 rounds of a game and the number of wins, losses and – in the case of RPS – draws was fixed based on the win-rate condition, and the order of occurrence of each outcome type was randomized within each condition (see Table 4.1).

Due to the blocks consisting of 90 rounds, the RPS condition with a 50% win-rate had a slightly uneven distribution of losses and draws: 45 wins, 23 losses and 22 draws. I did not consider this a major issue for analysing the data due to the fact that the number of non-win outcomes was still 50% and that the difference was very minor.

4.2.2.4 Procedure. At the beginning of the experiment, I explained the rules of the game and the structure of the experiment to each participant, allowing them to ask any questions about how to proceed. I instructed participants to try and maximize their score for each of the four blocks, and informed them that they would be playing against four separate opponents, one per each block. I also informed participants that there were no strict time limits to making choices, but that given the number of rounds they would be playing, they should not worry too much about a single choice. After giving the general instructions, I left the room. The experiment program started each block by informing the participant of the number of rounds and the type of game they would be playing in the block, followed by reminder of the rules of the game and the scoring.

Regardless of game type, each round began with the participant being prompted to make a choice, with no time limit to the decision-making. After the game choice was made there was a 500ms interval, after which the program presented the choices made by the participant and the opponent for 1000ms. After this, the text WIN, LOSE or DRAW was presented on screen for 1000ms, after which the participant's score was updated and the next round started after an interval of 1000ms. The score was increased by one point for each win and reduced by one point for each loss, with draws having no effect on the score.

After finishing all the game rounds, the experiment program informed the participant that the experiment was finished. I then debriefed the participant and thanked them for their time.

4.2.3 Results.

4.2.3.1 Reinforcement biases. Win-stay and lose-shift behaviour was analysed as in Experiments 3 and 4, with separate GLMMs first for the overall rate of a choice type, then with the same analysis on a binary variable denoting whether a participant had a significant bias. I analysed both game types within each model, as comparing the effects of win-rates regardless of game type, as well as examining the interaction between game type and win-rate, was crucial to one of the questions Experiment 5 is asking: do the baseline win-rates of RPS and MP lead to different rates of reinforcement biases?

This approach leads to potential main effects of game type that are "significant but not interesting".[1] That is, a statistically significantly higher rate of e.g. overall win-stay behaviour in MP than in RPS does not imply that participants are more biased towards win-stay in MP, as the baseline for bias is different in the two game types. However, splitting the analysis into two separate GLMMs, one per game type, would lead to issues I argue are greater: namely, they would prevent examining interaction effects or the main effect of win-rate regardless of game (the latter would then require four different models). Testing for an interaction effect with two separate models would require complicated equivalency tests, but these become unnecessary if the game types are included in the same model with an interaction term. Thus, one model is clearly more parsimonious.

[1] Note that this issue does not apply to the analysis of the binary individual bias variables.

Table 4.2. Win-stay and lose-shift decisions and likelihoods of individual participants having a win-stay or lose-shift bias (back-transformed estimated marginal means) in Experiment 5

Game	*Win-rate*	*WS overall*	*WS individual*	*LS overall*	*LS individual*
RPS	50.0%	32.5% (2.5%) [33.33%]	16.1% (6.9%)	75.0% (2.0%) * [66.7%]	25.3% (9.1%)
	33.3%	33.4% (2.6%) [33.3%]	12.1% (5.7%)	74.1% (2.0%) * [66.7%]	22.7% (8.6%)
MP	50.0%	67.6% (2.4%) * [50.0%]	50.2% (10.9%)	53.1% (2.3%) [50.0%]	17.8% (7.5%)
	33.3%	74.4% (2.2%) * [50.0%]	71.7% (9.3%)	51.4% (2.3%) [50.0%]	13.5% (6.3%)

Note: standard error in parentheses. Expected probabilities in square brackets. Observed probabilities are back-transformed using the emmeans package. Asterisks indicate that the expected rate of a strategy type falls below the 95% CI for the observed rate. Likelihoods of individual biases based on results obtained using binary variables based on the z-test.

4.2.3.1.1 Win-stay. I analysed the proportion of *win-stay* behaviour using a GLMM with a logit link function with game type *(RPS, MP)* and win-rate *(33.3%, 50.0%)* entered as predictors and participants entered as random effects. The model tested for both main effects and the interaction between the predictors. The model fixed the MP and 50% conditions as reference for their respective analyses and for the interaction. There was a significant main effect of game type (β = -1.47, SE(β) = 0.07, z = -21.94, $p < .001$) and of win-rate (β = 0.33, SE(β) = 0.08, z = 4.37, $p < .001$), as well as a significant interaction effect (β = -0.29, SE(β) = 0.11, z = -2.75, $p = .006$). Pairwise comparisons (Tukey) revealed significant differences between all conditions except for the comparison between the two RPS conditions ($p < .001$ for all other comparisons). See Table 4.2 for back-transformed probabilities. The difference between rates of *win-stay* behaviour between the RPS and MP conditions was to be expected even if participants exhibited no bias in either condition, given the different baseline levels for the two games (where completely random choice in RPS would lead to roughly 1/3 *win-stay* decisions whereas in MP it would lead to roughly 1/2 *win-stay* decisions). The rate of *win-stay* behaviour differed in the two MP conditions, with the 50.0% win-rate condition (M = 67.6%, SE = 2.4%) having a significantly lower rate of *win-stay* decisions than the 33.3% win-rate condition (M = 74.4%, SE = 2.2%). Notably, the rate of overall win-stay behaviour was non-biased in both RPS conditions, with the expected rate of 33.3% falling within the 95% CI, but both MP conditions had a significant win-stay bias, replicating the results of previous experiments. This suggests that *win-stay* behaviour is rarer in RPS than in MP. However, it does not seem that the effect is driven by a lower baseline win-rate in RPS, since this would imply that lowering the win-rate in MP should decrease *win-stay* behaviour as well. The results suggest the opposite, as the overall rate of win-stay in MP was *lower* when the win-rate was higher.

To examine differences in the likelihoods of individual *win-stay* bias between the conditions, I calculated a binary variable by testing each individual participant's proportion of *win-stay* behaviour in each condition against the game-specific chance level with a two-tailed one-sample z-test of proportion. I assigned this variable a value of 1 for participants whose proportion of *win-stay* behaviour was significantly higher than the baseline (50% for MP, 33.3% for RPS), and a value of 0 for participants with either a non-significant difference or a significantly lower proportion of *win-stay* choices. I then entered this variable as the dependent variable in the same GLMM as above. See Table 4.2 for probabilities of individual bias. There was a significant main effect of game type (β = -1.66, SE(β) = 0.57, z = -2.90, p = .004), a marginal main effect of win rate (β = 0.92, SE(β) = 0.53, z = 1.75, p = .08), and no significant interaction (β = -1.26, SE(β) = 0.79, z = -1.59, p = .111). The likelihood of participants expressing a *win-stay* bias was higher for MP (M = 61.0%, SE = 5.6% than for RPS (M = 14.1%, SE = 5.2%); a substantial difference. The marginal effect of win-rate was in the direction of higher rates of biased individuals in the 33% win-rate condition (M = 41.9%, SE = 5.8%) than in the 50% win-rate condition (M = 33.2%, SE = 7.2%). Thus, in addition to overall rates of *win-stay* behaviour, the trend of individual-level bias was also contrary to the hypothesis that lower win-rates decrease *win-stay* behaviour. I also conducted the same analysis using a binary variable calculated using the more conservative exact binomial test instead of the z-test for the rates of *stay* decisions after wins for each participant separately. The results of this analysis did not differ from the earlier analysis other than that the effect of win-rate was no longer marginal but p = .21. The numerical trend of a higher likelihood of *win-stay* biases in the 33.3% win-rate condition (M = 39.0%, SE = 6.2%) than in the 50% win-rate condition (M = 33.1%, SE = 7.3%) remained.

In sum, the overall rate of win-stay responding was significantly higher than what would be expected from completely random responding in MP but not in RPS. The likelihood of individuals expressing a statistically significant *win-stay* bias was higher for MP than for RPS, and the likelihood of individual-level bias in RPS was in general very low (under 20% in both win-rate conditions). This supports the notion arising from prior studies in this thesis and elsewhere that *win-stay* biases are generally more common for MP than for RPS. The findings that *win-stay* responding increased in MP when win-rate was lower than chance and that the marginal effect of win-rate on individual bias suggesting a higher likelihood of individual bias in the 33.3% win-rate condition,support flexibility in *win-stay*, but these results do not align with the results of the previous experiments.

4.2.3.1.2 Lose-shift. I analysed overall rates of *lose-shift* behaviour using a GLMM analysis identical to the one I used for *win-stay*. There was a significant main effect of game type (β = 0.97, SE(β) = 0.08, z = 11.93, p < .001), no significant main effect of win rate (β = -0.07, SE(β) = 0.57, z = -1.19, p = .24), and no significant interaction (β = 0.02, SE(β) = 0.11, z = 0.22, p = .83). The rate of *lose-shift* behaviour was significantly higher in RPS (M = 74.5%, SE = 1.8%) than in MP (M = 52.3%, SE = 2.2%). As for *win-stay*, a difference between the rates of *lose-shift* behaviour was expected even assuming fully random play given different baselines (66.7% for RPS and 50.0% for MP). However, there was a significant *lose-shift* bias in RPS, with the expected rate falling under the 95% CI for the observed rate, but no such bias in MP. See Table 4.2 for back-transformed probabilities.

I then ran the GLMM again with a binary variable coding for individual bias, similarly to the *win-stay* analysis (see Table 4.2). There was no significant main effect of game type (β = 0.45, SE(β) = 0.55, z = 0.81, p = .416), no significant main effect of win-rate (β = -0.33, SE(β) = 0.57, z = -0.57, p = .569), and no significant interaction (β = 0.18, SE(β) = 0.79, z = 0.23, p = .816). The results of this analysis did not change when I replaced the dependent variable with a binary bias variable that was based on the more conservative exact binomial test. In general, the likelihood of individuals expressing a lose-shift bias was relatively low for both game types (under 30% in all conditions), with a numerical trend toward higher incidences of bias in RPS.

In sum, the results suggest no effect of win rate on *lose-shift* behaviour. Moreover, the results do not indicate a difference between RPS and MP in the likelihood of people adopting a *lose-shift* bias, but nevertheless a significant overall *lose-shift* bias in the RPS conditions but not the MP conditions. It thus seems that while the number of people with the bias was similarly low for the two games, the biased participants expressed a stronger bias in RPS compared to MP. This replicates the findings of Experiments 3-4, with individuals with a *lose-shift* bias being a minority in MP and there being no differences between win-rate conditions in rates of individual *lose-shift* bias. Further, the result replicates the finding of an overall lose-shift bias in RPS (see Dyson et al., 2016), unlike Experiments 1-2. However, this overall bias in RPS was driven by a minority of biased individuals, unlike the win-stay bias in MP.

4.2.3.2 Reaction times. I analysed median reaction times (in milliseconds) for decisions at trial *n+1* separately for the two game types. I made the choice to separate the games for this analysis in order to see if the reaction time trend for wins, losses and draws in RPS from Experiments 1 and 2 would replicate (as there are no draws in MP, this analysis could not be conducted in one model). I entered median reaction times into a two-way repeated measures ANOVA with win rate (33.33%, 50.00%) and outcome at trial *n (MP: win, lose; RPS: win, lose, draw)* as factors for both separate analyses. Four participants were excluded due to having at least one block average median reaction time that was at least twice the group block average median (as in previous experiments), yielding a final sample of 44 participants for the reaction time analysis.

Table 4.3. Reaction times on trial *n* following wins, losses and draws on trial *n-1* (milliseconds) in Experiment 5, N = 42

Game	*Win-rate*	*Win*	*Lose*	*Draw*
RPS	50.00%	553 (44)	482 (41)	444 (28)
	33.33%	636 (63)	518 (39)	432 (24)
MP	50.00%	428 (20)	375 (18)	N/A
	33.33%	391 (17)	364 (16)	N/A

Note: standard error in parentheses.

For MP, there was no main effect of win-rate [F(1,43) = 1.79, MSE = 13993, p = .188, η_p^2 = .04], a significant main effect of outcome [F(1,43) = 35.82, MSE = 2024, p < .001, η_p^2 = .45], and a significant interaction effect [F(1,43) = 5.92, MSE = 1234, p = .019, η_p^2 = .12]. Post hoc comparisons revealed there were no differences within win reaction times or within lose reaction times between the two win-rate conditions (Tukey's HSD, p > .05 for all comparisons). Win reaction times were slower than lose reaction times for all comparisons except for the comparison between wins in the 33.3% win-rate block (391ms) and lose reaction times in the 50.0% win-rate block (375ms; p < .05 for all other win-lose comparisons). The interaction effect seems to have been driven by this lack of difference for one win-lose RT comparison. In general, the results indicate post-error speeding in both win frequency conditions and no overall effect of win frequency on RTs (see Table 4.3).

For RPS, Mauchly's test indicated violations of sphericity, and I used the Greenhouse-Geisser correction for effects with violations. The two-way repeated measures ANOVA found no significant main effect of win-rate [F(1,43) = 0.77, MSE = 111248, p = .386, η_p^2 = .02], a significant main effect of outcome [F(1.58, 67.71) = 14.55, MSE =

47889, $p < .001$, $\eta_p^2 = .25$], and no significant interaction effect [F(1.46, 62.89) = 1.88, MSE = 35545, $p = .170$, $\eta_p^2 = .04$]. Pairwise comparisons for the outcome condition revealed a significant difference between win and lose reaction times and win and draw reaction times (Tukey's HSD, p < .05 for both comparisons) and a marginal difference between lose and draw reaction times (Tukey's HSD, *p* = .07). Win reaction times (595ms) were slower than lose (500ms) or draw (438ms) reaction times, again indicating post-error speeding irrespective of overall win rate. Note also the numerical trend of longer reaction times in RPS compared to MP, similar to Experiments 1-2 compared to Experiments 3-4.

4.2.4 Discussion. Experiment 5 serves as a direct comparison between RPS and MP with respect to *win-stay* and *lose-shift* responding. A critical observation was that *stayshift* behaviour in RPS and MP under fixed chance level win-rates (33.3% wins, 50% wins, respectively) differed, as hypothesized. In general, there was an overall *win-stay* bias in MP but not RPS, and an overall *lose-shift* bias in RPS but not in MP. The results thus replicate earlier RPS studies (Dyson et al., 2016; Forder & Dyson) but not Experiments 1-2 in terms of reinforcement biases in RPS. The results also broadly replicate Experiments 3-4 in terms of the *win-stay* bias in MP being stronger than the *lose-shift* bias.

