|  |
| --- |
| **MSc Cognitive and Decision Sciences 2017-18**    Transfer of Learning through Opponent Modelling in Strategic Interaction  Candidate Number: TSJQ7  Supervisor: Dr Maarten Speekenbrink  Date: 13 August 2018  Word Count: 11944  **Project Assessment Form**  **Candidate: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  **Name of Supervisor: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |

Please use the table below in conjunction with the **PALS Guide for Class Descriptors for PGT Project Marking** which has been sent to you.

Please note that you are not required to give numerical grades in the table below, just tick the relevant box

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Grade** | **80%+** | **70-79%** | **60-69%** | **50-59%** | **40-49%** | **Below 40%** |
| **Classification** | **High Distinction** | **Distinction** | **Merit** | **Pass** | **Fail** | **Fail** |
| General |  |  |  |  |  |  |
| Literature Review |  |  |  |  |  |  |
| Method |  |  |  |  |  |  |
| Results |  |  |  |  |  |  |
| Discussion |  |  |  |  |  |  |

|  |
| --- |
| Additional comments: |

|  |  |
| --- | --- |
| **AGREED GRADE** |  |

**(Please leave your individual mark overleaf)**

**First Marker**

|  |  |
| --- | --- |
| Name |  |
| Signature |  |
| Date |  |
| Grade |  |

**Second Marker**

|  |  |
| --- | --- |
| Name |  |
| Signature |  |
| Date |  |
| Grade |  |

Table of Contents

[Abstract 6](#_Toc521928466)

[Introduction 7](#_Toc521928467)

[Traditional game theory 7](#_Toc521928468)

[Learning in strategic interaction: contribution of behavioural game theory 9](#_Toc521928469)

[Learning transfer across games 16](#_Toc521928470)

[The use of computer agents to study strategic interaction 19](#_Toc521928471)

[The present study 21](#_Toc521928472)

[Method 24](#_Toc521928473)

[Participants and design 24](#_Toc521928474)

[Task 24](#_Toc521928475)

[Experiment Procedure 26](#_Toc521928476)

[Results 29](#_Toc521928477)

[Performance 29](#_Toc521928478)

[Within and between-game learning 30](#_Toc521928479)

[Differences in within-Game learning between games 33](#_Toc521928480)

[Transfer of learning between games 35](#_Toc521928481)

[Moderation of within and between-game learning by type of opponent: 37](#_Toc521928482)

[Computational modelling of Human strategies 41](#_Toc521928483)

[Discussion: 46](#_Toc521928484)

[Conclusion 50](#_Toc521928485)

# **Abstract**

Learning transfer and generalisation are being studied extensively in cognitive psychology, education as well as artificial intelligence as they are considered one of the hallmarks of Human intelligence. In the domain of strategic interaction, most research on transfer of learning has used human participants as opponents and focused almost exclusively on action-based strategies to explain behaviour. By contrast, in this study, participants faced the same computer agent implementing reasoning-based level-k strategies in multiple games that have varying degrees of similarity. We find that the majority of human participants not only adapt their play to best respond to the opponent reasoning strategy, but also transfer this knowledge between games to different environments. Opponent modelling is offered as the most likely explanation of this ability, and it is shown that it is modulated by the similarity between environments as well as the degree of strategic sophistication of the computer agent.

# 

# **Introduction**

This research addresses how humans model their opponents in the context of strategic interaction, and more specifically while engaged in games repeatedly against the same opponent. Strategic interaction refers to situations where decisions are inter-dependent, and where decision-making needs to consider opponent decisions as well as the decision maker’s preferred outcomes. This has historically been studied in economics and mathematics using a multitude of games and social situations (Camerer, 2011), as well as real dilemmas such as principal-agent relations, negotiation and moral quandaries (Eisenhardt, 1989).

In this study, we are interested in Games. Games are composed of a set of players, each with a finite associated set of choices (also called pure strategies) and corresponding to each player, a payoff function which maps the set of choices to rewards. For instance, a 3x3 player game refers to games of two players, with each having the possibility to select one of 3 choices. Each combination of the two player’s selections will be mapped onto a reward vector describing the payoffs for each player. Strategies are a set of instructions, or decision rules, specifying which actions a player should take. Pure strategies define exactly what a player should do for any situation faced. A mixed-strategy by contrast, is a probability distribution over pure strategies, allowing players to select probabilistically which action to choose. This definition encompasses a wide variety of games and the mathematical study of how these games should be played under mathematical assumptions of rationality is the aim of traditional game theory. In the next section, we will review the foundational aspects of the field.

## **Traditional game theory**

The field of game theory developed initially as a branch of applied mathematics, whereby optimal strategies for playing the game were derived under a set of restrictive assumptions about the game and the players. The publication of “The Theory of Games and Economic Behavior" by Morgenstern & Von Neumman (1953) provided the axiomatic theory of expected utility, which allowed mathematical statisticians and economists to deal with decision-making under uncertainty. John Nash, in his 1951 paper, “equilibrium points in n-person games”, generalized this framework and provided a way to analyse the strategic interactions of several decision makers in order to predict their actions given a set of assumptions. The concept of Nash equilibrium is based on the insight that strategies should be chosen such that no player can do better by unilaterally deviating from it.

To derive elegant mathematical solutions to the puzzle of finding optimal playing strategies in these multiplayer games, the equilibrium concepts such as that developed by John Nash (1950) or the strong equilibrium of Aumann (1959) are based on a strong set of assumptions. Namely that the joint strategy space, the payoff structure, and the rationality of the players are all common knowledge. The latter assumption for instance translates to the idea that players are rational, in the sense that they will strive to maximise their utility and can do so without cognitive or other limitations, and that this is known by all players. Furthermore, players know that everyone knows everyone is rational, that everyone knows this, and so on.

Nash equilibrium and subsequent advances in game theory didn’t encounter success in psychology as psychologists have known for some time that the underlying assumptions are far from an accurate description of how humans typically interact. Empirical results confirm that players do not behave like rational self-interested utility maximisers, and that deviations from the predictions of game theory are large and consistent enough that it wouldn’t be prudent to deem them inexplicable irrationalities (Doucouliagos, 1994; Binmore, 1987; Camerer, 1997). More specifically, multiple studies conclude that humans don’t play according to the Nash equilibrium predictions (Brown & Rosenthal, 1990; Rosenthal et al., 2003). One notable issue stems from what it implied for human’s ability to mentalise about others, known in cognitive psychology as “theory of mind” (Premack & Woodruff, 1978). The rationality assumption and the fact that it is common knowledge implies that participants can use theory of mind ad infinitum to infer their opponent’s intentions (Osborne & Rubinstein, 1994; Bicchieri, 1988). Later work in game theory (Daskalakis et al., 2009) has shown that calculating Nash equilibrium is an intractable algorithmic problem in many situations, challenging enough for powerful machines, let alone humans with inherent cognitive constraints.

Simon (1972) put forward the idea of bounded rationality as a more accurate basis for modelling human decision making. It’s the insight that when humans make decisions, their “rationality” is necessarily limited by the complexity of the problem at hand, the cognitive limitations of the mind, as well as the time available to them. Therefore, humans will seek to “satisfice”, meaning they will opt for a “good enough” solution based on a simplified model of the environment rather than an optimal one. Economists and cognitive psychologists have taken inspiration from this idea to build models of decision making that fit better with the experimental data on strategic interaction in games.