There was a significant difference in *win-stay* choices between the win-rate conditions in MP, but no such differences in RPS, and no differences in *lose-shift* choices between the win-rate conditions for either game type. Moreover, the significant difference observed between the win-rate conditions in MP was such that the rate of *win-stay* was *lower* when the win-rate was *higher* (50%). Similarly to the results of Experiments 3 and 4, this suggests that *win-stay* may vary more than *lose-shift*, but the bias does not simply increase through increased wins. Here, participants significantly increased their rate of *win-stay* in the only condition where the win-rate was *lower* than the baseline for the game.

Moreover, the marginal effect of win-rate on individual-level bias also seemed to suggest higher likelihoods of individuals adopting a *win-stay* bias when win-rate was lower. Taken at face value, these results could indicate that *win-stay* behaviour increases when a participant is losing more than they feel they should, but no such pattern was observed in similar situations in Experiments 3 and 4. In Experiment 3, the rate of win-stay responding was lower in the 50% win-rate *descending* condition than in both the 70% and 30% win-rate *continuous success* and *continuous failure* conditions. On the other hand, Experiment 4 had four conditions each with a 50% win-rate in different slopes or randomly distributed, and win-stay responding was lowest in the condition where the participants experienced an almost flat success slope with local win-rates remaining at 50% most of the time (even though Experiment 4 also contained a condition with a similar success slope to that of the descending condition in Experiment 3). Thus, each of Experiments 3-5 have shown variability in *win-stay* but not *lose-shift*, but there is no clear pattern to the variability of *win-stay* in these experiments, suggesting the variability in the experiments reported in this thesis may simply be statistical noise, or a result of some complex interaction that the present experiments are unable to examine.

There was also a trend towards post-error speeding in each condition, contrary to my hypothesis of finding less post-error slowing specifically in the higher win-rate conditions. This result is inconsistent with Experiments 3 and 4. Based on this, it seems likely that the additional task of responding to the confidence measures in Experiments 3 and 4 on each trial of the game (not present in Experiment 5) may have interrupted the task and thus any RT effects. However, the trend of post-error speeding in the MP conditions in Experiment 5 suggests that the interruption in Experiments 3-4 likely did not mask post-error slowing but rather post-error speeding (see Dudschig & Jentzsch, 2009; Jentzsch &

Dudschig, 2009). The trend observed in Experiment 5 matches with the trend observed in Experiments 1-2 in the unexploitable condition, where participants were on average winning at the baseline chance rate of the game (RPS). Note again that only the RPS x 50.00% condition in Experiment 5 had an above-chance win-rate, which nevertheless does not seem to have translated into post-error slowing due to losses becoming a rarer event (see e.g. Danielmeier & Ullsperger, 2011) and/or the participant perceiving the opponent as more exploitable. In Experiments 1 and 2, post-error slowing was predicted by the win-rate in the exploitable condition, but participants in these experiments could achieve win-rates much higher than 50.00% of the trials (up to 78.89% in Experiment 1 and 83.33% in Experiment 2). Thus, the manipulation in Experiment 5 may not have been strong enough to induce post-error slowing. If that is the case, however, this at the very least does not support a *linear* relationship between success rate and reaction times in situations where participants cannot do anything to increase their win-rate. A higher win-rate, and one that is achieved via lawful behaviour, may be needed to induce post-error slowing.

The overall biases in the two games seem to have stemmed from different numbers of people actually expressing a bias. For the *win-stay* bias, there was both a significant overall bias to choose *win-stay* in MP but not in RPS, as well as a higher likelihood for an individual to have a *win-stay* bias in MP than in RPS . For the *lose-shift* bias, there was a significant overall bias to choose *lose-shift* in RPS but not in MP . However, the likelihood of individuals having a *lose-shift* bias did not differ significantly between MP and RPS . Despite only a minority of the sample expressing the *lose-shift* bias on an individual level, these individuals had a strong enough *lose-shift* bias for it to show up as a significant overall bias in RPS. These results are somewhat complicated in terms of assessing the flexibility or lack thereof of the two types of reinforcement bias. In terms of participants actually adopting a bias, there seems to be more variability for *win-stay* than *lose-shift*. However, for the minority of participants who did adopt a *lose-shift* bias, the strength of this bias seems to have been different based on game type, suggesting that something in RPS causes *lose-shift* choices to be more frequent compared to the baseline than in MP. This, further, suggests some flexibility in the *strength of the lose-shift bias*, if not the *likelihood* of the bias itself. Further, the results of Experiment 5 support my inference from the results of Experiments 3-4 that a true *stayshift* bias (with both biased *win-stay* and *lose-shift* responding) is relatively rare, given the low likelihood of individuals expressing the *lose-shift* part of the bias.

These differences in reinforcement biases between the games also seem to be the result of factors other than their baseline win-rates, and thus there is something else about the games that cause these differences. If the baseline win-rates were the main factor behind the differences in player biases between the games, participants in similar win-rate conditions should have expressed similar biases regardless of game type. Specifically, participants in the 50.0% win-rate RPS condition (mimicking the baseline win-rate of MP), compared to the 33.3% win-rate RPS condition, should have had higher rates of *win-stay* choices and/or a higher likelihood of having an individual level *win-stay* bias. Similarly, participants in the 33.3% win-rate MP condition (mimicking the baseline win-rate of RPS), compared to the 50.0% win-rate MP condition, should have had higher rates of *lose-shift* choices and/or a higher likelihood of having an individual level *lose-shift* bias. One possibility is simply that the addition of one more choice type in RPS affects choice probabilities in favour of shifting after losses; participants may e.g. underestimate the likelihood that an opponent would repeat their choice after winning due to more than one shift option. Alternatively, the participant may be less likely to have a *win-stay* bias in RPS because the amount of information to track in RPS is greater than in MP due to more choice and outcome options, and this may reduce the biasing effects of memory (as argued by Rapoport & Budescu, 1992).

A second possibility (that will be explored in Experiment 6) is that the rate of autocorrelation (i.e. the rate of repetition) in the games affects the differences in win-stay and lose-shift biases (see Scheibehenne et al., 2011; Wilke et al., 2014, for examples of manipulating autocorrelation). That is, assuming an opponent that does not have an item bias, the baseline likelihood of e.g. a tails following a tails in MP is 50%, but the baseline likelihood of e.g. a rock following a rock in RPS is 33.33%. While this does not matter in terms of what the equilibrium strategy is because the likelihoods for repetition and different types of shifts are always equal assuming a non-biased opponent, it may matter in terms of how participants perceive the game task. As repeating one's own move after a win is optimal *if and only if* the opponent is also more likely to repeat their move in that situation, perceiving the opponent as more likely to repeat their moves in MP than in RPS may increase win-stay behaviour in MP. Manipulating the win-rate in the present experiment did not directly manipulate autocorrelation, as the randomization of outcomes across the blocks caused no systematic reward to any choice strategy. That is, while participants may have won more or less than expected in a given condition, this did not translate to the opponent repeating their moves more or less often. Directly manipulating autocorrelation by increasing or decreasing the rate of repetitions from its baseline rate would, on the other hand, affect the opponent's exploitability. In the case of decreased autocorrelation (less response repetition), following the *stayshift* rule would cause the players to lose more often than expected by pure chance. Conversely, in the case of increased autocorrelation (more response repetition), following the *stayshift* rule would cause the players to win more often than expected by pure chance. As McKay and Efferson (2010) have pointed out, to test if a supposed cognitive bias can actually be meaningfully called a bias, one should see if the bias shows up in situations when reducing the bias would actually lead to better performance. Therefore, Experiment 6 examined behaviour under different conditions of autocorrelation in a final examination of the variability and flexibility of standard reinforcement learning heuristics.

4.3 Experiment 6

4.3.1 Introduction. In Experiment 5, I replicated in a single sample the finding that RPS and MP, two simple zero-sum competitive games, induce different kinds of reinforcement biases in players when the opponents cannot be exploited. Specifically, RPS was characterized by a *lose-shift* bias, whereas MP was characterized by a *win-stay* bias. These biases did not seem to stem from the kinds of win-rates that are typical to the games when the opponent is playing randomly (33.3% for RPS and 50.0% for MP). Instead, they might stem from the degree of baseline response repetition in these games (also 33.33% for RPS and 50% for MP), i.e. the autocorrelation of the opponents' responses. In Experiment 6, I attempted to test whether the difference in biases between the two game types would also lead to differences in learning when participants played against exploitable opponents that either supported (positive autocorrelation) or punished (negative autocorrelation) the biases. The *stayshift* heuristic in general is optimal when an opponent's moves are positively autocorrelated, that is, when the likelihood of a repetition is greater than the likelihood of any type of shift: the opponent is more likely to repeat their choice no matter what, and thus repeating one's own winning moves and shifting away from losing moves leads to a higher chance of winning. When the opponent's moves are negatively autocorrelated, that is, when the likelihood of a repetition is lower than the likelihood of any type of shift, the player should respond by using the opposite *shiftstay* rule.

Scheibehenne et al. (2011) and Wilke et al. (2014) showed, in experiments using simulated (binary choice) slot machines with positive autocorrelation, negative autocorrelation and random patterns, that people have trouble detecting and exploiting negative autocorrelation unless it is quite high (up to 80% likelihood of alternation in a binary choice game). This suggested that the *stayshift* bias can be difficult to learn away from even when participants are incentivized to do so by the game structure. If it is the case that RPS and MP lead to different kinds of biases, this could also mean that players would have different kinds of difficulties when attempting to exploit an opponent with a negatively autocorrelated pattern. Additionally, the differences in biases could lead to differences in learning to exploit an opponent with a positively autocorrelated pattern. That is, if players tend towards a *lose-shift* bias in RPS, they may find it easier to increase their *lose-shift* responding relative to *win-stay* during positively autocorrelated conditions, and vice versa for MP. Scheibehenne et al. (2011) noted that their participants were better at applying the correct rule in their autocorrelation conditions after wins than after losses, consistent both with responses after wins being more flexible, but also with win-stay being the more common form of the bias in binary choice tasks.

Note that in the typical experimental situation of an opponent playing randomly, the two games have different baseline rates of autocorrelation. In RPS, the likelihood of a repetition is 33.33%, whereas in MP it is 50%. Thus, increasing or decreasing the rate of repetition must happen relative to the baseline. In Experiment 6, I included three autocorrelation conditions for each game type: baseline, positive and negative. I decided on a conservative step of 16.67% percentage points for increases and decreases of the repetition rate for both game types. This made it possible to create conditions for both game types where the modified repetition rate was the baseline repetition of the other game type: 50% in the positively autocorrelated RPS condition, and 33.3% for the negatively autocorrelated MP condition. By setting the conditions up like this, it was possible to compare two conditions with an identical likelihood of repetitions but a different game type in order to test the effect of the baseline autocorrelation in the two games. Does the 33.3% likelihood of repetitions in a random RPS condition itself lead to increased *lose-shift* specifically? Similarly, does the 50% likelihood of repetitions in a random MP condition lead to increased *win-stay* specifically? If the finding of different rates of *win-stay* and *lose-shift* between the game types replicates, but the identical autocorrelation conditions do not resemble each other in terms of *stayshift* biases, it would suggest that RPS leads to different decision biases simply due to having three options (as the potential explanation of baseline win-rate was not supported by the results of Experiments 3-5).

By using de facto exploitable opponent conditions where only a specific type of behaviour would lead to increased wins, this experimental set-up also allowed for a return to the analysis of post-error slowing against exploitable and unexploitable opponents (c.f. Experiments 1 and 2). I expected participants to exhibit post-error slowing in the positive but not the baseline or negative autocorrelation conditions, as based on Scheibehenne et al. (2011) and Wilke et al. (2014), participants should be most likely to be able to exploit the positively autocorrelated opponents.

4.3.1.1 Hypotheses. My hypotheses for Experiment 6 were:

1) An overall win-stay bias in MP but not in RPS in the baseline condition (per Experiment 5)

2) An overall lose-shift bias in RPS but not in MP in the baseline condition (per Experiment 5)

3) Higher likelihood for individuals to adopt a win-stay bias in MP than in RPS in the baseline condition (per Experiment 5)

4) Higher rates of optimal behaviour in the positively autocorrelated than the negatively autocorrelated conditions (per Scheibehenne et al., 2011; Wilke et al., 2014)

5) Specifically higher rates of win-stay in MP and lose-shift in RPS in the positively autocorrelated conditions

6) Post-error slowing in the positively and negatively autocorrelated conditions, modulated by win-rate (per Experiments 1-2)

7) Post-error speeding in the baseline conditions (per Experiments 1-2 and 5)

4.3.2 Method.

4.3.2.1 Participants. Thirty-seven participants (N = 37, 32 female, M_{age} = 23.81, SD_{age} = 4.35) from the University of Sussex participant pool and visiting students were recruited via posters, flyers and advertising in summer school workshops. Two participants were excluded after testing due to having later indicated that they had participated in earlier similar studies from the lab, leading to a final sample of 35. I was initially supposed to continue data collection until a sample of 48 viable participants, four for each of the counterbalancing conditions of the experiment (see below), were recruited, but had to stop due to time pressure. Each participant was paid a flat £6 reward for participation. Informed consent was obtained from all participants before testing, and the experiment was approved by the School of Psychology at the University of Sussex (ER/JS/753/12).

4.3.2.2 Materials. The stimulus materials and their presentation were identical to those of Experiment 5.

4.3.2.3 Design. The experiment was 2x3 within-subjects design, with game type *(RPS, MP)* and level of autocorrelation *(baseline, positive, negative)* as factors. The conditions were counterbalanced so that an individual participant would first play through all three RPS blocks or all three MP blocks, followed by the three remaining blocks. The order of the autocorrelation conditions was always identical between the game types, and varied between the counterbalancing orders. There were 12 different counterbalancing orders (6 different permutations of the three autocorrelation conditions, and 2 different orders of the game type conditions).

I defined the positivity or negativity of autocorrelation in relation to the baseline of each game type (33.3% for RPS and 50.0% for MP). That is, a positive autocorrelation condition always represented a likelihood of the opponent repeating their move that was higher than baseline. For both game types, I made the deviations from baseline in steps of 16.7% points (see Table 4.4). Thus, the RPS baseline condition had a 33.3% likelihood of the opponent repeating their previous move; the RPS positive autocorrelation condition had a 50.0% likelihood; and the RPS negative autocorrelation condition had a 16.67% likelihood. Likewise, the MP baseline condition had a 50% likelihood of the opponent repeating their previous move; the MP positive autocorrelation condition had a 66.7% likelihood; and the MP negative autocorrelation condition had a 33.3% likelihood.

Regardless of the condition, the first move the computer opponent made was randomized. After this, on each round, the opponent either repeated the previous move or shifted according to the likelihood of repetition in that condition. The likelihoods for different rounds were independent. For RPS, in the case of the opponent shifting, the likelihood of shifting in either direction *(upgrade* or *downgrade)* was 50%, and was independent for each case of a shift.

Table 4.4. Rate of repetition by the opponent conditions in Experiment 6

Game	*Autocorrelation*	*Repetition likelihood*
RPS		
	Positive	50.0%
	Baseline	33.3%
	Negative	16.7%
MP		
	Positive	66.7%
	Baseline	50.0%
	Negative	33.3%

4.3.2.4 Procedure. The experimental procedure and instructions to given to participants were identical to Experiment 5. After finishing all of the game blocks, I debriefed participants, thanked for their time and paid them the flat participation fee of £6.

4.3.3 Results.

4.3.3.1 Win-rates. I analysed the success of the participants by entering the wins and non-wins from each trial into a GLMM with game type *(RPS, MP)* and autocorrelation *(baseline, positive, negative)* entered as within-subjects factors and participants entered as random effects. The win variable was a binary variable with a value of 1 for each win and a value of 0 for each loss or draw for each trial in the experiment. The model fixed the MP and baseline conditions as references. As in Experiment 5 (see section 4.2.3.1), I analysed both game types within the same model in order to be able to test for interaction effects, which would be cumbersome to test with separate models on the two game types.

Table 4.5. Win-rates and likelihood of individual above-chance win-rate (back-transformed estimated marginal means) in Experiment 6

Game	*Autocorrelation*	*Win-rate overall*	*Individual success*
RPS	Positive	36.2% (0.9%) *	7.2% (4.6%)
	Baseline	33.0% (0.89%)	1.5% (1.7%)
	Negative	34.0% (0.9%)	5.1% (3.7%)
MP	Positive	55.9% (0.9%) *	17.2% (7.8%)
	Baseline	48.8% (1.0%)	1.5% (1.7%)
	Negative	50.6% (1.0%)	5.1% (3.7%)

Note: standard deviation in parentheses. Asterisks indicate that the chance-level win-rate falls below the 95% CI for the observed rate. Likelihoods of individual success based on results obtained using binary variables based on the z-test.

There was a significant main effect of game type collapsed across autocorrelation ($\beta = -0.66$, $SE(\beta) = 0.05$, $z = -12.70$, $p < .001$), with MP (M = 51.8%, SE = 0.6%) having a higher win-rate than RPS (M = 34.4%, SE = 0.6%), as expected given different baseline win-rates (33.3% for RPS and 50.0% for MP). The *positive* autocorrelation condition also differed significantly from the *baseline* autocorrelation condition ($\beta = 0.28$, $SE(\beta) = 0.05$, $z = 5.57$, $p < .001$), with higher average win-rates in the *positive* (M = 46.0%, SE = 0.7%)

compared to the *baseline* autocorrelation condition (M = 40.9%, SE = 0.7%). The *negative* autocorrelation condition did not differ significantly from the *baseline* autocorrelation condition (β = 0.07, SE(β) = 0.05, z = 1.39, p < .001). There was a marginal interaction effect between game type and the *positive* autocorrelation condition (β = -0.14, SE(β) = 0.07, z = -1.93, p = .054) but no significant or marginal interaction between game type and the *negative* autocorrelation condition (β = -0.02, SE(β) = 0.07, z = -0.33, p = .741). The marginal interaction seems to have stemmed from a larger difference in win-rate between the *baseline* and *positive* autocorrelation condition in MP than in RPS (pairwise comparison significant only for MP, Tukey's HSD, p < .05). For both MP and RPS, the chance level win-rate of the game fell below the 95% CI of the observed win-rate only in the *positive* autocorrelation condition (see Table 4.5). Both the *baseline* and *negative* autocorrelation condition had chance-level win-rates, supporting the hypothesis that participants would perform better against opponents with a bias towards repeating their choices rather than alternating. The marginal interaction effect suggests participants may have been more successful against positively autocorrelated opponents in MP compared to RPS, which would lend support to the hypothesis that participants would find adopting a win-stay bias easier in MP than in RPS.

As a secondary test of success, I calculated a binary variable denoting above-chance win-rates for each participant. I calculated this variable using a two-tailed one sample z-test of proportion, testing each participant’s win-rate against the chance level for each game (33.3% for RPS and 50% for MP). I assigned the variable a value of 1 if a participant had a win-rate significantly above chance and a value of 0 if the participant’s win-rate did not significantly differ from chance or was significantly below chance. I then entered this variable as the dependent variable into the same GLMM as above (see Table

4.5). Here, there was no significant main effect of game type (β = 0.00, SE(β) = 1.47, z = 0.00, *p* = 1). The *positive* autocorrelation condition differed significantly from the *baseline* autocorrelation condition (β = 0.26, SE(β) = 1.16, z = 2.26, *p* = .024), with a higher likelihood of individual success in the *positive* (M = 12.2%, SE = 5.1%) than in the *baseline* autocorrelation condition (M = 1.5%, SE = 1.3%). The negative autocorrelation condition did not differ significantly from the baseline autocorrelation condition (β = 1.26, SE(β) = 1.23, z = 1.03, *p* = .305). There was no significant interaction between game type and the *negative* autocorrelation condition (β = 0.00, SE(β) = 1.73, z = 0.00, *p* = 1) or between game type and the *positive* autocorrelation conditon (β = -0.99, SE(β) = 1.64, z = -0.60, *p* = .549). The results of this analysis did not change when I replaced the dependent variable with a binary bias variable that was based on the more conservative exact binomial test.

The results of the individual success analysis further support the hypothesis of participants performing better against opponents with a bias towards repeating rather than alternating. However, note that the likelihood of individual success was very low for all conditions (see Table 4.5), suggesting that a very small minority of participants were actually successfully exploiting even in the *positive* autocorrelation condition. The highest individual win-rates achieved in RPS were 54.4% *(positive)* and 44.4% *(negative)*, and the highest win-rates in MP were 64.4% *(positive)* and 63.3% *(negative)*. Note also the numerical trend with a higher rate of successful individuals in the positive autocorrelation condition for MP compared to RPS, with rates of successful individuals being equal between games for other autocorrelation conditions. This trend is non-significant, but is again in the hypothesized direction of more success in adopting a *win-stay* bias in MP than in RPS (this trend was similar regardless of how the binary success variable was calculated).

4.3.3.2 Reinforcement biases. I analysed the rate of win-stay and lose-shift choices in a similar way to the analyses in Experiment 5, with separate with separate GLMMs first for the overall rate of a choice type, then with the same analysis on a binary variable denoting whether a participant had a significant bias. As in Experiment 5 (see section 4.2.3.1), I used a single model in order to be able to examine interaction effects, which would be cumbersome to analyse if I had analysed the two game types in two separate models. See Table 4.6 for descriptives.

Table 4.6. Win-stay and lose-shift decisions and likelihoods of individual participants having a win-stay or lose-shift bias (back-transformed estimated marginal means) in Experiment 6

Game	*Autocorr.*	*WS overall*	*WS individual*	*LS overall*	*LS individual*
RPS	Positive	60.8% (4.5%) * [33.3%]	63.0% (13.9%)	75.5% (3.2%) * [66.7%]	21.7% (8.8%)
	Baseline	45.0% (4.7%) * [33.3%]	33.3% (13.4%)	70.3% (3.6%) [66.7%]	21.7% (8.8%)
	Negative	38.0% (4.5%) [33.3%]	24.4% (11.4%)	65.0% (3.9%) [66.7%]	13.0% (6.5%)
MP	Positive	83.8% (2.6%) * [50.0%]	93.8% (4.3%)	56.8% (4.1%) [50.0%]	39.7% (11.4%)
	Baseline	69.1% (4.0%) * [50.0%]	53.2% (14.8%)	52.2% (4.1%) [50.0%]	25.0% (9.5%)
	Negative	58.6% (4.5%) [50.0%]	28.7% (12.5%)	42.1% (4.1%) [50.0%]	4.3% (3.1%)

Note: standard error in parentheses. Expected probabilities, under the null assumption of no bias, in square brackets. Asterisks indicate that the expected rate of a strategy type falls below the 95% CI for the observed rate. Likelihoods of individual biases based on results obtained using binary variables based on the z-test.

4.3.3.2.1 Win-stay. I entered the rates of win-stay behaviour from each participant into a GLMM with a logit link, with game type *(RPS, MP)* and autocorrelation *(baseline, positive, negative)* entered as within-subjects factors and partdicipants entered as random effects. The model fixed the MP and *baseline* conditions as references. The model indicated a significant main effect of game type collapsed across autocorrelation (β = -1.00, SE(β) = 0.09, z = -11.16, $p < .001$), with MP having a higher rate of *win-stay* (M = 70.5%, SE = 3.6%) than RPS (47.9%, SE = 4.4%). The model also indicated that the *baseline* autocorrelation condition differed significantly from both the *positive* (β = 0.84, SE(β) = 0.09, z = 9.82, $p < .001$) and *negative* (β = -0.46, SE(β) = 0.08, z = -5.77, $p < .001$) autocorrelation conditions. The *baseline* autocorrelation condition (M = 57.1%, SE = 4.3%) had a lower rate of *win-stay* behaviour than the *positive* autocorrelation condition (M = 72.3%, SE = 3.5%), and a higher rate than the *negative* autocorrelation condition (M = 48.3%, SE = 4.4%). There were no significant or marginal interaction effects (every $p > .1$).

As the expected rate of *win-stay* assuming players played randomly is different between the game types, the main effect of game type was to be expected. However, contrary to Experiment 5, the chance level of *win-stay* behaviour fell below the lower 95% CI of the observed rate for both MP and RPS. The *negative* autocorrelation condition had no overall *win-stay* bias for either game type (RPS: M = 38.0%, SE = 4.5%; MP: M = 58.6%, SE = 4.5%), whereas the *positive* autocorrelation condition had an overall bias for both game types (RPS: M = 60.8%, SE = 4.5%; MP: M = 83.8%, SE = 2.6%). Since the rate of win-stay responding in the *negative* autocorrelation condition was neither significantly higher or lower than chance , the results support the notion that optimal responding is harder for participants in the face of an alternation bias. That is, the overall pattern suggests a tendency toward the optimal response in the *positive* autocorrelation condition but not in the *negative* autocorrelation condition, where a below-chance level of win-stay responses would signal optimal play. It thus seems that participants on average had a bias in both game types in the baseline and increased or decreased their *win-stay* behaviour accordingly with *positive* or *negative* autocorrelation but did not reach a group-level significant trend of optimal play in the *negative* autocorrelation condition. See Table 4.6 for back-transformed probabilities.

For the likelihoods of individual participants having a *win-stay* bias, the model indicated no significant main effect of game type (β = -0.82, SE(β) = 0.65, z = -1.26, p = .206). There was also no significant difference between the *baseline* and *negative* autocorrelation conditions (β = -1.04, SE(β) = 0.66, z = -1.57, p = .116), but there was a significant difference between the *baseline* and *positive* autocorrelation conditions (β = 2.58, SE(β) = 0.78, z = 3.31, p < .001). There was no significant interaction between game type and the *negative* autocorrelation condition (β = 0.60, SE(β) = 0.93, z = 0.65, p = .518) or between game type and the *positive* autocorrelation condition (β = -1.35, SE(β) = 0.99, z = -1.37, p = .169). The *positive* autocorrelation condition had a significantly higher likelihood of an individual having a *win-stay* bias (M = 78.4%, SE = 8.0%) than the *baseline* autocorrelation condition (M = 43.2%, SE = 11.8%). Again, contrary to Experiment 5, there were no significant differences in the rates of *win-stay* bias based on game type (see Table 4.6 for back-transformed probabilities). There was, however, a numerical trend in the direction of a higher likelihood of individual win-stay biases in MP than in RPS for all conditions, consistent with Experiment 5, especially notable in the *positive* autocorrelation condition. The results of this analysis did not change meaningfully when I replaced the dependent variable with a binary bias variable calculated based on the more conservative exact binomial test. All the numerical trends remained the same with this alternative bias variable as well.

Overall, the results do not support MP specifically being characterized by a *win-stay* bias (contra Experiment 5) and instead suggest a significant overall bias in both game types. However, the numerical trend in the likelihoods of individual participants having a win-stay bias does match the notion of a win-stay bias being less frequent in RPS than in MP; this effect not reaching significance may be an effect of the low sample size. Participants increased and decreased their *win-stay* responding when it was optimal to do so, but due to an initial bias in the *baseline* condition, the rate of responding the *negative* autocorrelation condition was not significantly different from chance.

4.3.3.2.2 Lose-shift. I analysed overall rates of *lose-shift* responding and the binary bias variable for *lose-shift* identically to *win-stay* (see Table 4.6). For overall rates of *lose-shift,* the model indicated a main effect of game type ($\beta = 0.78$, $SE(\beta) = 0.09$, $z = 8.59$, $p < .001$). Here, the rate of *lose-shift* was higher in RPS (M = 70.3%, SE = 3.4%) compared to MP (M = 50.4%, SE = 3.9%), as expected based on the different rates of *lose-shift* behaviour even assuming players played randomly (50.0% for MP and 66.7% for RPS). There was also a significant difference between the *baseline* and *positive* autocorrelation conditions ($\beta = 0.19$, $SE(\beta) = 0.08$, $z = 2.38$, $p = .017$) and a significant difference between the *baseline* and *negative* autocorrelation conditions ($\beta = 0.16$, $SE(\beta) = 0.13$, $z = -5.36$, $p < .001$). There was no significant interaction between game type and the *negative* autocorrelation condition ($\beta = 0.16$, $SE(\beta) = 0.13$, $z = 1.26$, $p = .20$) or between game type and the *positive* autocorrelation condition ($\beta = 0.08$, $SE(\beta) = 0.13$, $z = 0.581$, $p = .561$). Overall, participants decreased their rate of *lose-shift* choices in the *negative* autocorrelation conditions (M = 53.5%, SE = 3.8%) compared to the *baseline* autocorrelation conditions (M = 61.2%, SE = 3.7%), and increased the rate in the *positive* autocorrelation conditions (M = 66.2%, SE = 3.5%). Once again, contrary to Experiment 5, there was no overall significant *lose-shift* bias in either of the *baseline* conditions, indicating that on average neither game type led to an overall *lose-shift* bias. There was no bias in either of the *negative* autocorrelation conditions either, and a bias for RPS but not for MP in the positive *autocorrelation* condition. This fits with the notion of RPS being more conductive to the lose-shift rule, and participants finding it easier to increase its use in RPS compared to MP.