## **Learning in strategic interaction: contribution of behavioural game theory**

Much research on boundedly rational behaviour in games comes from behavioural game theory, a nascent field that is primarily interested in what players “actually” do. It tests and tries to explain deviations from traditional game theory predictions in experimental settings, using methods of experimental economics and psychology. In this section, we will review models from behavioural game theory, competing to explain deviations from traditional game theory predictions. We will discuss the contributions of five main models of behaviour, from models that categorise player’s strategic sophistication such as level-k theory and cognitive hierarchy models, to approaches that use associative learning processes to explain observed behaviour in games such as reinforcement learning. We will conclude the review with models that explicitly formalise how players base their decisions on beliefs about the likely actions of their opponents, such as fictitious play and patter detection models.

### **Level-k theory:**

The concept of bounded rationality in games finds its natural expression in “level-k” theory, used to explain systematic deviations from Nash equilibrium in single shot games by assuming participants have heterogenous and inconsistent beliefs about the opponents they face (Nagel, 1995; Stahl and Wilson, 1994, 1995; Ho et al., 1998). It is also a useful framework for understanding how players might model their opponents as it categorises decision makers into various groups depending on the depth of their reasoning about others. The level player is non-strategic and plays randomly or follows a salient strategy given the environment (Arad & Rubinstein, 2012). level player best responds assuming all his opponents are Similarly, a level player best responds to the belief that all other players are level players, and so on. The level-k theory is naturally linked to theory of mind abilities in humans. People routinely represent other people’s mental states so as to predict their behaviour, something akin to level-2 thinking. It is also common for humans to model what other people think about them and make decisions based on that or go even further (e.g. Alice thinks that Bob knows that she is aware of his indiscretions).

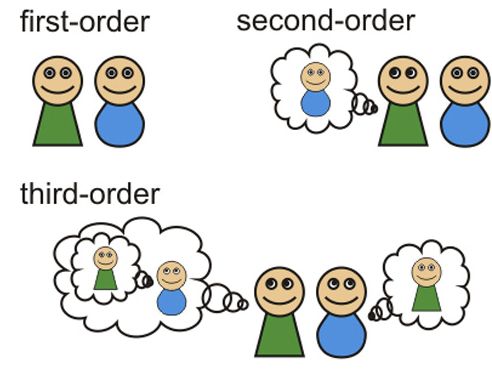
**

Figure 1: Graphical description of various levels of theory of mind

There is a vast literature using level-k models to successfully fit participants’ choices in a variety of strategic games (Crawford & Irriberri, 2007; Östling et al., 2011; Arad and Rubinstein, 2012) or attempting to identify the distribution of levels in the population (Stahl & Wilson, 1995; Costa-Gomes & Crawford, 2006). However, Georganas *et al*. (2015) show the difficulty of finding consistency in these estimates of level-k for players undertaking different classes of games. This lack of consistency could potentially be explained by the fact that the estimated level of participants reflects their beliefs about their opponent’s sophistication rather than an innate cognitive bound (Agranov et al., 2012). Since different game types can elicit different beliefs about opponent’s sophistication, the estimated levels will naturally vary from on type of game to the other.

### **Cognitive hierarchy models**

A similar model from the behavioural game theory literature is known as the cognitive hierarchy model (Camerer et al. 2004). The key idea here is that people assume that their opponents are drawn from a distribution of lower level-k reasoning abilities and as such the best response depends on the nature of that distribution. Levels 0 and 1 are the same as in level-k model, level 2 and higher are assumed however to play as if the opponents are a mixture of lower levels, and best respond against a weighted average of what these lower-level other players do.

A major limitation of level-k and cognitive hierarchy models is that they have mostly been used to test whether players actions are consistent with game theory predictions in one-shot games, thereby ignoring the learning dynamics in the off-equilibrium path that would occur with repeated play. Studying repeated games and learning is important for multiple reasons: First, because they represent a better model of real world interaction in multiple situations. Most of our social interaction are with people that we face repeatedly, and our decisions on how best to act, cooperate or compete are informed by our prior experience with that person. Decisions about whether to compete or cooperate with a colleague or which commute is best for avoiding traffic jams are all examples of repeated games. Second, repeated games are a powerful framework since they capture the idea that a player will have to consider the impact of his or her current action on the future actions of other players, therefore giving a more important role to concepts such as mental representations of opponents that are closely related to work on theory of mind. Third because empirically, if theoretical equilibrium is observed, it is rarely reached in the initial stages of play, and only develops after players have had the opportunity to observe and learn from their opponent’s actions (Mookherjee & Sopher, 1994; Erev & Roth, 1998).

### **Reinforcement Learning**

A successful approach to explaining players actions in repeated games over time uses reinforcement learning, whereby players are more likely to repeat previous strategies that yielded a good payoff (reinforcement) and less prone to choosing actions that did not yield a high payoff, weakening the likelihood they will be played again. Originally developed as an artificial intelligence learning algorithm (Sutton & Barto, 1998), this approach is akin to “rote” or “implicit” learning (Holyoak and Spellman, 1993) whereby players learn mappings between actions and rewards, without an explicit model of the strategic structure of the game. Erev & Roth (1998) show that a single parameter reinforcement learning model does a better job explaining repeated play than the assumption of instantaneous Nash equilibrium. Moreover, allowing for the decision to rely on a model of the opponent leads to an improvement in the performance of the model.

### **Belief learning**

This idea of opponent modelling is at the heart of another successful approach to model learning in repeated games. Belief learning, generally speaking, is a set of models where each player updates a belief about what others will do in the future based solely on the history of their interaction and chooses the best response to that belief (Camerer,2011).

A prominent model within this category is “fictitious play”: an algorithm that was originally proposed by Brown (1951) as a way to approximate Nash equilibrium. However, after proof emerged that not all environments converge to the Nash equilibrium using fictitious play, the approach was used as a way to explain off-equilibrium learning. After each round of the game, players current belief is simply the average of all previously observed actions distributions. Players then estimate expected payoffs for each strategy given these beliefs and select the strategies with the highest expected payoffs more frequently. This approach weighs all past actions by the opponent equally. Another variant, known as “weighted fictitious play”, weighs more recent observations higher. It was used by Cheung & Friedman (1997) to successfully capture the heterogeneity of players and explain observed play.

### **Experience weighted attraction models**

Since belief learning models contend that players beliefs and actions are largely determined by the frequency of past plays by the opponent, they implicitly assume that players do not reason about the actions they themselves previously played. What matters most for these models is what the opponent did in the past, not what the player chose. On the other hand, reinforcement learning does not explicitly consider beliefs about the opponent’s strategy and learns “implicitly” a relationship between chosen actions and subsequent rewards. These models don’t encompass any information about foregone payoffs or regrets from not playing another strategy. There are clear advantages to each approach and limitations to using either exclusively as it is not an efficient use of all the information available to players during the game.

Based on this insight, Camerer and Ho (1999) formulated a hybrid model that encompasses both belief and reinforcement learning: The Experience Weighted Attraction (EWA) model. This approach assumes that players put a varying weight, *d*, on foregone payoffs from unchosen strategies (in the extreme cases, belief learning would be *d*=1 while reinforcement learning *d*=0). Camerer et al. (2002) further developed the EWA model by adding the possibility of sophisticated learning and teaching. Sophisticated learning means that players consider the fact that their opponents are also learning, while teaching reflects the related idea that players can pick actions in order to deceive their learning opponents into erroneously inferring a particular strategy, before changing their style of play to their “real” strategy to exploit this induced inference.