For likelihoods of an individual participant having a *lose-shift* bias, there was no significant main effect of game type (β = -0.184, SE(β) = 0.607, z = -0.303, p = .762). There was also no significant difference between the baseline and positive autocorrelation conditions (β = 0.681, SE(β) = 0.590, z = 1,154, p = .249), but there was a significant difference between the baseline and negative autocorrelation conditions (β = -1.994, SE(β) = 0.792, z = -2.518, p = .012). There was also no significant interaction between game type and the *negative* autocorrelation condition (β = 1.38, SE(β) = 1.02, z = 1.36, p = .175) or between game type and the *positive* autocorrelation condition (β = -0.68, SE(β) = 0.85, z = -0.80, p = .424). In the *negative* autocorrelation conditions (M = 8.7%, SE = 4.0%), participants had a significantly lower likelihood of bias than in the *baseline* autocorrelation conditions (M = 23.4%, SE = 7.4%). There was a numerical trend toward a higher likelihood of individuals expressing a *lose-shift* bias in MP than in RPS in the *positive* and *baseline* autocorrelation conditions, but not in the *negative* autocorrelation condition. The results of this analysis did not change meaningfully when I replaced the dependent variable with a binary bias variable calculated based on the more conservative exact binomial test. All the numerical trends remained the same with this alternative bias variable as well.

Overall, the results do not seem to support RPS being characterized by an overall *lose-shift* bias contrary to MP, but participants were still better at increasing their rate of lose-shift responding in RPS than in MP when it was optimal. Participants increased and decreased their lose-shift responding when it was optimal to do so, but only reached a rate significantly different from baseline in the positive autocorrelation condition in RPS. The results suggest no overall bias in either game type in the baseline, and numerically lower likelihoods of people expressing a *lose-shift* bias than a *win-stay* bias.

4.3.3.3 Reaction times. I analysed median reaction times for decisions at trial *n+1* as in Experiment 5, separately for the two game types using a two-way repeated measures ANOVA, with the autocorrelation condition *(baseline, positive, negative)* and outcome at trial *n (MP: win, lose; RPS: win, lose, draw)* as within-subjects factors for both separate analyses. Six participants were excluded due to having at least one block average median reaction time that was at least twice the block average median (as in previous experiments), yielding a final sample of 29 participants for the reaction time analyses.

For RPS, Mauchly's test indicated violations of sphericity, and I used the Greenhouse-Geisser correction for effects with violations. There was no significant main effect of the autocorrelation condition [$F(1.46, 40.92) = 3.39$, $MSE = 268019$, $p = .058$, $\eta_p^2 = .11$], a marginal main effect of outcome [$F(2, 56) = 2.74$, $MSE = 68724$, $p = .073$, $\eta_p^2 = .09$], and no significant interaction [$F(2.57, 71.83) = 0.26$, $MSE = 51631$, $p = .825$, $\eta_p^2 < .01$]. Contrary to my hypothesis, participants did not slow down after losses even in the exploitable conditions; the marginal trend of the effect of outcome was towards post-error speeding, not post-error slowing (see Table 4.7).

For MP, Mauchly's test indicated violations of sphericity, and I used the Greenhouse-Geisser correction for effects with violations. There was no significant main effect of the autocorrelation condition [F(1.67, 46.65) = 0.32, MSE = 13383, *p* = .692, η_p^2 = .011], a significant main effect of outcome [F(1, 28) = 5.87, MSE = 6772, *p* = .022, η_p^2 = .17], and a marginal interaction [F(2, 56) = 3.27, MSE = 2349, *p* = .054, η_p^2 = .11]. Participants had faster lose (M = 353, SE = 132) than win (M = 383, SE = 157) reaction times, contrary to the hypothesis that opponent exploitability should lead to post-error slowing. The marginal interaction seems to have been driven by the difference between win and lose reaction times having been greater in the *positive* autocorrelation condition compared to the *baseline* or *negative* autocorrelation conditions. This trend even further suggests that exploitability in and of itself did not lead to post-error slowing, as the participants seem to have been most able to exploit the opponent in the *positive* autocorrelation condition, where lose reaction times were the *lowest.* See Table 4.7 for descriptive statistics.

To repeat the analysis of correlation between post-error slowing and individual win-rates from Experiments 1-2, I calculated a Pearson correlations for the difference between an individual's median lose and win reaction times and their win-rate separately for each of the *positive* and *negative* autocorrelation conditions. None of these correlations were statistically significant (every *p* > .05). This result likely stems from low overall and individual win-rates (see above) in the present experiment (c.f. Experiments 1-2).

Table 4.7. Average median reaction times (milliseconds) for choices on trial *n+1* after different outcome types on trial *n* in Experiment 6, N = 29

Game	*Autocorrelation*	*Win*	*Lose*	*Draw*
RPS	Positive	543 (61)	482 (60)	451 (54)
	Baseline	509 (56)	403 (50)	434 (54)
	Negative	666 (98)	601 (109)	583 (97)
MP	Positive	386 (32)	332 (25)	N/A
	Baseline	389 (34)	361 (28)	N/A
	Negative	373 (21)	365 (21)	N/A

Note: standard error in parentheses.

4.4 General Discussion

In Experiments 5 and 6, I attempted to replicate the findings regarding reinforcement biases of the prior 4 experiments and examine potential reasons for them. Experiments 1-2 found no evidence of either a *win-stay* or a *lose-shift* bias in RPS. In contrast, Experiment 3-4 provided evidence for both biases in MP, with *win-stay* being more common. In Experiment 5, I compared behaviour in both game types in the same sample. Here, I replicated the results of earlier studies on RPS (Dyson et al., 2016; Forder & Dyson, 2016), suggesting the presence of only a *lose-shift* and no *win-stay* bias in RPS. I also found a *win-stay* but no *lose-shift* bias in MP. Note that while Experiments 3 and 4 found both biases in the average data, lose-shift was numerically weaker in MP in both Experiment 3 and Experiment 4, and did not always reach significance. Moreover, the likelihood of an individual participant having a *lose-shift* bias in MP was quite low across Experiments 3 and 4, and thus the result of no significant lose-shift bias in Experiment 5 is not entirely surprising. The results of Experiment 5 thus broadly fit with Ivan et al.'s (2018) and Gruber and Thapa's (2016) findings suggesting that both for humans and rats, the *lose-shift* response is distinct from the *win-stay* response, and that *lose-shift* is increased by cognitive demands and reduced by task delays. Assuming that MP has lower cognitive demands compared to RPS – fewer choice options for the player and fewer opponent choice alternatives to anticipate – this likely reduces *lose-shift* in MP and increases it in RPS. However, this does not explain why *win-stay* responding in RPS is rarer than in MP.

Given that the likelihood of individuals having a *lose-shift* bias was not significantly affected by the game type in Experiment 5, and that the likelihood of a *win-stay* bias was higher in MP, the results of Experiment 5 suggest that people are simply less likely to have biases in RPS in general. If and when they did, however, the *lose-shift* bias seems to have been stronger, as reflected by the group-level rate of *lose-shift* behaviour being significantly above 66.67%. The fact that participants were overall less biased in RPS would fit with Rapoport and Budescu's (1992) notion of working memory demands reducing deviations from randomness. It would thus seem, based on Experiment 5, that MP has higher rates of *win-stay* bias due less cognitive demands leading to more deviations from randomness, while simultaneously reducing the rate of *lose-shift* (Ivan et al., 2018). That is, higher memory demands of a game may induce *lose-shift* specifically while reducing other types of deviations from random play.

Additionally, Experiment 5 found only post-error speeding, consistent with the notion that post-error slowing requires an increased win-rate caused by actually exploiting an opponent, the latter of which was absent in all conditions in Experiment 5. This further supports the notion of *lose-shift* being more automatic, as participants took less time to think after losses and when there was a trend in decisions following losses, it was towards shifting.

In sum, the results of Experiment 5 replicated the previously observed differences in reinforcement biases between the game types, while not supporting the notion that baseline win-rates would affect this.

However, all of this is questioned by the results of Experiment 6. Here, the average data showed an overall *win-stay* bias but no overall *lose-shift* bias for both game types in the baseline condition, where the opponent could not be exploited in any way. This was contrary to my hypothesis of, based on prior experiments in this thesis and literature (Dyson et al., 2016; Forder & Dyson, 2016), that RPS specifically should have no overall *win-stay* bias. The results more fit with Alós-Ferrer and Ritschel's (2018) findings of an overall *win-stay* bias in a three-option game that differed from RPS in not having a neutral outcome option. Moreover, when the opponent could be exploited by a participant making more *stayshift* choices, participants in both game types increased their *win-stay* responding, contrary to my hypothesis that participants would be less likely to do this in RPS given the initial lack of a *win-stay* bias observed previously. It must be noted at this point that Experiment 6 suffered from a lack of participants and data collection for Experiment 6 had to be cut before the planned sample of 48 participants had been collected, which could have prevented some effects from reaching significance.

The results of Experiment 6 did support the notion that participants had more difficulties in increasing *lose-shift* responding in MP than they did in RPS: the only case of participants overall making significantly more than chance *lose-shift* choices was in the positively autocorrelated RPS condition. While participants in both game types increased and decreased their rate of win-stay and lose-shift responding appropriately in the autocorrelation conditions, it seems that in only RPS did they actually reach an average level above statistical noise. However, note that this effect was again driven by a minority of participants. The *lose-shift* rule thus seemed to be generally harder for participants to apply even when it was optimal (replicating Scheibehenne et al.'s, 2011 observations). The participants also increased and decreased their *stayshift* responding appropriately despite very few participants ever achieving above-chance wins: thus, the participants could clearly perceive the opponents' exploitability, but the number of rounds or the increases in autocorrelation may have been too low to allow for higher win-rates. Moreover, the numerical trend was still in favour of the likelihood of individual *win-stay* biases being higher in MP compared to RPS even in the baseline condition (53.15% vs. 33.30%, respectively; see Table 4.6). However, the trend in *lose-shift* was also such that it was higher in MP than in RPS, though with a smaller difference (25.01% vs 21.72%, respectively). In sum, despite its limitations, the results of Experiment 6 still tentatively support a differentiation in *win-stay* and *lose-shift* biases, with the former being numerically more common among individuals in MP, and the latter being generally rarer than the former. This further supports the notion that a "true" *stayshift* bias is relatively rare.

Finally, there seemed to be no post-error slowing as a function of opponent exploitability. In MP, there was a significant effect in the opposite direction of post-error speeding; for RPS, there was a marginal effect also in the direction of post-error speeding. This was contrary to my hypothesis of post-error slowing in the exploitable conditions (based on Experiments 1-2). Moreover, there were no significant correlations between individual win-rates and the difference between win and lose reaction times in the exploitable conditions, unlike in the exploitable conditions of Experiments 1-2. The result likely stems from low win-rates across the exploitable conditions: as mentioned above, while participants learned to exploit their opponents, few participants actually achieved above-chance wins.

For a descriptive comparison, the highest win-rate any individual participant achieved in RPS in Experiment 6 was 54.44% (in the positive autocorrelation condition), whereas the highest RPS win-rates in Experiments 1-2 were 78.89% and 83.33%, respectively. Thus, the likeliest explanation for the difference between the results of Experiments 1-2 and 6 is that win-rates simply need to be higher for participants to start slowing down after losses. Due to the methodological issues in Experiments 3-4 potentially masking RT effects, the present studies can not rule out the notion of post-error slowing being caused simply by infrequent errors (Danielmeier & Ullsperger, 2011).

4.5 Conclusion

In sum, the results of Experiments 5 and 6 replicate some general patterns of reinforcement biases observed in the earlier experiments, but also provide conflicting results regarding the differences between the two game types. In general, responding after losses seems to be more random than responding after wins, despite shorter reaction times following losses. The reaction time results, taken together with the results of Experiments 1 and 2, support the notion of high win-rates against exploitable opponents, rather than perceived exploitability in itself, being a key predictor of post-error slowing. The implications of these results will be discussed in more detail in Chapter 5.

CHAPTER 5: General Discussion

5.1. Summary of Experiments

Across six experiments, I examined the expression of reinforcement biases (*win-stay* and *lose-shift*, hence *stayshift*) and reaction time trends in two simple zero-sum competitive games, Rock, Paper, Scissors (RPS) and Matching Pennies (MP). Across the experiments, participants played both game types against different computer opponents. The experimental conditions varied in terms of win-rates, success slopes, and opponent exploitability. Here, I shall first summarize the results of each empirical chapter (Chapters 2-4), and then discuss the results and their implications as a whole.

In Experiments 1 and 2 (Chapter 2), I examined players' decisions and reaction times when playing against both *unexploitable* (randomly playing) and *exploitable* opponents in RPS. The *exploitable* opponents in Experiments 1-2 specifically had a predictable of play that consisted of cyclic shifting. The shifts were set up in such a way that the optimal strategy was a specific type of shift after wins in both experiments, thus misaligning with the *win-stay* bias. Additionally, the optimal choice after draws in Experiment 1 and after losses in Experiment 2 was to stay, further misaligning the optimal responses from the *stayshift* bias. The aim was to examine how well participants would play as a function of the prior outcome and whether the optimal move aligned with reinforcement or not. I failed to replicate the previous finding (Dyson et al., 2016; Forder & Dyson, 2016) of a *lose-shift* (and *draw-shift*) bias in RPS in the *unexploitable* conditions. The participants in Experiments 1 and 2 also exhibited no overall *win-stay* bias in the *unexploitable* condition: the participants had no outcome-dependent predictable pattern of play. Against exploitable opponents, the participants in both experiments made the most optimal choices after wins, despite the optimal strategy in both being a form of *win-shift*, and despite the optimal strategies after losses and draws at times aligning with reinforcement rules. In the *exploitable* conditions, participants also exhibited post-error slowing as a function of how well they did against the opponents. However, this increased pausing before decisions following losses relative to wins did not seem to translate to better performance after losses, as both the overall sample and the subgroup of participants who reached an above-chance win-rate had the lowest rates of optimal decisions specifically on trials that followed a loss.