### **Pattern detection:**

Other ways to model opponents have been extensively explored, mostly in the economics literature. An important category are models that are based on sequence prediction or pattern recognition strategies. These are used to identify opponent play that may be based on specific sequences that were played in the past and aims to identify and explore these patterns. They key difference with fictitious play is that instead of assuming that past actions are independent over time, and calculating simple or weighted averages of past actions, it specifically looks for autocorrelations in these past actions of the opponent and makes the implicit assumption that these will be repeated. In other words, we are looking at the frequency of sequential actions (rock, rock, paper is an example of a sequence with length 3 in the rock, paper, scissors game for instance) rather than the frequency of individual actions.

Sonsino & Sirota (2003), using the Battle-of-the-Sexes game, find that the majority of participants in their experiment converged to patterns of Nash equilibria (different NE played in subsequent rounds), consistent with players recognising patterns in the history of play. Spiliopoulos (2013) extends the weighted fictitious play approach discussed earlier by keeping track of sequences of action of length in the observed history of the game. This approach uses the observation of the past rounds to estimate the marginal probabilities of the next play. Using a repeated 2x2 game with a unique MSNE, their study shows how a learning model allowing for two period pattern detection outperforms EWA in both fitting the experimental data and in out of sample prediction of participants’ behaviour.

To sum up, we have reviewed various models that aim to explain human behaviour in repeated games. The evidence is inconclusive on which models are best able to account for the variety of observed deviations from equilibrium predictions (Camerer, 2011). One way to evaluate the validity of these various approaches is to look at whether they would be consistent with evidence of how humans transfer previously acquired learning. In the next section, we will show that there is a growing body of literature supporting the observation that players can generalise, under certain conditions, learning from previous games and transfer it to relatively novel situations. Using this insight as a yardstick for model soundness, we will compare the predictions of the various approaches outlined in this section with the empirical evidence on generalisation and transfer in strategic settings.

## **Learning transfer across games**

The rules of the game, the possible actions and potential rewards, define a particular environment through which players interact when they engage in any specific strategic game. In the real world, people might engage the same opponent in a variety of environments, so that they might be able to generalise what they have learned about their opponent in one environment to inform their decision making across other environments. It is fair to say that successfully transferring previously acquired knowledge to a new area is one of the hallmarks of human intelligence and will be a key building block of any general artificial intelligence (Taylor et al., 2008). Thus, it is important to examine the issue of whether and how human players engage in transferring the learned strategy of the opponent from one game to another.

There is a sizeable body of literature confirming that humans do transfer learning between games played sequentially (Ho et al. (1998); Stahl (2000); Rankin et al. (2000); Rick and Weber (2010)). An important transfer of learning distinction in the literature is between what is deemed superficial transfer arising from descriptive similarity of games and deep transfer that stems from a strategic similarity. Camerer & Knez (2000) call descriptive similarity between two games the correspondence in choice labels, number of action choices, identity of players, and presentation format. In contrast, strategic similarity is defined by similar payoffs from action combinations, similar equilibrium properties or corresponding socially salient characteristics of payoffs (such as need for fairness, punishment, pareto-optimality…).

In their study, Camerer and Knez have participants play two games repeatedly, for 5 rounds each, against another randomly matched participant. They used three games, the first two being descriptively and strategically similar, while the last two only had descriptive similarity. Transfer in this study was characterised by the propensity of players to cooperate in the latter game (prisoner’s dilemma) if they had learnt to choose the efficient outcome in the earlier coordination game (weak link game). Their results show that participants do transfer learning between prisoner’s dilemma and weak-link games more readily in the presence of descriptive similarity, and that the effect size is quite large, increasing cooperation in the prisoner’s dilemma game from 15-30% to 70%. However, when there was only structural similarity between the two games played, the transfer was much weaker.

Juvina (2014) looked at transfer of learning between two games (prisoner’s dilemma followed by a game of chicken), and whether superficial or deep similarity (equivalent to notions of descriptive and strategic similarity in the previous study) was driving the transfer of learning. In a clever experimental design, players were matched in groups of two and played the games sequentially, engaging with the games in one direction (PD first) or the opposite direction (chicken game first). The order of playing the game dictated whether superficial similarity would lead players to transfer learning towards an optimal or sub-optimal solution in the latter game. They find that the two types of similarity affect transfer. However, unlike the results from Camerer & Knez (2000), their analysis shows that deep similarities are the main driver of optimal strategies transfer across games. In contrast, surface similarity on its own may help or hinder optimal play depending on whether it primes players toward optimal or sub-optimal solutions in the latter game.

Similarly, exploring learning spillovers within and across structurally similar games, Mengel & Sciubba (2014) matched players in groups of four and made them play five rounds of an initial game, then ten rounds of either a strategically similar or different game against an opponent randomly matched from the same group that played the initial game. They find that the likelihood of reaching Nash equilibrium or coordination in the second game was significantly higher when players had experience with a strategically similar game compared to when they had played a structurally different game previously. In fact, playing a dissimilar game resulted in slower convergence to Nash equilibrium even when compared to a base scenario where no previous interaction had occurred (negative transfer).

While the distinctions above are interesting, most of these studies fail to offer a formal explanation of this transfer or a modelling framework that can explain the experimental observation of transfer between games and generalise it to extensive classes of games. A notable exception is the effort by Haruvy and Stahl (2012) to specify a model of learning where players learn abstract rules that they can generalise and transfer across dissimilar games, rather than action choices that can only be used within the same game. Participants played ten games, presented as 4x4 normal form (matrix payoffs). Their results suggest that not only players exhibit substantial learning transfer across games, they also show a near perfect transference of rule propensities rejecting thereby the null hypothesis of no rule learning. Other cues for rule learning were that their degree of strategic sophistication increased and that within game-learning was much faster in the latter rounds.

From the review so far, it appears that strategies that incorporate some model of the opponent can successfully fit experimental data and have more explanatory power than simple Nash equilibrium strategies or naïve reinforcement learning. However, an important caveat is that none of these experiments were designed with the primary objective to investigate opponent modelling in humans. Indeed, most studies in behavioral game theory were performed with human participants playing against other human participants. In the next section, we will briefly review studies using computer agents as opponents in games, showing that players do indeed adapt their strategies to those of the opponent and confirming computer agents as a valid experimental paradigm to study learning within games.

## **The use of computer agents to study strategic interaction**

While the approach of matching groups of human players repeatedly to play economic games has ecological validity in recreating environments where social decision making can be studied, it complicates the detection and modelling of strategies used by participants. It is harder to focus on an individual and how her strategies are changing and adapting to the opponent’s play if we cannot experimentally control the behaviour of the opponent. Therefore, studies seeking to model human strategy learning have replaced the human opponent with a computerised agent, allowing researchers to manipulate the opponent’s behaviour and observe how the human participants adapt their behaviour.

The earliest studies investigating human strategy learning when facing computer agents were conducted by Messick (1967). He made participants play a 3x3 repeated zero sum game against 3 computer agents with different strategies. He concluded that players were relying on simple heuristics such as win-stay / lose-shift in a significant proportion of trials, and that the percentage of time this simple heuristic was played was much higher when facing the fictitious play opponents. Coricelli (2005) followed similar design with human subjects playing a 4x4 game with a unique MSNE. This study came to similar a conclusion, namely that humans choose different strategies against the opponents, better responding to the expectation of over-alternation by increasing the autocorrelation in their choices.

Shachat & Swarthout (2004) made human participants face computer opponents playing various mixed strategies in a zero-sum asymmetric matching pennies game. They found that the players changed their strategies towards exploiting the deviations from the MSNE, and that this exploitation was very likely if the deviation from the MSNE play was at least 15%. They also noted significant heterogeneity in the extent of exploitation of the non-optimal play between subjects.

Theses empirical precedents show that human players do learn to exploit their computer opponents and adapt their strategies successfully to the agent they are facing, even when the agent was using relatively sophisticated strategies to potentially exploit the human opponent. It therefore appears that using computer agents as an experimental paradigm is a valid way to study strategy learning in humans, with the advantage of offering significant experimental control since the agent’s playing rules are known to the experimenter. The above also implies that successfully manipulating the agent’s strategy by changing the degree of sophistication and the nature of rules the computer uses (Nash, iterated dominance or level-k) would allow the exploration of participants ability to detect, exploit and eventually generalise various rules and strategies.

# **The present study**

If participants engage in a form of opponent modelling that goes beyond solely learning the contingencies of their opponents’ play, then this type of learning should allow participants to generalise their knowledge about the opponent playing rules, which would manifest as a learning transfer between games.

The review of the literature shows that transfer of learning studies tend to use human participants playing against other human participants. This paradigm vastly diminishes the experimental control of the studies, since human opponents are also capable of learning and will dynamically change their strategies in subsequent interactions, blurring any evidence of learning transfer from experimental data. Therefore, in order to measure learning transfer, and manipulate the degree of sophistication of the opponent, the use of computer agents is needed.

On the other hand, we saw that studies that looked at how players adapt to their opponent strategies when facing computer agents have mostly looked at the ability of players to detect and exploit *action-based* learning rules. They either used deterministic strategies that did not consider the human participant’s play, as in the case of pre-determined mixed strategy algorithms, or strategies that only depended on the player’s frequency of past plays, not considering the computer’s own past plays, or the degree of strategic reasoning of the human player about the computer, such as fictitious play ( based on past actions of the opponent) and reinforcement learning (based on the rewards of past actions of the player).

Thus, none of the studies we reviewed in the literature of human vs computer players used a reasoning-based computer agent, such as a level-k strategy, or a cognitive hierarchy model. Haruvy and Stahl (2012) show that these rules can be learned by humans and are generalised and transferred to dissimilar games. In this study, we therefore propose to fill this gap in the literature, by making human participants play against computer opponents that use reasoning-based playing strategies, consistent with various degrees of level-k theory. While the computer opponent will remain the same for each participant, the study will consist of three games played sequentially. The games will have various degrees of similarity which will allow us to explore the effects of descriptive vs strategic similarity on the transfer of learning in these games.

The above experimental paradigm serves to explore various questions related to the ability of human participants to learn agent’s strategy through repeated interaction, whether this learning is transferred and to what extent the transfer is modulated by the degree of sophistication of the opponent as well as the similarity between games. More specifically, we will aim to test three research hypotheses as articulated below:

### **Between Game Transfer of learning**

The main research question we are exploring here is whether playing an opponent repeatedly across different games leads participants to generalise their learning about the opponent strategy from one game to another. If participants are indeed generalising their learning, then scores on games 2 and 3 should be higher than score on game 1, even before learning has occurred in the former. There are multiple ways we will explore this question. More specifically, we will compare player’s average scores on the first half of each game. Next, as a robustness check, we will compare score on the five earliest rounds (2 to 6, excluding the first round where the computer agent chooses its action randomly since it doesn’t have access to the human’s last play).

### 

### **Effect of agent’s degree of sophistication on learning transfer**

Another important effect to look into is that of the level of strategic sophistication of the computer agent. We believe that the level-2 agent will be harder to play against than the level-1 agent, as it follows a more sophisticated strategy. Therefore, we will compare the overall average scores for each game of players facing level-1 agents vs those facing level-2 agents. We will also explore whether within-game learning and transfer of learning are affected by the degree of sophistication of the agent.

### **Effect of game similarity on learning transfer**

We expect the degree of similarity between the games to positively impact transfer, since it should be easier to transfer knowledge between structurally and descriptively similar games than between only structurally similar games This will tie into the literature on implicit vs explicit learning. Explicit learning of the strategy will be reflected in the possibility of transferring knowledge to different games and should be reflected in the feedback measure sf confidence and difficulty and more importantly in the written comments of participants at the end of the experiment. We should see a significant performance drop in the early rounds of the descriptively dissimilar game for players who associatively learned the winning patterns as this type of learning is considered “implicit” and therefore harder to transfer.

# **Method**

## **Participants and design**

A total of 52 (28 female, 24 male) participants were recruited on the Prolific Academic platform to take part in the experiment.The mean age of participants was 31.2 years**,** and it took them on average around 18 minutes to complete the experiment. Participants were paid a fixed fee of £2.5 plus a bonus dependent on their performance which averaged £1.06. The study used a 2 (computer opponent levels k1 and k2) by 3 (games of rock-paper-scissors, fire-water-grass and the mod game) design, with repeated measures on the second factor.

## **Task**

Participants in the experiment each played 3 games sequentially against the same computer opponent. The computer opponent either used a k1 or k2 strategy. The three games were rock-paper-scissors, fire-water-grass, and the mod game.

### **Rock paper scissors**

Rock-paper-scissors is a very popular children’s game, especially as a random way to make a choice between two people or as a way to settle a dispute. A typical Rock-paper-scissors game (hereafter RPS) is a 3x3 zero sum game, with a cyclical hierarchy between possible actions: rock blunts scissors, paper wraps rock, and scissors cut paper. If one player choses an action which dominates their opponent’s action, the player wins (receives a reward of 1) and the other player loses (receives a reward of -1). Otherwise it is a draw and both players receive a reward of 0. It has a unique MSNE consisting of randomly playing one of the three options each time. RPS has generated a fairly important literature in experimental economics where it was used to study strategic interaction in repeated games (Batzilis et al., 2014) and evolutionary learning processes (Hoffman et al. 2015).

### **Fire-Water-Grass**

This is a variant of Rock-Paper Scissors that has the exact same strategic structure, with three choices having a cyclical dependency: Fire burns grass, water extinguishes fire, and grass absorbs water. This game is the same as Rock-paper-scissors in all but action labels. We are interested in exploring whether learning is transferred in a fundamentally similar game where the only difference is in the description of the choice actions. If one player choses an action which dominates their opponent’s action, the player wins (receives a reward of 1) and the other player loses (receives a reward of -1). Otherwise it is a draw and both players receive a reward of 0.

### **The Mod Game**

First used by Frey & Goldstone (2013) as an experimental paradigm to test cyclical behaviour in strategic interactions, the mod game is a generalization of rock-paper-scissors. In the variant we use, 2 participants concurrently pick a number between 1 and 5. To win in this game, a participant needs to pick a number exactly 1 higher than the number chosen by the opponent. For example, if a participant thinks their opponent will pick 3, they ought to choose 4 to win the round. To make the strategies cyclical as in RPS, the game stipulates that the lowest number (1) beats the highest number (5), so if the participant thinks the opponent will play 5, then the winning choice is to pick 1. This game has a structure similar to RPS in which every action is dominated by exactly one choice. All other possible combination of choices that are not consecutive are considered ties. A win would add 1 point to the score of the player, while a loss deduces one point and a tie does not affect the score. Similar to RPS, the MSNE is to play each action with equal probability in a random way. From now on, this game will be referred to as NUMBERS in the analysis.

## **Experiment Procedure**

After seeing a page with general information about the study and signing a consent form, participants were given instructions about the experiment. The interface for the experiment was coded as a webpage using the Node.js platform for the back-end allowing for better real-time interaction between the server and the players, as well as the possibility of multiple players connecting to the server at the same time (not to play against each other, but to play against their randomly matched computer opponents).