In Experiments 3-4 (Chapter 3), I examined players' decisions, reaction times and confidence of future wins in a series of manipulated blocks of MP. None of the opponents in Experiments 3-4 had any set exploitable pattern of play: instead, they were set to allow the player to win a certain number of rounds in certain set orders (success slopes) no matter what the player did. The aim of the experiments was to examine whether changing win-rates, without any pre-set pattern in the opponents' moves, would lead to changes in *stayshift* responding or reaction times, and how participants' confidence in wins developed. Overall, the results supported no clear effects of above or below chance win-rates (*continuous success, continuous failure*) or success slopes (*ascending, descending*) in MP affecting the biases. Specifically, the likelihood of an individual expressing either type of bias was not affected by the experimental conditions in either experiment, and the likelihood of an individual having a *win-stay* bias was greater than them having a *lose-shift* bias. In terms of overall responding, there was a significant overall *win-stay* bias for each condition, with a weaker but statistically significant lose-shift bias in all conditions except the random control condition in Experiment 4. The results regarding confidence following different success slopes in the early and late parts of a block, when controlled for in terms of the local win-rate at which player confidence was measured, suggested that the most likely predictors of confidence at a given time was the local win-rate, and/or the local trend in win-rate prior to the measurement point. Participants were also generally overconfident across the experiments. Given very little variation in *stayshift* responding and much more variation in on-line confidence as a function of local win-rates and off-line confidence as a function of overall win-rates, this further suggests that confidence likely plays little role in *stayshift* biases. Finally, the experiments found no post-error slowing or speeding in any condition, but this result was confounded by a delay in game responses caused by the intervening confidence measures on each trial.

Due to the differences in *stayshift* responding I observed between the two game types and conflicting literature on whether this should be expected, in Experiments 5-6 (Chapter 4) I compared behaviour between the games within the same sample of participants and within the same experimental design. By comparing the two games within the same paradigm, I could also control for any potential effects of methodological differences between Experiments 1-2 and Experiments 3-4. Most critically, Experiments 5-6 removed the confidence measures used in Experiments 3-4 to remove their potential confounding (interruption) effect on reaction times. The aim of Experiment 5 was to replicate the results regarding reinforcement biases from Experiments 1-4 under a design of fixed win-rates. Experiment 5 broadly replicated the previous findings of the *win-stay* bias specifically being more common in MP than the lose-shift bias on an individual level, and overall *win-stay* responding being above chance only in MP. Additionally, only a minority of participants had a *lose-shift* bias in either game type, with the overall rate of *lose-shift* responding being above chance only in RPS. Experiment 5 also found no support for the hypothesis that the differences in bias between the games were caused by the different baseline win-rates of the games (33.33% for RPS and 50.00% for MP).

Similarly, the aim of Experiment 6 was to again replicate the difference in biases between the two games, this time with both *unexploitable* and *exploitable* computer opponents. This design allowed for testing two separate questions. First, given differences in biases between the two game types, whether participants would also learn to increase or decrease *stayshift* responding differently when that was the optimal way of playing due to positively or negatively autocorrelated choice patterns from the opponents. Second, whether the difference observed between the game types could be explained in terms of the different baseline autocorrelation players would normally face in a typical experimental scenario with a randomly playing opponent. However, Experiment 6 did not replicate the previously observed difference between game types: here, the overall rate of *win-stay* responding was above chance for both game types. The results did lend some support to the hypothesized different rates of learning, as only in RPS was *lose-shift* responding above chance on the group level when lose-shift was the strategically optimal choice. In both game types, participants increased and decreased their rate of *stayshift* responding when it was optimal to do so (increased during positive autocorrelation, decreased during negative autocorrelation). However, the overall rate of either *win-stay* or *lose-shift* responding differed significantly from chance (in the appropriate direction) only in the positive autocorrelation conditions, supporting the hypothesis that learning is easier when the optimal strategy is aligned with reinforcement. Finally, both experiments found a trend of post-error speeding in all experimental conditions, suggesting that the null result of Experiments 3-4 likely stemmed from the methodological issues outlined above. Further, given the results of Experiments 1-2 showing post-error slowing as a function of win-rates in a context where participants achieved wins above 70% in RPS, the results of Experiment 6, where win-rates where much lower (54.44% in RPS and 64.44% in MP at highest), suggest that relatively high win-rates are needed for post-error slowing. The finding that the participants did increase or decrease *lose-shift* responding when it was optimal seems to suggest that the speeding after losses did not necessarily lead to suboptimal decisions.

5.2 Flexibility and Variability of Reinforcement Biases

In sum, the results of the six experiments provide conflicting information about how strong the *stayshift* bias is in the baseline experimental situation where an opponent cannot be meaningfully exploited. Table 5.1 provides a summary of the presence of different biases in the six experiments in the overall group data. The table includes only the most clearly "random" conditions of each experiment. Experiment 3 is excluded from this table as it had only different success slope conditions with no randomized condition (see section 5.4 for further discussion on the differences in randomization between the experiments). Note that there was an overall *lose-shift* bias in MP in Experiments 3-4 in all of the conditions not included in the table. However, this bias was of a smaller magnitude than the *win-stay* bias, that is, *lose-shift* responding in Experiments 3-4 was consistently closer to chance on the group level than *win-stay* responding. Taken together with the low number of individuals in both game types expressing a *lose-shift* bias across Experiments 3-6, the explanation for the inconsistency in the results regarding a group-level *lose-shift* bias is likely that the bias is simply rare on an individual level. The results thus suggest that participants across the experiments were responding more randomly after losses than they were after wins. This seems to have been the case both for the random and quasi-random conditions across experiments: the fixed win-rate and success slope conditions of Experiments 3-5, as well as the exploitable conditions in Experiments 1, 2 and 6. In Experiments 1-2, participants had the least optimal responses for decisions made immediately following a loss. In Experiment 6, the rate of overall *lose-shift* responding was significantly different from chance only in the positively autocorrelated RPS condition, whereas the rate of overall *win-stay* responding reached a significant bias in both games. Moreover, the likelihood of individual participants expressing a *lose-shift* bias when it was optimal was numerically lower than the likelihood of individuals expressing a *win-stay* bias in Experiment 6.

Table 5.1. The presence of overall (group-level) win-stay and lose-shift biases in the six experiments in the random/unexploitable conditions

Game	*Experiment*	*Randomization*	*WS*	*LS*
RPS	1	Randomized flat distribution of opponent choices	No bias	No bias
	2	Randomized flat distribution of opponent choice	No bias	No bias
	5	Randomized flat distribution of outcomes	No bias	Bias
	6	Fully random opponent choice	Bias	Bias
MP	4	Randomized flat distribution of outcomes	Bias	No bias
	5	Randomized flat distribution of outcomes	Bias	No bias
	6	Fully random opponent choice	Bias	No bias

This overall trend of more random responding after losses marks a clear difference between the two parts of the *stayshift* bias. If *lose-shift* is generally less common, and people tend to respond more randomly after losses than they do after wins, calling the bias "win-stay, lose-shift" is a misnomer. The finding of low rates of *lose-shift* responding further supports Ivan et al.'s (2018) notion of *lose-shift* as an impulsive response that is, under normal circumstances, suppressed by executive functions in adults. As mentioned in the general discussions for Chapters 3 and 4, the number of people expressing both bias types is logically at most the number of individuals expressing the *lose-shift* bias, which was a minority in each of the experiments (3-6) that assessed individual-level biases. This also matches with the observations by Rapoport and Budescu (1992), who found that a minority of their participants playing MP against each other had a *stayshift* bias. Further, this result replicates Scheibehenne et al.'s observation (2011) that *lose-shift* responding, even when participants were incentivized to engage in it by a positively autocorrelated pattern in the game task, was less frequent overall than *win-stay* responding (i.e. responding after losses was more random). If a true *stayshift* bias is very rare, considering reinforcement biases as they are observed in humans as an expression of the Hot Hand Fallacy or a positive recency effect (as seen in Scheibehenne et al., 2011; Wilke & Barrett, 2009; Wilke et al., 2014) loses support. A positive recency effect on the expectations players have of their opponents' choices would mean that players should expect repetitions in general, which would show up as roughly equal rates of *win-stay* and *lose-shift* (as both are optimal assuming the opponent is going to repeat their choice).

Why is win-stay more common than lose-shift, and why is responding after losses more likely to be random? A potential explanation for is that the two reinforcement rules are differentially aligned with decision inertia (see e.g. Alós-Ferrer & Ritschel, 2018), i.e. simply the preference for repeating actions. The response of staying after winning a round is action in accordance to both immediate reinforcement and decision inertia, but shifting after losses aligns only with reinforcement and goes against inertia. Note that inertia on its own would fail to explain why group-level lose-shift behaviour was so rarely different from simply random. That is, if an individual-level lose-stay bias (aligning with inertia) were common, it should also manifest as below-chance lose-shift behaviour, which was not the case for any of my experimental conditions except for the exploitable condition of Experiment 2, where lose-stay was the optimal choice 70% of the time (note that this effect did not replicate in Experiment 6). For experimental conditions where there was no optimal strategy (i.e. the differently formulated random conditions; see Table 5.1), there was never a group-level *lose-stay* bias. The finding that responding after losses tends to be more random than responding after wins makes sense under the assumption that the effects of inertia and reinforcement are additive: if they draw the player in different directions, the end result is less bias in any direction. Thus, reinforcement matters in terms of biases, but reinforcement is not the only factor driving behavioural tendencies.

The interplay of reinforcement and inertia would also explain why participants in Experiment 6 seem to have found it easier to increase their rates of *win-stay* than their *lose-shift* behaviour when that was the optimal thing to do, and were less able to exploit an opponent that could be exploited with the opposite win-shift and lose-stay strategies. That is, when win-stay is optimal, both inertia and reinforcement are already by default nudging players towards the optimal decision; when lose-shift is optimal, there is still an inherent conflict between inertia and reinforcement. When win-shift is the optimal decision rule, both inertia and reinforcement are nudging the player in the wrong direction; but when lose-stay is optimal the situation is similar to when lose-shift is optimal, with inertia and reinforcement pulling in different directions. This may explain both the difference in biases after specific outcomes as well as differences in learning optimal responses to specific outcomes. However, it is unclear why Experiments 1 and 2 seem to suggest that lose-stay may be easier for participants to learn in RPS than a specific kind of lose-shift strategy (one that goes against myopic best reply). This result would seem to imply that inertia is a stronger driver of responses after losses than reinforcement, but no other result in my experiments supports this notion. The interplay between inertia and reinforcement also does not help explain the differences in biased responding observed between two-choice and three-choice games: I will return to this later in this section.

An important and consistent finding in my experiments was that while the *win-stay* bias was more common than the *lose-shift* bias on an individual level, it was still relatively rare for most of the experimental conditions throughout the experiments. The highest likelihood of individuals expressing a win-stay bias, 71.7%, was observed in Experiment 5 in the MP condition where win-rates had been manipulated to be only 33.3% but random. This is a high rate, but it was also in an unusual unexploitable condition due to the manipulated win-rate, and the results of the experiment suggested increased win-stay behaviour as a function of win-rate. For unexploitable conditions where win-rates had not been manipulated to be above or below the normal baseline of a game, the highest likelihood of individuals expressing the bias (observed in the baseline MP condition of Experiment 6) was essentially a coin-flip: 53.2%. This raises questions about how, for all the baseline random conditions where I tested individual-level biases, such a large proportion of the participants did not have a *win-stay* bias. Given hypotheses about the evolutionary roots of the *stayshift* bias (see Scheibehenne et al., 2011; Wilke & Barrett, 2009; Wilke et al., 2014), it is interesting that neither *win-stay* or *lose-shift* biases reached a large majority on the individual level (unless an experimental manipulation caused the game to deviate from baseline in some way). Scheibehenne et al. (2011, pp. 332) specifically reported that “most participants” in their experiment tended to use a *stayshift* decision rule. However, this statement seems to be based on a collapsing over *win-stay* and *lose-shift* responses (as Scheibehenne et al., 2011, also noted that *lose-shift* was less frequent than *win-stay*). More critically, their paper does not report an actual analysis of individual-level biases. Thus, the more correct claim would be that most of the participants' responses, rather than most participants' responses, were biased towards *stayshift*. A crucial property of the experiments reported in this thesis is the analysis of individual-level biases and differentiating between *win-stay* and *lose-shift*. By doing this, the experiments are able to show that there is considerable variation in reinforcement biases both between individuals and between the two different types of reinforcement bias.

The experiments reported here are unable to answer conclusively the question of what caused the variation between individuals. Based on the results of Experiment 2, it does not seem that individual differences in working memory or executive control would predict these differences. Specifically, working memory span did not predict the rate of optimal decisions on trials immediately following any outcome type in the exploitable condition of Experiment 2, and executive control did not predict fewer deviations from randomness unlike I had hypothesized (based on Terhune & Brugger, 2011). Further, the biases do not seem to emerge as a function of different success slopes participants may experience as a function of random chance in the long run, based on Experiments 3-4. Based on Experiment 5, there is some evidence suggesting that random variation of win-rates that participants experience in random games could increase or decrease biased responding, but this effect was only observed for MP. Individual differences in sensitivity to reinforcement are a potential candidate as an explanation for the variance. That is, the FRN, an ERP linked to reinforcement learning (Holroyd & Coles, 2002), has been shown to be increased in people with higher rates of reinforcement-based errors (Achtziger et al., 2015). However, this effect was mediated by financial incentive, only appearing under high compared to low financial incentives (Achtziger et al., 2015). Given the lack of financial incentives in most of the experiments presented in this thesis, and a lack of effects of financial incentives in Experiment 2 where they were offered, it seems that reinforcement biases appear even without financial incentives, but not uniformly across participants. Wilke et al. (2014) found that habitual gamblers were more prone to the *stayshift* bias – however, their study included a financial incentive as well. Further, even if the participants expressing biases in the present experiments happened to be habitual gamblers (gambling background was not checked during recruitment), it would still be unclear whether gambling tendencies are predicted by more sensitivity to reinforcement or vice versa.

When it comes to the variation between the different reinforcement bias types, it seems that biases were simply rarer in RPS than in MP on the individual level, in addition to being more inconsistent in RPS than in MP on the group level (see Table 5.1). This effect does not seem to have been driven either by win-rates or the rates of repetition or alternation inherent in the games in the baseline scenario of a randomly playing opponent. Thus, by elimination, the effect is most likely caused by RPS simply having more choice and/or outcome options than MP, meaning a player needs to track more information. This aligns with Rapoport and Budescu's (1992) and Wagenaar's (1991) argument that higher demands for working memory reduce deviations from randomness by lessening the effects of prior trials on decisions. Additionally, given that a *lose-shift* bias was more consistently observed in RPS than in MP in Experiments 5-6, it seems that *lose-shift* may work in different ways from *win-stay*, increasing as a function of working memory load (see Gruber & Thapa, 2016; Ivan et al., 2018). The lower rates of *win-stay* responding and more consistent *lose-shift* responding in RPS compared to MP may also stem from shifting in RPS being more salient due to more shift options than in MP. In either case, it seems that the presence or absence of reinforcement biases in people's behaviour is dependent on the type of decision task and the amount of options and information the participants have available to them. Given that even the extremely simple addition of one more choice and one more outcome type seems to be enough to reduce the likelihood of a *win-stay* bias, it would follow that the bias is likely quite minimal in more complex competitive scenarios. However, if the lose-shift bias increases as a function of cognitive demands, it may be more common in even more complex scenarios.