Participants were informed they would play three different games against the same computer opponent, namely: Rock-paper-scissors, fire-water-grass and the mod game. Each participant plays all three games consecutively and in the same order described above. Participants were told that the opponent cannot cheat and will choose its actions simultaneously without knowledge of the participant’s choice. A total of 50 rounds of each game was played with the player’s score displayed at the end of each game. The score was calculated as the number of wins minus the number of losses. Ties did not affect the score. In order to incentivise the participants to maximise the number of wins against the opponents, players were paid a bonus at the end of the experiment that was proportional to their final score. At the end of the experiment, the overall score of the players was translated into the bonus by making each point worth £0.02. This bonus is significant as players could increase the total payoff from the experiment by up to 60% assuming they’d won all rounds against the computer opponent.

Players were matched randomly against one of two computer agents that played according to a pre-programmed strategy in 90% of rounds and chose an action randomly on 10% of rounds. The strategies were respectively named after the level of iterated reasoning programmed into the agent. We called “level-1 agent” a computer that believes its human opponent is level-0. In this case, a level 0 means that the human is naïve and simply repeats last round’s choice. Therefore, a level-1 agent will play (90% of the time) a strategy that would have beaten the choice of the human player in the past round. Meaning, if the opponent had played Rock in the past round, 90% of the time the level-1 agent will play paper and 10% of the time it would choose its response randomly out of the three possible options.

By contrast, a level-2 computer agent assumes that the human player is a level-1 opponent, as explained above. Therefore, a level-2 computer agent believes the human player will choose in the next round the action that would have beaten the computer’s choice in the last round and will best respond to that. For instance, assuming the level-2 computer agent had played rock in the most recent round, it will “think” that the human opponent (as a level-1) will try to beat that by playing paper this round, and it will best respond to that by playing scissors.

After detailed instructions for each game, participants underwent a small test (3 questions) to ascertain they have understood the rules of the games and the duration of the play. These questions probe whether they have understood and memorised the hierarchy between the options, such as asking which action would beat “rock” in the first game.

In each round of a game, participants were instructed to choose one the “weapons” by clicking on the appropriate icon. In the meantime, the computer agent chose its “weapon” based on the play in the previous round (in the first round, it chose randomly from the available actions). The computer’s choice was made after a random time delay between 0.5 and 3 seconds, to simulate how a human opponent would behave in these circumstances. After each round, the participants were given feedback: they were reminded of their own choice, informed of the opponent’s choice, as well as the outcome of the round (win, tie or loss). The number of wins, losses and ties were displayed at the top of the screen for each game, and this scoreboard was reinitialised to zero at the onset of a new game. An example of the interface for the rock-paper-scissors game is provided in Figure 2.

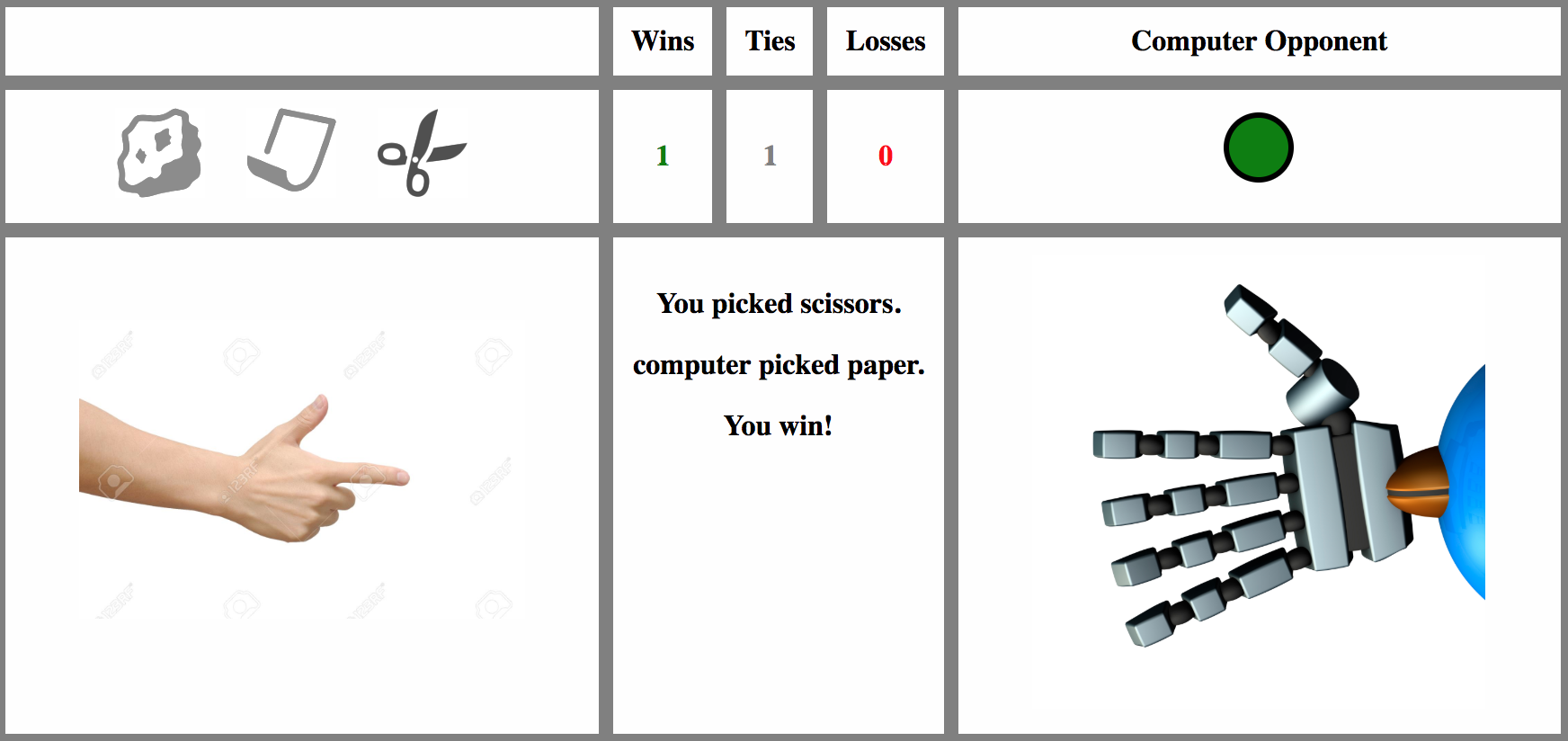


Figure 2: Screenshot of the feedback at the end of a round of Rock-Paper-Scissors

At the end of the third game, participants were given information about their overall score and asked 3 questions about the opponent’s strategy. The first question asked, on a scale from 1 (very easy) to 100 (very hard), how difficult was it for them to learn the opponent’s strategy and beat it. The second question asked them to rate, on a scale of 1 (not at all confident) to 100 (very confident) their confidence in the assessment that they had learned what the opponent strategy is. The third question asked participants to briefly describe the strategy they though the opponent was playing, as well as any other feedback on the experiment.

# **Results**

All subsequent analyses use as a dependent variable the participants average score per round, calculated by counting the number of wins across games minus the number of losses, ignoring ties, and dividing the results by the number of rounds. Average scores will therefore lie between -1 (player loosing systematically all his rounds) and +1 (player winning systematically all his rounds).

## **Performance**

An initial assessment of the distribution of scores per game (see Figure 3) shows a high variability of scores, with a few negative average scores and some as high as 0.9.

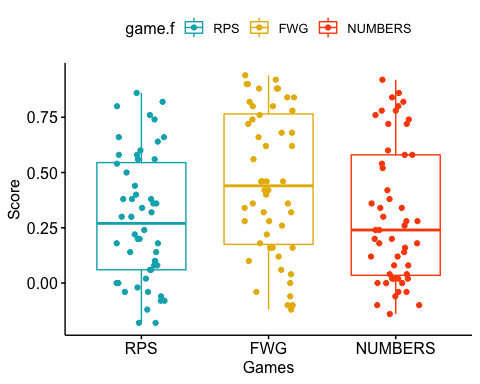


Figure 3: Distribution of scores by game with the mean score and standard error boxes

Looking at the various scores, the RPS game had the lowest average score across participants (*M = 0.289, SD = 0.348*) followed by NUMBERS (*M = 0.31, SD = 0.347*) and finally the FWG game had the highest average score (*M = 0.454, SD = 0.354*).

A player choosing actions randomly against the computer agent should be expected to have an equal number of wins, losses and ties, meaning an average score of 0. We can test the average scores for each game against a hypothesised value of 0 for a random player using parametric one sample t-tests. All scores are significantly higher than 0 and random play is rejected (*RPS: t(51) = 7.26, p-value = 2.1e-08* ; *FWG: t(51) = 10.04* , *p-value = 1.10e-13 ; NUMBERS: t(51) = 7.17, p-value = 2.87e-09).*

## **Within and between-game learning**

To analyse within and between game learning, we used a 2 (condition: level-1, level-2) by 3 (game: RPS, FWG, NUMBERS) by 2 (block: first half, second half) repeated measures ANOVA with the first factor varying between participants. As the number of participants facing level-2 agents was slightly higher than those facing level-1 agents (28 vs 24), the ANOVA is unbalanced. We therefore used the “afex” (Singmann & al., 2016) package in R, as it allows us to handle unbalanced designs and produces type III sum of squares by default. The full results of this analysis are provided in Table 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Effect | Df1 | Df2 | MSE | F | GES | p-value |
| Condition | 1 | 50 | 0.34 | 5.44 \* | .05 | .02 |
| Game | 1.87 | 93.62 | 0.11 | 8.54 \*\*\* | .05 | .0005 |
| Condition\*Game | 1.87 | 93.62 | 0.11 | 0.75 | .005 | .47 |
| Block | 1 | 50 | 0.05 | 22.5 \*\*\* | .03 | <.0001 |
| Condition\*Block | 1 | 50 | 0.05 | 0.16 | .0002 | .69 |
| Game\*Block | 2.00 | 99.90 | 0.04 | 6.93 \*\* | .02 | .002 |
| Condition\*Game\*Block | 2.00 | 99.90 | 0.04 | 3.89 \* | .010 | .02 |

*Table 1: Output of 2(condition) by 3(game) by 2(block) Repeated Measures ANOVA*

There was a main effect of Game ( *F(2,100) = 8.54, ges[[1]](#footnote-1) = 0.05, p = .0005*), showing that average scores varied significantly over the games. Post-hoc pairwise comparisons showed that performance in the FWG game was significantly higher than in the RPS game ( t*(100) =3.78, p = 0.0008* ), and the performance in NUMBERS was significantly lower than FWG game ( *t(100) = -3.32 , p = 0.0024*). The score in RPS was not significantly different from the score in NUMBERS however ( *t(100) = 0.45 , p = 0.65*).

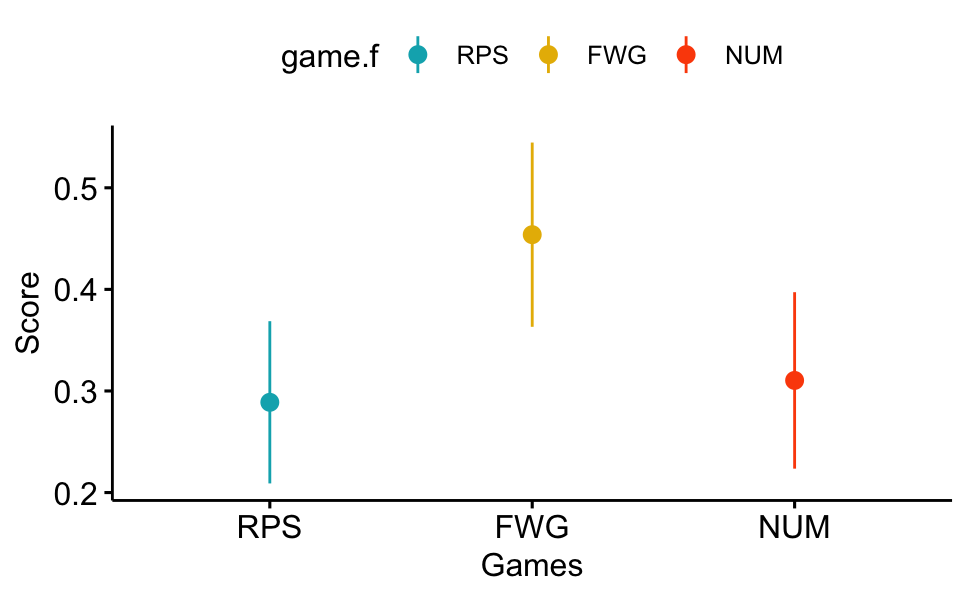


Figure 4: Average scores per game with 95% confidence intervals

The main effect of Block ( *F(1,50) = 22.51 , p < .001, ges = 0.03*) shows that the average score in the first half of games (*M = 0.29*) was significantly lower than in the second half of the games played (*M = 0.40*), which translates to within-game learning : as players get more interaction with the agent, they learn how to win more often.

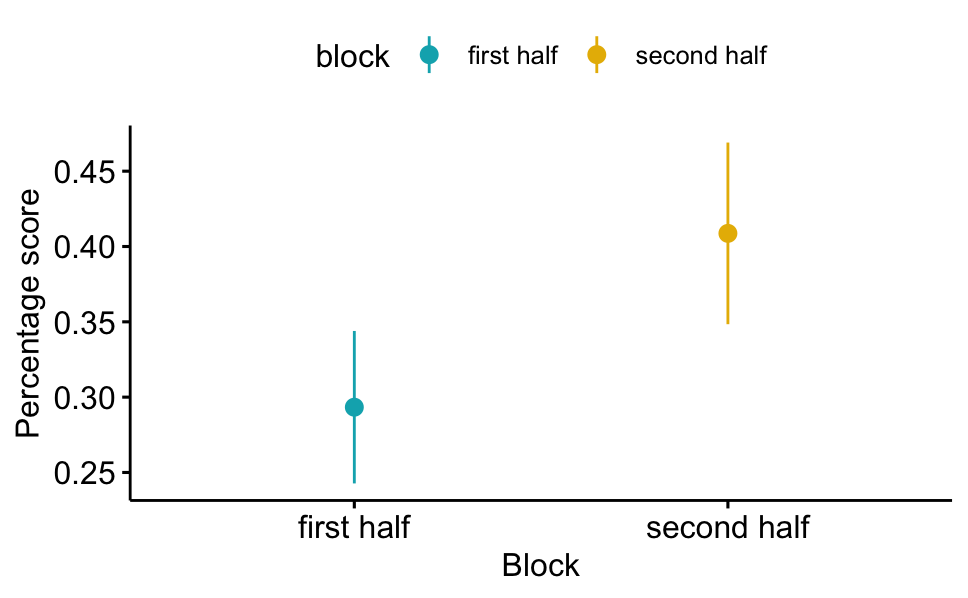


Figure 5: Average scores for each block with 95% confidence intervals

The main effect of Condition (*F(1,50) = 5.44, p = .024, ges = 0.05*) indicates that scores were higher against the level-1 player (*M = 0.43*) than against the level-2 player (*M = 0.27*). This means that it was harder for participants, on average, to learn the strategy of the more sophisticated opponent (level-2) compared to that of the comparatively less sophisticated agent (level-1).