While all of the above suggests variability in terms of the *stayshift* bias, there is a consistency in the results in that whenever there was a trend towards a bias based on a game's outcome on the previous trial, it was in the direction of *win-stay* and/or *lose-shift*. That is, the overall rates of responding never showed a significant *win-shift* or *lose-stay* bias in any of the six experiments reported here. This indicates that while reinforcement biases may have been rare on an individual level, biases *against* reinforcement were consistently rarer and/or weaker than reinforcement biases (or, in the cases of no overall bias, equally rare and/or weak). Further, based on the results of Experiment 6, these biases are not trivial (see McKay & Efferson, 2010), given that participants could increase reinforcement-based responding to reach a group-level significant bias, but were less capable of decreasing it when incentivized to do so. Thus, the results support *win-stay* and, to a lesser degree, *lose-shift* being default decision rules, which many people at least in samples consisting mostly of Western university students were able to avoid in zero-sum games with mixed equilibria.

Based on my results, I would urge researchers to be cautious about interpreting group average statistics about biased decisions in absence of individual-level bias analyses. If we wish to understand cognitive biases and their potential implications on learning, on human rationality, or on understanding our evolutionary history, we need to know how common “irrational” behaviour truly is among our samples, let alone the general population. Aggregate statistics of responses to a task may imply a bias, but these statistics may mask the existence of very different patterns of responses (see e.g. Worthy et al., 2012, on an examination of two qualitatively different approaches to a gambling task). This also applies to research more generally in the field of cognitive biases and learning outside of the realm of specifically reinforcement biases: in any task, the average of what people do does not necessarily describe what most people do.

Moreover, given my observations of differences in the incidence of different types of bias as a function of game type, it is important not to draw strong conclusions of how reinforcement biases would affect behaviour in very different task environments. Understanding and predicting such biases requires knowledge of the factors that affect them in different ways and recognizing those factors in different tasks: memory demands or the number of choices, the interaction of inertia and reinforcement (and myopic best reply), and most likely several other things. Still, given that reinforcement biases are clearly not trivial, and that they can appear even without financial incentives, they likely affect several areas of competitive action and learning at least for a sub-population of people. Thus, the interplay of reinforcement and inertia likely affects more complex competitive areas such as competitive sport or gaming. Due to increased cognitive demands, it may be that lose-shift becomes the more common bias: a potential area for more research.

5.3 Reinforcement Biases and Speed of Responding

A key question behind the reaction time analyses in the experiments following Experiments 1-2 was whether post-error slowing was dependent on simply win and loss frequencies or the exploitability of the opponent (see Danielmeier & Ullsperger, 2011; Dutilh et al., 2012). That is, would participants slow down after losses in a context of high win-rates simply as a response to infrequent losses regardless of whether there was a winning strategy, or only after losses against opponents that the participants had learned to exploit? Due to a task interruption effect in Experiments 3-4 that tested high win-rates without opponent exploitability, the results lack a clear comparison between high win-rates under *exploitable* and *unexploitable* conditions. However, the reaction time patterns observed in the rest of the experiments support the role of high win-rates (and thus infrequent losses) as a driver of post-error slowing.

In the experiments where play responses were not interrupted with additional prompts to the participants (Experiments 1-2, 5-6), I observed post-error slowing only in situations where opponents could be exploited *and* the participants were able to exploit them to a high degree, that is, in Experiments 1-2. Specifically, this post-error slowing was predicted by an individual's win-rate, with higher win-rates predicting post-error slowing. In all situations across Experiments 1-2 and 5-6 where win-rates were on average at chance, either due to being fixed (Experiment 5) or due to randomized opponent choices (quasi-random in Experiments 1-2, fully random in Experiment 6), participants exhibited post-error speeding. Additionally, in Experiment 5, when win-rates were above chance in the fixed 50% win-rate RPS condition, participants also exhibited post-error speeding. In Experiment 6, even when participants could exploit the opponent, there was a significant post-error speeding effect. However, as mentioned in Chapter 4, participants in Experiment 6 achieved lower win-rates than participants in Experiments 1-2 on the individual level, the highest RPS win-rate in Experiment 6 having been 54.44% and the highest RPS win-rates in Experiments 1-2 having been 78.89% and 83.33%, respectively. Thus, it seems that win-rates need to be relatively highly above chance for post-error slowing to occur (see also Figures 2.5 and 2.9 in Chapter 2).

Interestingly, participants in Experiment 6 increased and decreased *lose-shift* responding appropriately even though reaction times were faster after losses. Thus, the more impulsive decisions made after losses did not fully prevent participants from learning. However, the rate of *lose-shift* responding changed less than that of *win-stay* responding: participants' responses after losses were generally more random than their responses after wins. This quick responding following losses may be the factor that prevented greater rates of learning. It could also explain the puzzling finding of post-error slowing in Experiments 1-2 co-occurring with the participants making the least optimal choices after losses. That is, the participants may have initially not paid much attention to what to do after losses and responded impulsively. After learning the optimal response after wins and achieving a high win-rate, losses became a rare event causing participants to pause, but as they had not learned the appropriate the strategy, responding was suboptimal. Another possibility is that after learning an optimal strategy and achieving a high win-rate, participants were more likely to find losses (caused by the exploitable opponents in Experiments 1-2 at times making random moves) as a sign of a change of pattern, and thus engaged in exploratory behaviour.

Assuming that longer reaction times are a marker of increased cognitive control (see Botvinick et al., 2001; Mackie et al, 2013), the results of the six experiments do not support the notion that this potential increased cognitive control helps people make better choice in all situations, as whether a choice is "good" or not is highly contextual. When the game opponents could not be exploited, participants were generally quicker in their responses after losses than after wins and also less likely to express the *lose-shift* bias than the *win-stay* bias on an individual level. This would normatively be considered a positive – the participants are playing closer to the mixed equilibrium strategy, despite shorter reaction times and thus assumedly less cognitive control. Likewise, more predictable *win-stay* responding (at least in MP) despite longer reaction times after wins suggests that assumedly increased cognitive control did not allow participants to play randomly in situations where they could not exploit the opponent. When opponents could be exploited in Experiment 6, participants increased or decreased their rate of *win-stay* responding more, coinciding with higher reaction times after wins. However, in Experiments 1-2 the participants engaged in post-error slowing yet made the most optimal choices after wins, again suggesting a dissociation.

Finally, two separate findings suggest that random responding after losses was not dependent on reaction times. First, the same trend of more random responding after losses than wins was observed in Experiments 1-2 and 6 when the participants could in fact exploit the opponents, despite differences in reaction time trends between these experiments. Second, despite no reaction time differences between decisions made after wins and losses in Experiments 3-4 due to task interruption, the participants in these experiments had a similar trend of reinforcement biases in MP as the participants in Experiments 5-6, with a higher likelihood of *win-stay* than *lose-shift* responding. In sum, either increased reaction times are not a reliable marker of cognitive control, or they are but cognitive control does not always predict more optimal responding even in situations where there is one single choice option (rather than a distribution of choice options) that participants should choose in order to be optimal. The optimality of play decisions seems to depend on the prior outcome and the optimal decision rule's alignment with reinforcement and inertia, separate from how quickly participants respond. Thus, the link between reaction times an optimal responding seems to be highly contextual. Again, my results underline the contextual nature of cognitive phenomena; slowing may be a sign of increased cognitive control only under specific circumstances.

5.4 Limitations and Ideas for Future Studies

There were design differences in the “random” opponents of the experiments that contained a “random” condition (Experiments 1-2, 4-6; see Table 5.1), with only Experiment 6 having a random condition with each trial having truly independently equal probabilities of any choice type by the opponent. In Experiments 1 and 2, the random opponent condition consisted of an opponent sampling without replacement from a flatly distributed set of the choice options. Sampling without replacement from a flat distribution was also how the *exploitable* opponents in Experiments 1-2 made their random choices for the trials during which they did not follow the predictable rule (30% of the trials). This decision was made in order to minimize the risk of any item biases in the random opponent’s choices, but the design may have led to *local* item biases in the random blocks for individual participants. However, if individual participants noticed these potential item biases in the random opponent and attempted to exploit them, this should have led to an increase in win-stay behaviour (repeat the move that the opponent is weak against due to their bias) and at the very least no reduction in lose-shift behaviour (as losing against an item-biased opponent should lead to a shift to the item the opponent is biased towards playing). If participants tracked the number of every move type the opponent had produced and then pre-emptively started overplaying the choice option that the opponent had been least likely to produce, this may have led to local increases in overall stay behaviour in the latter half of the block. This could have thus increased the rate of lose-stay choices and thus reduced the bias in the aggregate data. However, note that this design for the random opponent was identical to the design used by Dyson et al. (2016) and Forder & Dyson (2016), who reported a significant bias towards lose-shift in RPS (with a similar analysis method as the one used in Experiments 1 and 2). Thus, the design of the random opponent in and of itself does not seem to explain the lack of reinforcement biases in Experiments 1 and 2. The more plausible explanation would seem to be, based on Experiments 5 and 6, that the *lose-shift* bias is simply relatively rare among individuals in general regardless of game type, and thus may simply have failed to reach group-level significance in Experiments 1 and 2. Likewise, for the minority of random choices made by the exploitable opponents in Experiments 1-2, it does not seem that this could explain suboptimal responding after losses, as responding was more random after losses consistently across the experiments regardless of randomization method.

In Experiments 4 and 5, the random conditions were defined in terms of win-rate. This decision was made as the questions addressed in these two experiments necessitated controlled win-rates; in the random conditions, the win-rates were kept at chance-level. Thus, the randomization was conducted by randomizing a sequence consisting of equal numbers of each of outcome type for these conditions, equal to sampling without replacement from a flatly distributed set. Thus here, like in Experiments 1 and 2, the design choice may have led to situations where wins or losses (or draws) were clustered, but it is unclear how this could have led to noticeable changes in *stayshift* behaviour. As the general results of Experiment 3 and 4 suggested very little effect on reinforcement biases from different success slopes, any concern over the way in which outcomes in Experiments 4 and 5 were randomised seems unwarranted.

Experiment 6, the most direct test of the flexibility of biased responding in RPS and MP, suffered from a small sample due to time constraints. It was the only experiment reported here with independent equal probabilities for each choice the random opponents made, and also the only experiment that reported a significant *win-stay* bias in RPS (similarly to Alós-Ferrer & Ritschel, 2018). It is unclear how the methodological difference between Experiment 6 and the other experiments could cause the anomalous (among the six experiments) observation of a *win-stay* bias in RPS. It is unlikely that the difference would be caused by slight fluctuations in win-rate in the random condition, as this was the case in Experiments 1-2 as well. Moreover, the randomization of Experiments 5 would have caused slight fluctuations in the randomness of the random opponents' responses as well, as the responding happened as a function of what the participant chose in order to ensure a flat distribution of outcomes. This suggests that fluctuations in the opponents' response distribution also did not cause win-stay responding in RPS. However, the randomization in Experiment 5 did lead to a dynamic between the participants' and the opponents' choices, unlike in Experiment 6, though this contingency was not something the participants could in any way exploit. It is possible that some subtle effect of both the outcomes and the opponents' responses being completely independent of the participants' choices may increase win-stay responding in RPS, but why or how this would happen is unclear based on the present experiments. One should also note that the smaller sample size in Experiment 6, in addition to more cells in its analyses than in the other experiments, may give rise to false positives or false negatives – the anomalous observation may simply be an error caused by a small sample. In either case, future studies could assess the effects of different randomization methods and thus slight differences in the dependencies participants may observe and their effects on the emergence of reinforcement biases.

The reaction time analyses reported here only differentiated between decisions made after wins, losses and draws. They did not differentiate, however, between different decision types – that is, they do not tell us about the speed of *stay* or *shift* responses. This decision was made due to practical concerns about statistical power and to maintain similarity between the analyses. The decision was informed by the results of Forder and Dyson (2016), who found no differences between reaction times for decisions based on whether they aligned with reinforcement or not, and only found a main effect of outcome types. Due to this analysis method, all of the interpretations presented here about the connection between reaction times and reinforcement biases are indirect. It is possible that e.g. post-error speeding combined with more random responding after losses masks a pattern of *lose-shift* responses specifically being even faster than *lose-stay* responses. Likewise, it may be that slow responses after wins were driven by extremely slow *win-stay* responses. Alós-Ferrer and Ritschel (2018) found results that support this alternative interpretation. Specifically, in an RPS-like game, they found that reaction times for *win-stay* were faster than reaction times for any type of *win-shift*, and that the reaction times for *lose-shift*, specifically in the direction of the myopic best reply (or Cournot's best reply: see Dyson, 2019) were faster than for other decisions following losses. Given that Alós-Ferrer and Ritschel (2018) used a modified RPS-like game with no draws but instead wins and losses of different magnitudes, it is not clear whether this difference between their findings and those of Forder and Dyson (2016), who used a standard RPS game, is due to methodological differences or something else. In either case, future studies could assess reaction times in both RPS and MP, against both randomly and non-randomly playing opponents, to further the understanding of when biased responses are fast or slow.

The analyses of player confidence from Experiments 1-4 allowed no inferences other than that players were slightly more confident overall when playing against exploitable opponents (Experiments 1-2) and that players were generally slightly overconfident about their wins (Experiments 3-4). Given the way the analyses were run, they cannot clearly answer questions about the relationship between individual confidence and reinforcement biases on a trial-by-trial basis. As biased *lose-shift* responding was quite rare among both RPS and MP, and biased *win-stay* responding was more common at least in MP, a simple hypothesis for future studies would be that participants are more confident about reinforcement-based responding after wins than after losses. An analysis on this level would also help answer the question of whether individuals with strong *stayshift* biases are more confident, or whether they make the reinforcement-based responses automatically with confidence being separate from the likelihood of responding in a certain way.

In all of the experiments reported here, participants were not informed about the types of opponents they were playing against or whether to expect opponents playing randomly. Participants were simply informed that each of the opponents in the different blocks was a different opponent and that they might play the game differently from one another. In each experiment, the participants were also told that they should try and maximize their score, which would increase with each win and decrease with each loss (with draws, in the case of RPS conditions, having no effect on the score). These decisions were made to avoid expectancy effects and to maintain similarity between experiments: the aim was to have each experimental condition work as a separate learning task, with participants having to figure out if the opponent could be exploited or not. This approach could not eliminate any prior assumptions the participants may have had about zero-sum games. During debriefings throughout the six experiments, for example, a few participants indicated having chosen to try and ignore their score, outcomes, or the opponents' choices completely in order to play in a way that they assumed was the "correct" one. One participant in Experiment 6 indicated having decided to simply choose one option repeatedly in MP, expressing that they believed this was the “correct” way to act in a random game as their choices would not matter in the end. Another participant in Experiment 6 indicated having often repeated their decision in order to “prove it was right”. Thus, some participants may have avoided learning in conditions where learning would have been possible.

In the case of participants making repeated choices of only one item rigidly, the participant would be an outlier in terms of having both a high *win-stay* and a *lose-stay* bias as defined by statistical analyses, while clearly also having a choice pattern that is qualitatively different from what is usually meant by these biases. However, excluding participants like this from the data would be problematic, and neither repeated single-item responding nor participants reporting potentially irrational beliefs about the games were part of the exclusion criteria. Excluding these participants would be problematic as the aim of the six experiments was specifically to examine behaviour that could be considered irrational (reinforcement biases) – excluding certain types of irrationality but not others would be counterproductive if the experiments are to give a general picture of how common reinforcement biases are. Moreover, one can easily imagine similar, potentially irrational and idiosyncratic prior beliefs about MP or RPS (rather than lower-level properties of e.g. the reinforcement learning system or working memory) leading to reinforcement biases as well.

This is not to say that participants simply making one response type across several blocks may not have been simply unmotivated; only that discerning this is not without problems. However, for the specific case of single-item responding inflating the rate of win-stay and reducing the rate of lose-shift in the overall data, these “straight-liners” seem to not have been the primary cause for the differences between the two reinforcement bias types. There were a total of zero participants staying after wins or losses for 100% of the time in the unexploitable conditions of Experiments 1-2, a total of one throughout the conditions of Experiment 3, a total of three throughout the conditions of Experiment 4, a total of two throughout the conditions of Experiment 5, and a total of one in the random conditions of Experiment 6 (the aforementioned participant who indicated having behaved this way because they thought it was correct). In sum, these participants were such a minority that they could not have driven the effects observed here. In any case, future studies could also move towards a direction of examining individual differences in beliefs about zero-sum games specifically and testing whether in fact only participants with irrational beliefs also play irrationally, rather than the *stayshift* bias emerging regardless of e.g. a statement of a belief in the games being random. Other possible individual differences to measure could be gambling propensity and general sensitivity to reinforcement on the neural level (see section 5.1).

Another limitation for the current studies is caused by the within-subjects design. As each participant went through each experimental condition in an experiment, there may be carry-over effects from one condition to another. The conditions were counter-balanced in each experiment to mitigate this issue, and the experiment instructions emphasized to the participants that each block they played would be against a different game opponent that may play the game differently from the other opponents. However, these design decision in and of themselves may have led to the dilution of some effects. For example, it is possible that win-stay or lose shift biases are very common (or rare) in the initial stages of a participant getting used to the game task, and then decrease (or increase) as the participants play through a total of several hundreds of game trials. As the main variable of interest regarding reinforcement biases was the rate of win-stay or lose-shift choices throughout a block of trials, the analyses I have reported cannot detect temporal effects within a block. There is some evidence from the prior literature against the notion of changing rates of reinforcement biases as more rounds are played: Wilke and Barrett (2009) found no significant differences in *stayshift* responding as a function of time in a series of 100 trials of a binary choice task. However, the measure used by Wilke and Barrett (2009) combined rates of win-stay and lose-shift into one single measure, and there was a non-significant trend towards a decrease as a function in this measure. It is possible that this trend was due to an increase in either win-stay or lose-shift: however, it could not be due to a selective decrease in either without a slightly larger increase in the other. Thus, there is tentative evidence that the difference between win-stay and lose-shift is not due to participants decreasing their use of either rule as a function of time, but there is a possibility of an increase of either win-stay or lose-shift or a large decrease in one accompanied by a smaller decrease in the other.

Relatedly, the present experiments did not examine higher-order repetition effects, i.e. the relationship between a player's previous moves on their next decision. The focus of my studies was the stayshift heuristic, its predictors and its effects on learning. Dyson et al. (2016) found that different forms of shifts in RPS were more likely to lead to continued shifting rather than staying or random choices when playing against unexploitable opponents. Given that Dyson et al. (2016) found that shifting was also more likely following losses or draws, their results supported a notion of continued or cyclic biased decisions following outcomes other than wins. The present results cannot answer the question of whether this was the case with the minority of participants who exhibited a lose-shift bias, but this opens up possibilities for future experiments. Given the earlier finding of cyclic shifting on average, and a finding in the present experiments of only a handful of participants with a lose-shift bias, future studies could delve deeper into the dynamic of these biased participants. In addition to analyses of individual-level biases, future studies should examine whether the temporal patterns in rates of win-stay or lose-shift are different between biased and non-biased participants. To assess the aforementioned issues with longer-term temporal effects (carryover from one block to the next), an experiment could simply consist of several blocks of e.g. 90-120 rounds of a single game type with a fully randomly playing computer opponent in each block. Researchers could then test not only for differences in temporal trends between biased and non-biased participants (and potential individual differences behind these biases), but also whether a potential temporal increase or decrease carries from one computer opponent to the next.

The majority of the present experiments, with the exception of Experiment 2, did not examine the effects of financial rewards on reinforcement biases. Additionally, the incentive in Experiment 2 may have been too weak to bring out an effect. However, based on the results reported in this thesis, biases based on reinforcement and/or inertia seem to happen without financial incentives for succeeding in the experimental task. This suggests firstly that these biases may simply be a general function of learning and very flexible to abstract rewards, and secondly that experiments in cognitive bias in games may not absolutely require such incentives (see Read, 2005). To fully examine the effects of concrete rewards on biases, researchers should not only conduct studies comparing the effects of low and high rewards, but also comparing the effects of rewards to behaviour that is not financially incentivized. As discussed above, it is possible that the participants most likely to express biases on an individual level may be individuals highly sensitive to reward in general; it is not clear whether the rate of individuals with a bias would increase or not with added financial incentive. In order to understand the potential larger implications of reinforcement biases outside of the laboratory, researchers need to be careful in examining all the factors that increase or decrease such biases.

5.5 Conclusion

The results of the six experiments replicate some earlier findings but also provide some conflicting results. The results suggest a difference in reinforcement-based biased responding between RPS and MP, likely stemming from higher memory demands in the former leading to less *win-stay* and potentially more *lose-shift* responding. The results suggest that at least in samples consisting mostly of young, Western university students, reinforcement biases in zero-sum games on the individual level are relatively rare, with lose-shift biases being less likely than win-stay biases. This further suggests that a true win-stay, lose-shift bias was very rare among my participants, and thus future studies should take care to examine biased responding separately for different outcomes. Overall group-level biases did not replicate consistently, likely stemming from the low frequency of individuals with biases in any of the experiments, suggesting for future studies that assessing reinforcement biases on an individual level may give clearer results. However, there was never an overall trend in the opposite direction of *lose-stay* or *win-shift*, suggesting that reinforcement-based biases are more common.

Moreover, reinforcement biases are not trivial, as overall participants were better at exploiting their opponents when the opponents could be exploited by increasing reinforcement-based responding rather than decreasing it. This was true more for win-stay than lose-shift, again highlighting the difference between the two. The difference between the two bias types likely stems from the inherent difference in the roles of reinforcement and inertia in relation to wins and losses: after a win, inertia and reinforcement are aligned, whereas after losses they are in conflict. Due to this difference, it seems incorrect to consider reinforcement biases an instance of a positive recency effect (i.e. the Hot Hand Fallacy) as that would imply roughly equal win-stay and lose-shift responding. Future experiments should address potential individual differences such as beliefs about zero-sum games or reward sensitivity as predictors of the high variability between people. Moreover, the non-triviality of the biases is emphasised by the fact that I observed biased responding even in the absence of any financial rewards. This suggests that the biases are a general behavioural tendency, and opens up potential future studies dissociating the effects of rewards from the base level of biased responding.

The results support the notion of post-error slowing in zero-sum games happening only when losses are a rare event. Whether participants slowed down or not also does not seem to have been a predictor of whether they made choices optimally even when they could exploit an opponent. A common trend of post-error speeding when win-rates were low and/or opponents were making their choices randomly, combined with often more random responding after losses, seems to imply normatively "better" decisions after losses despite quicker reaction times. For optimal behaviour after wins, there was a similar dissociation, with participants being generally better at exploiting opponents on rounds following a win despite differences in reaction time trends between different exploitable conditions across the experiments. Future experiments could delve deeper into the specific speed of reinforcement-biased responding and other types of responding under *unexploitable* and *exploitable* conditions to illuminate what predicts learning or impulsive decisions.

In sum, the results show substantial variance and relative infrequency in a supposedly common cognitive bias, and dissociations both between two types of biased responding that are often considered a single property and between decision impulsivity and “optimal“ decisions. The results call for caution against assumptions of reinforcement biases being ubiquitous, but also demonstrate that if and when people are biased, the biases do not simply disappear when people are incentivized to act against the bias.

APPENDICES

Appendix 1

Anthropomorphism questionnaire items (from Epley, Akalis, Waytz, & Cacioppo, 2008).

“My opponent had a mind of its own”

“My opponent had intentions”

“My opponent had free will”

“My opponent had consciousness”

“My opponent experienced emotions”

“My opponent was attractive”

“My opponent was efficient”

“My opponent was strong”

Game Engagement Questionnaire (GEQ) items (from Brockmyer et al., 2009).

“I lost track of time”

“Things seemed to happen automatically”

“I felt different”

“I felt scared”

“The game felt real”

“If someone would have talked to me, I wouldn’t have heard them”

“I got wound up”

“Time seemed to kind of stand still or stop”

“I felt spaced out”

“I wouldn’t have answered if someone talked to me”

“I couldn’t tell that I was getting tired”

“Playing seemed automatic”

“My thoughts went fast”

“I lost track of where I am”

“I played without thinking about how to play”

“Playing made me feel calm”

Modified co-presence questionnaire items (from Nowak & Biocca, 2003 and Forder & Dyson, 2016).

From Nowak & Biocca (2003):

“My opponent was intensely involved in our game”

“My opponent seemed to find our game stimulating”

From Forder and Dyson (2016):

“I felt as though my opponent had a strategy that was based on the moves I was making”

“I felt as though my opponent had a strategy that was based on the moves it was making”

“My opponent exhibited a human-like strategy”

“I felt like my opponent was somehow cheating”

“I found this game of RPS rewarding to play”

Appendix 2

Experiment 2 questionnaire items and descriptive statistics.

	Unexploitable		Exploitable	
	Low value	High value	Low value	High value
"I think the opponent was responding to my moves"	6.700 (2.662)	6.500 (2.864)	4.700 (3.172)	5.300 (3.421)
"I think the opponent played according to a pattern"	5.675 (2.787)	5.125 (2.794)	8.275 (2.909)	8.275 (2.773)
"I think my opponent was somehow cheating"	3.225 (2.626)	3.350 (2.905)	2.300 (2.066)	2.825 (2.541)
"I think my opponent changed their strategy at some point during the block"	6.900 (2.872)	5.775 (2.869)	6.350 (3.527)	6.675 (3.075)
"I think my opponent could predict what I was doing"	6.225 (2.833)	5.600 (3.241)	4.225 (2.806)	4.400 (2.816)
"I think my opponent was playing randomly"	3.800 (2.452)	3.800 (2.594)	2.825 (2.171)	2.550 (1.663)

Note: standard devation in parentheses.

Appendix 3

Rotter's Locus of Control Scale

For each question select the statement that you agree with the most

1. a. Children get into trouble because their parents punish them too much.

b. The trouble with most children nowadays is that their parents are too easy with them.

2. a. Many of the unhappy things in people's lives are partly due to bad luck.

b. People's misfortunes result from the mistakes they make.

3. a. One of the major reasons why we have wars is because people don't take enough interest in politics.

b. There will always be wars, no matter how hard people try to prevent them.

4. a. In the long run people get the respect they deserve in this world.

b. Unfortunately, an individual's worth often passes unrecognized no matter how hard he tries.

5. a. The idea that teachers are unfair to students is nonsense.

b. Most students don't realize the extent to which their grades are influenced by accidental happenings.

6. a. Without the right breaks one cannot be an effective leader.

b. Capable people who fail to become leaders hive not taken advantage of their opportunities.

7. a. No matter how hard you try some people just don't like you.

b. People who can't get others to like them don't understand how to get along with others.

8.a. Heredity plays the major role in determining one's personality

b. It is one's experiences in life which determine what they're like.

9. a. I have often found that what is going to happen will happen.

b. Trusting to fate has never turned out as well for me as making a decision to take a definite course of action.

10. a. In the case of the well prepared student there is rarely if ever such a thing as an unfair test.

b. Many times exam questions tend to be so unrelated to course work that studying in really useless.

11. a. Becoming a success is a matter of hard work, luck has little or nothing to do with it.

b. Getting a good job depends mainly on being in the right place at the right time.

12. a. The average citizen can have an influence in government decisions.

b. This world is run by the few people in power, and there is not much the little guy can do about it.

13.a. When I make plans, I am almost certain that I can make them work.

b. It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow.

14. a. There are certain people who are just no good.

b. There is some good in everybody.

15. a. In my case getting what I want has little or nothing to do with luck.

b. Many times we might just as well decide what to do by flipping a coin.

16. a. Who gets to be the boss often depends on who was lucky enough to be in the right place first.

b. Getting people to do the right thing depends upon ability. Luck has little or nothing to do with it.

17.a. As far as world affairs are concerned, most of us are the victims of forces we can neither understand, nor control.

b. By taking an active part in political and social affairs the people can control world events.

18. a. Most people don't realize the extent to which their lives are controlled by accidental happenings.

b. There really is no such thing as "luck."

19. a. One should always be willing to admit mistakes.

b. It is usually best to cover up one's mistakes.

20.a. It is hard to know whether or not a person really likes you.

b. How many friends you have depends upon how nice a person you are.

21. a. In the long run the bad things that happento us are balanced by the good ones.

b. Most misfortunes are the result of lack of ability, ignorance, laziness, or all three.

22. a. With enough effort we can wipe out political corruption.

b. It is difficult for people to have much control over the things politicians do in office.

23. a. Sometimes I can't understand how teachers arrive at the grades they give.

b. There is a direct connection between how hard 1 study and the grades I get.

24. a. A good leader expects people to decide for themselves what they should do.

b. A good leader makes it clear to everybody what their jobs are.