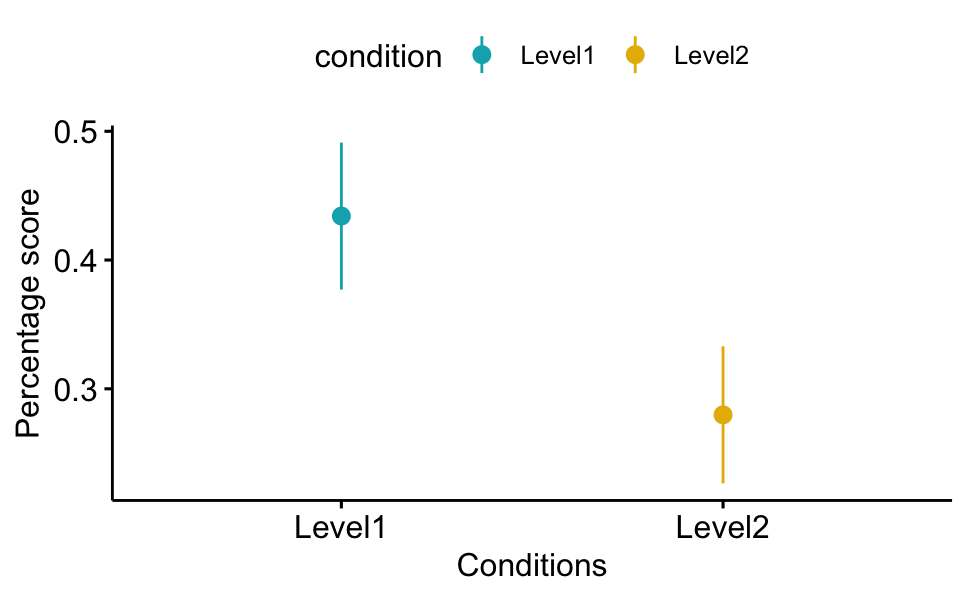


Figure 6: Average score per condition with 95% confidence intervals

Finally, the analysis showed a significant interaction effects of block by game ( *F(2,100) = 6.92 , p = .002, ges = 0.02*), indicating that within-game learning differed between the games, as well as a three-way interaction between condition, game, and block *( F(2,100) = 3.88 , p = .023, ges = 0.01*). Below, we will look deeper into these interactions through post-hoc tests involving pairwise comparisons.

## **Differences in within-Game learning between games**

Figure 7 in an interaction plot showing the marginal score means by the within-subject factors: block and game. As can be seen, second-half scores are always higher than first-half scores, indicating that players having more success against an opponent the more experience they have playing against them in a game (this is consistent with the main effect of block reported above). However, the graph indicates that such within-game learning is most marked for the first game (RPS), and less so for the two consecutive games (FWG and NUMBERS).

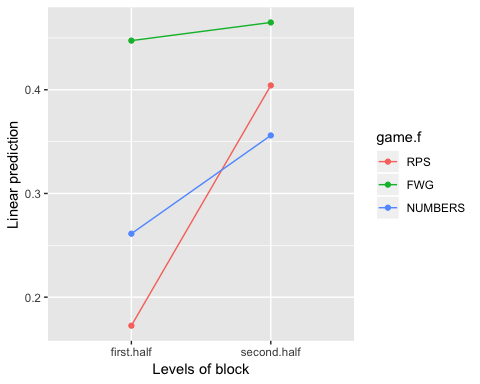


Figure 7: Interaction plot for average scores per game and block

To test for the statistical significance of this pattern, we performed post-hoc pairwise comparisons through the lsmeans package (Lenth, 2015) in R and used the Holm-Bonferroni correction to constrain the family wise error rate to 5%.

For the initial (RPS) game and last (NUMBERS) games, the score on the second half of the game is significantly higher than on the first half (*RPS: t(149.97) = 5.59, p < .0001; NUMBERS: t(149.97) = 2.28, p = 0.047*). For the second (FWG) game, the scores between the two halves are not significantly different from each other ( *t(149.97) = 0.42 , p = .67* ). Therefore, there is evidence for within game learning in RPS and NUMBERS, but not in FWG. The reason within game learning is low in the latter two games might be due to high scores in the early rounds, which would reflect learning transfer from RPS to these two games. In the next section, we will focus on how to measure and test for the existence of transfer of learning.

## **Transfer of learning between games**

### **Measuring transfer by comparing first-half block scores**

With regards to across games transfer, one way to measure it is to compare first-half scores in each game, which reflect performance that is relatively untainted by within-game learning. The first-half score in the FWG game is significantly higher than first-half RPS score *( t(172.82) = 5.18, p < 0.0001*) and only marginally higher than second-half scores in the RPS game ( *t(173.47) = 0.81, p =1.00* ). This indicates a successful transfer of learning from the RPS game to the more similar game (FWG) since participants early performance in a game that was new to them is significantly higher than initial performances in the precedent game RPS and no less than their performance at the end of RPS, after significant interaction with the opponent.

Looking at NUMBERS for evidence of transfer, we see that the first half score of NUMBERS is significantly lower than that of FWG ( *t(172.82) = -3.51, p = 0.006*), and in line with early RPS scores (*t(172.82) = 1.67 , p = 0.52*). This approach, looking at block scores, suggests that players, on average, fail to transfer the learning from the first two games to the more dissimilar game (NUMBERS) even after having had a longer interaction with the opponent (100 rounds of playing two very similar games).

### **Measuring transfer by performance in initial rounds of the games**

One issue with the way we measured transfer, is that we compared scores on the initial 25 rounds (half the total number of rounds per game) for each participant. It is possible that 25 rounds give plenty of opportunity for within game learning and therefore are not as an accurate measure of transfer between games.

Therefore, as a robustness check, we focus on participants’ scores in the 5 earliest rounds only, excluding the very first round for which the agent is programmed to play randomly as it has no data on prior rounds on which to build its response. A group of players with no experience of the game are expected to have scores not significantly different from 0. Any significantly positive group average scores would therefore reflect prior learning from past experiences.

In figure 9, the average score across participants by game for rounds 2-6 are plotted. Scores are also averaged across levels of condition.