25. a. Many times I feel that I have little influence over the things that happen to me.

b. It is impossible for me to believe that chance or luck plays an important role in my life.

26. a. People are lonely because they don't try to be friendly.

b. There's not much use in trying too hard to please people, if they like you, they like you.

27. a. There is too much emphasis on athletics in high school.

b. Team sports are an excellent way to build character.

28. a. What happens to me is my own doing.

b. Sometimes I feel that I don't have enough control over the direction my life is taking.

29. a. Most of the time I can't understand why politicians behave the way they do.

b. In the long run the people are responsible for bad government on a national as well as on a local level.

Scoring:

One point for each of the following:

2.a, 3.b, 4.b, 5.b, 6.a, 7.a, 9.a, 10.b, 11.b, 12.b, 13.b, 15.b, 16.a, 17.a, 18.a, 20.a, 21.a, 22.b, 23.a, 25.a, 26.b, 28.b, 29.a.

A high score = External Locus of Control

A low score = Internal Locus of Control

REFERENCES

Achtziger, A., & Alós-Ferrer, C. (2013). Fast or Rational? A Response-Times Study of Bayesian Updating. *Management Science*, *60*(4), 923–938. https://doi.org/10.1287/mnsc.2013.1793

Achtziger, A., Alós-Ferrer, C., Hügelschäfer, S., & Steinhauser, M. (2015). Higher incentives can impair performance: Neural evidence on reinforcement and rationality. *Social Cognitive and Affective Neuroscience*, *10*(11), 1477–1483. https://doi.org/10.1093/scan/nsv036

Alós-Ferrer, C., Hügelschäfer, S., & Li, J. (2016). Inertia and Decision Making. *Frontiers in Psychology*, *7*(February), 1–9. https://doi.org/10.3389/fpsyg.2016.00169

Alós-Ferrer, C., & Ritschel, A. (2018). The reinforcement heuristic in normal form games. *Journal of Economic Behavior and Organization*, *152*, 224–234. https://doi.org/10.1016/j.jebo.2018.06.014

Arthur, W. B. (1994). Inductive reasoning and bounded rationality. *The American Economic Review*, *84*(2).

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory and Cognition*, *32*(8), 1369–1378. https://doi.org/10.3758/BF03206327

Ayton, P., Hunt, A. J., & Wright, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making*, *2*(4), 221–238. https://doi.org/10.1002/bdm.3960020403

Baddeley, A. D. (1966). The capacity for generating information by randomization. *The Quarterly Journal of Experimental Psychology*, *18*(2), 119–129. https://doi.org/10.1080/14640746608400019

Bar-Hillel, M., Peer, E., & Acquisti, A. (2014). 'Heads or tails?'-a reachability bias in binary choice. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(6), 1656–1663. https://doi.org/10.1037/xlm0000005

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, *12*(4), 428–454. https://doi.org/10.1016/0196-8858(91)90029-I

Barron, G., & Leider, S. (2010). The Role of Experience in the Gambler's Fallacy. *The Journal of Behavioral Decision Making*, *23*, 117–129. https://doi.org/10.1002/bdm

Block, J., & Kremen, A. M. (1996). IQ and ego-resiliency: Conceptual and empirical connections and separateness. *Journal of Personality and Social Psychology*, *70*(2), 349–361. https://doi.org/10.1037/0022-3514.70.2.349

Bornstein, G., Gneezy, U., & Nagel, R. (2002). The effect of intergroup competition on group coordination: An experimental study. *Games and Economic Behavior*, *41*(1), 1–25. https://doi.org/10.1016/S0899-8256(02)00012-X

Botvinick, M. M., Carter, C. S., Braver, T. S., Barch, D. M., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*. https://doi.org/10.1037/0033-295X.108.3.624

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., Mcbroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire : A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, *45*(4), 624–634. https://doi.org/10.1016/j.jesp.2009.02.016

Burger, J. M. (1986). Desire for control and the illusion of control: The effects of familiarity and sequence of outcomes. *Journal of Research in Personality*, *20*(1), 66–76. https://doi.org/10.1016/0092-6566(86)90110-8

Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: “Gambler’s fallacy” versus “hot hand.” *Psychonomic Bulletin & Review*, *11*(1), 179–184.

Charness, G., & Levin, D. (2005). When optimal choices feel wrong: A laboratory study of bayesian updating, complexity, and affect. *American Economic Review*, *95*(4), 1300–1309. https://doi.org/10.1257/0002828054825583

Collins, A. G. E., & Frank, M. J. (2013). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis, *35*(7), 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x.How

Coventry, K. R., & Norman, A. C. (1998). Arousal, erroneous verbalizations and the illusion of control during a computer-generated gambling task. *British Journal of Psychology*, *89*, 629–645.

Crysel, L., Crosier, B. S., & Webster, G. (2013). The Dark Triad and Risk Behavior. *Personality and Individual Differences*, *54*(1), 35–40.

Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology*, *2*(SEP), 1–10. https://doi.org/10.3389/fpsyg.2011.00233

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110. https://doi.org/10.1038/nn.3279

De Vries, R. E., De Vries, A., & Feij, J. A. (2009). Sensation seeking, risk-taking, and the HEXACO model of personality. *Personality and Individual Differences*, *47*(6), 536–540. https://doi.org/10.1016/j.paid.2009.05.029

Dudschig, C., & Jentzsch, I. (2009). Speeding before and slowing after errors: Is it all just strategy? *Brain Research*, *1296*, 56–62. https://doi.org/10.1016/j.brainres.2009.08.009

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454–465. https://doi.org/10.3758/s13414-011-0243-2

Dyson, B. J. (2019). Recursive Thought in Non-Transitive Game Strategy. *Games*, *10*(32), 1–14.

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*, 1–7. https://doi.org/10.1038/srep20479

Ejova, A., Navarro, D. J., & Delfabbro, P. H. (2013). Success-slope effects on the illusion of control and on remembered success-frequency. *Judgment & Decision Making*, *8*(4), 498–511.

Epley, N., Akalis, S., Waytz, A., & Cacioppo, J. T. (2008). Creating Social Connection Through Inferential Reproduction. *Psychological Science*, *19*(2), 114–120. https://doi.org/10.1111/j.1467-9280.2008.02056.x

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*, 1–8. https://doi.org/10.1038/s41562-016-0002

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*. https://doi.org/10.1038/srep33809

Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *American Statistician*, *60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Gigerenzer, G., Goldstein, D. G., Owen, G., Nordgaard, J., Jansson, L., Sæbye, D., … West, R. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1192/bjpo.bp.115.000224

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295–314. https://doi.org/10.1016/0010-0285(85)90010-6

Goodie, A. S., Doshi, P., & Young, D. L. (2012). Levels of theory-of-mind reasoning in competitive games. *The Journal of Behavioral Decision Making*, *25*, 95–108. https://doi.org/10.1002/bdm

Gronchi, G., & Sloman, S. a. (2008). Do causal beliefs influence the Hot-Hand and the Gambler’s Fallacy? *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*, 1164–1168.

Gruber, A. J., & Thapa, R. (2016). The memory trace supporting lose-shift responding decays rapidly after reward omission and is distinct from other learning mechanisms in rats. *ENeuro*, *3*(6), 1–14. https://doi.org/10.1523/ENEURO.0167-16.2016

Ivan, V. E., Banks, P. J., Goodfellow, K., & Gruber, A. J. (2018). Lose-shift responding in humans is promoted by increased cognitive load. *Frontiers in Integrative Neuroscience*, *12*(March), 1–9. https://doi.org/10.3389/fnint.2018.00009

Jentzsch, I., & Dudschig, C. (2009). Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *Quarterly Journal of Experimental Psychology*, *62*(2), 209–218. https://doi.org/10.1080/17470210802240655

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Lange, R. V., & Tiggemann, M. (1981). Dimensionality and reliability of the Rotter l-E Locus of Control scale. *Journal of Personality Assessment*, *45*(4), 398–406. https://doi.org/10.1207/s15327752jpa4504_9

Langer, E J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311–328. https://doi.org/10.1037//0022-3514.32.2.311

Langer, Ellen J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, *32*(6), 951–955. https://doi.org/10.1037/0022-3514.32.6.951

Lee, D., Conroy, M. L., McGreevy, B. P., & Barraclough, D. J. (2004). Reinforcement learning and decision making in monkeys during a competitive game. *Cognitive Brain Research*, *22*(1), 45–58. https://doi.org/10.1016/j.cogbrainres.2004.07.007

Lee, D., McGreevy, B. P., & Barraclough, D. J. (2005). Learning and decision making in monkeys during a rock-paper-scissors game. *Cognitive Brain Research*, *25*(2), 416–430. https://doi.org/10.1016/j.cogbrainres.2005.07.003

Lee, K., & Ashton, M. C. (2014). The Dark Triad, the Big Five, and the HEXACO model. *Personality and Individual Differences*, *67*, 2–5.

https://doi.org/10.1016/j.paid.2014.01.048

Losecaat Vermeer, A. B., & Sanfey, A. G. (2015). The effect of positive and negative feedback on risk-taking across different contexts. *PLoS ONE*, *10*(9), 1–13. https://doi.org/10.1371/journal.pone.0139010

Ma, Q., Jin, J., Meng, L., & Shen, Q. (2014). The dark side of monetary incentive. *NeuroReport*, *25*(3), 194–198. https://doi.org/10.1097/WNR.0000000000000113

Mackie, M.-A., Van Dam, N. T., & Fan, J. (2013). Cognitive Control and Attentional Functions. *Brain and Cognition*, *82*(3), 301–312. https://doi.org/10.1016/j.bandc.2013.05.004.Cognitive

Matute, H. (1995). Human Reactions to Uncontrollable outcomes: further evidence for superstitions rather than helplessness. *The Quarterly Journal of Experimental Psychology Section B*, *48*(2), 142–157. https://doi.org/10.1080/14640749508401444

McKay, R., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior*, *31*(5), 309–319. https://doi.org/10.1016/j.evolhumbehav.2010.04.005

Mehta, B. J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, *84*(3), 658–673.

Miller, J. B., & Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica*, *86*(6), 2019–2047. https://doi.org/10.3982/ecta14943

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, *36*(1), 48–49.

https://doi.org/10.1073/pnas.36.1.48

Neuringer, A. (1986). Can people behave “randomly?”: The role of feedback. *Journal of Experimental Psychology: General*, *115*(1), 62–75. https://doi.org/10.1037/0096-3445.115.1.62

Neurobehavioral Systems. (n.d.). Cognitive Experiments II v5. http://www.neurobs.com/menu_presentation/menu_teac

Neurobehavioral Systems. (n.d.). Cognitive Experiments IV v2. http://www.neurobs.com/menu_presentation/menu_teaching/exp_pack?pack_id=9

Notebaert, W., Houtman, F., Opstal, F. Van, Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition*, *111*(2), 275–279. https://doi.org/10.1016/j.cognition.2009.02.002

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *12*(5), 481–494. https://doi.org/10.1162/105474603322761289

Parkes, K. R. (1985). Dimesionality of Rotter’s locus of control scale: An application of the “very simple structure” technique. *Personality and Individual Differences*, *6*(1), 115–119. https://doi.org/10.1016/0191-8869(85)90036-4

Rabbitt, P. M., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology*, *29*(December 2014), 727–743. https://doi.org/10.1080/14640747708400645

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rapoport, A., & Budescu, D. V. (1992). Generation of random series in two-person strictly competitive games, *121*(3), 352–363.

Read, D. (2005). Monetary incentives, what are they good for? *Journal of Economic Methodology*, *12*(2), 265–276. https://doi.org/10.1080/13501780500086180

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, *80*(1, Whole No. 609).

Scheibehenne, B., Wilke, A., & Todd, P. M. (2011). Expectations of clumpy resources influence predictions of sequential events. *Evolution and Human Behavior*, *32*(5), 326–333. https://doi.org/10.1016/j.evolhumbehav.2010.11.003

Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00474

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*, 1482–1492. https://doi.org/10.1111/cogs.12115

Sundali, J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making*, *1*(1), 1–12. http://journal.sjdm.org/jdm06001.pdf

Terhune, D. B., & Brugger, P. (2011). Doing better by getting worse: Posthypnotic amnesia improves random number generation. *PLoS ONE*, *6*(12), 8–11. https://doi.org/10.1371/journal.pone.0029206

Thompson, S. C., Armstrong, W., & Thomas, C. (1998). Illusions of control, underestimations, and accuracy: A control heuristic explanation. *Psychological Bulletin*, *123*(2), 143–161. https://doi.org/10.1037/0033-2909.123.2.143

Thorndike, E. L. (1911). *Animal Intelligence. Experimental studies.* New York: Macmillan Company.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Tyszka, T., Zielonka, P., Dacey, R., & Sawicki, P. (2008). Perception of randomness and predicting uncertain events. *Thinking and Reasoning*, *14*(1), 83–110. https://doi.org/10.1080/13546780701677669

Unsworth, N., Heitz, R. P., & Engle, R. W. (2005). An automated version of the operation span task. *Behaviour Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/bf03192720

Verbruggen, F., Chambers, C. D., Lawrence, N. S., & McLaren, I. P. L. (2017). Winning and losing: Effects on impulsive action. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 147–168. https://doi.org/10.1037/xhp0000284

Wagenaar, W. A. (1991). Randomness and randomizers: Maybe the problem is not so big. *Behavioral Decision Making*, *4*, 220–222.

Wang, Z., Xu, B., & Zhou, H.-J. (2014). Social cycling and conditional responses in the Rock-Paper-Scissors game. *Scientific Reports*, *4*, 5830. https://doi.org/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. https://doi.org/10.1016/S1389-0417(00)00014-0

Wilke, A., & Barrett, H. C. (2009). The hot hand phenomenon as a cognitive adaptation to clumped resources. *Evolution and Human Behavior*, *30*(3), 161–169. https://doi.org/10.1016/j.evolhumbehav.2008.11.004

Wilke, A., Scheibehenne, B., Gaissmaier, W., McCanney, P., & Barrett, H. C. (2014). Illusionary pattern detection in habitual gamblers. *Evolution and Human Behavior*, *35*(4), 291–297. https://doi.org/10.1016/j.evolhumbehav.2014.02.010

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*(6), 2074–2081. https://doi.org/10.1037/a0038199

Wolford, G., Newman, S. E., Miller, M. B., & Wig, G. S. (2004). Searching for patterns in random sequences. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *58*(4), 221–228. https://doi.org/10.1037/h0087446

Worthy, D. A., Hawthorne, M. J., & Otto, A. R. (2012). Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. https://doi.org/10.3758/s13423-012-0324-9