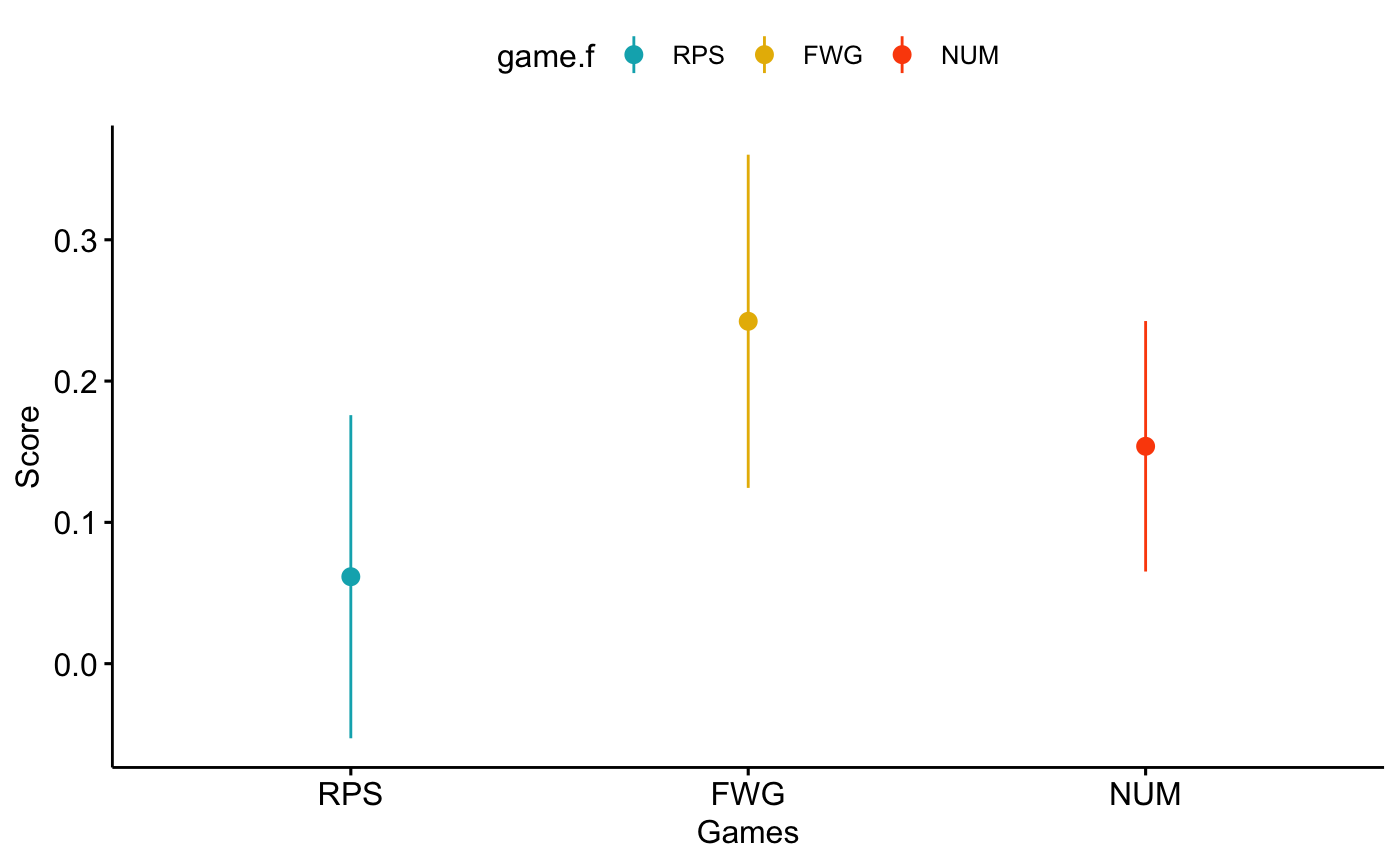


Figure 8: Average scores on rounds 2 to 6 across conditions for each of the 3 games

The RPS game had the lowest average score across participants (*M = 0.06, SD = 0.41*) followed by the numbers game (*M = 0.15, SD = 0.31*) and finally the FWG game had the highest average score across participants (*M = 0.24, SD = 0.42*). We test the average scores for each game against a hypothesised value of 0 for a non-experienced player using parametric one sample t-tests. As expected for the initial RPS game, the average score is not significantly different from 0 ( *t(148.85) = 1.04 , p = 0.89*). In FWG, the score is significantly higher than 0 ( t*(148.85) = 4.58 , p < 0.0001*). Unexpectedly, this is also the case for NUMBERS ( *t(148.85) = 3.00, p = 0.0092*).

Our robustness check results using the earliest rounds confirm that transfer of learning is indeed occurring between games, with a strong effect between the very similar FWG and RPS games. For the dissimilar game, NUMBERS, this approach offers evidence for transfer of learning, albeit a weaker one given the lower score, which is at odds with the performance analysis using blocks of games. The reason for the discrepancy is likely the high score on the first block of RPS (*M = 0.17, SD =0.25*) which is also statistically significant *( t(161.75) = 3.70, p = 0.0003*). This score artificially inflates the “control” condition of RPS and therefore underestimates transfer.

We therefore conclude that our results from the early rounds analysis are likely to be a more representation of whether learning transfer occurs. Thus, we will focus on early rounds scores (2-6) to measure transfer as comparing first-half block scores might be tainted by within game learning.

## **Moderation of within and between-game learning by type of opponent:**

In the previous analysis, we had averaged across conditions, pooling together playing who faced both a level-1 and level-2 agents. In this section, we are interested in separating the evidence for within and between game transfer by type of opponent faced.

For the within-game learning, the significant three-way interaction between condition, game, and block indicates a difference in learning by game for the two types of opponent faced. Figure 9 shows marginal means of players scores for first and second block for each game, distinguishing players facing level-1 computer agents (left panel) from those facing level-2 agents (right panel)[[2]](#footnote-2).

### **Within-game learning by type of opponent:**

For level-1 facing players, most of the learning happens in the first game (RPS) where scores on the second half are significantly higher than those on the first half ( *t(149.97) = 4.82, p < 0.0001*). First half and second half scores are not statistically significantly different for FWG ( *t(149.97) = 0.41, p = 1.00)* and NUMBERS *( t(149.97) = -0.05 , p = 1.00*) showing that no significant incremental within-game learning occurs in these two games. With regards to players facing level-2 agents, although the performance is lower, the pattern of within-game learning within RPS is still significant ( *t(173) = 3.02, p = 0.01* ).

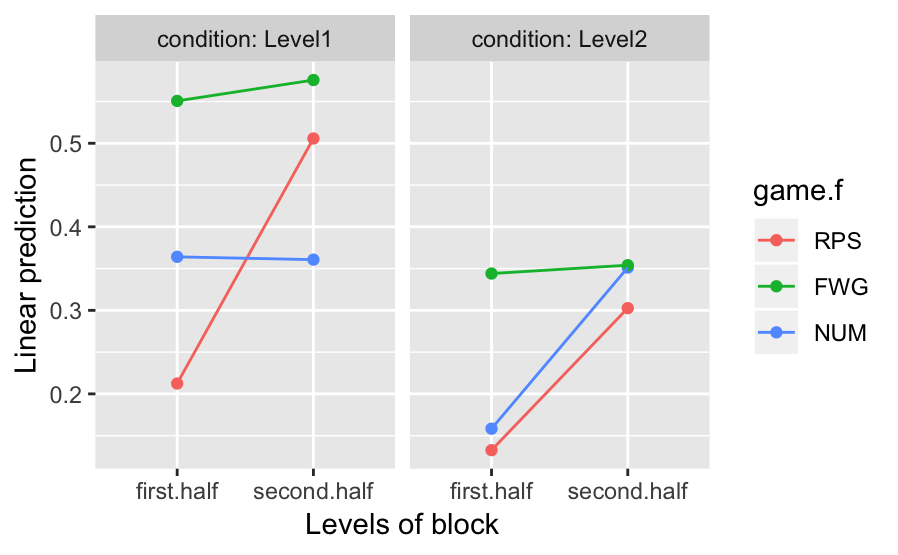


Figure 9: Marginal means of players scores by block and type of opponent

### **Moderation of between-game learning by type of opponent and game similarity**

We will use average scores from rounds 2-6 as a proxy for transfer of learning by type of opponent between RPS and both the similar (FWG) and dissimilar game (NUMBERS). Figure 10 shows the marginal mean scores for each game for both level-1 and level-2 facing players. Graphically we can see that the pattern is dissimilar between level-1 and level-2 players, and we suspect transfer to be positively associated with similarity and negatively with degree of sophistication of the agent. To test these hypotheses, we run statistical tests on scores by game and opponent against the null hypothesis of 0 (no transfer)

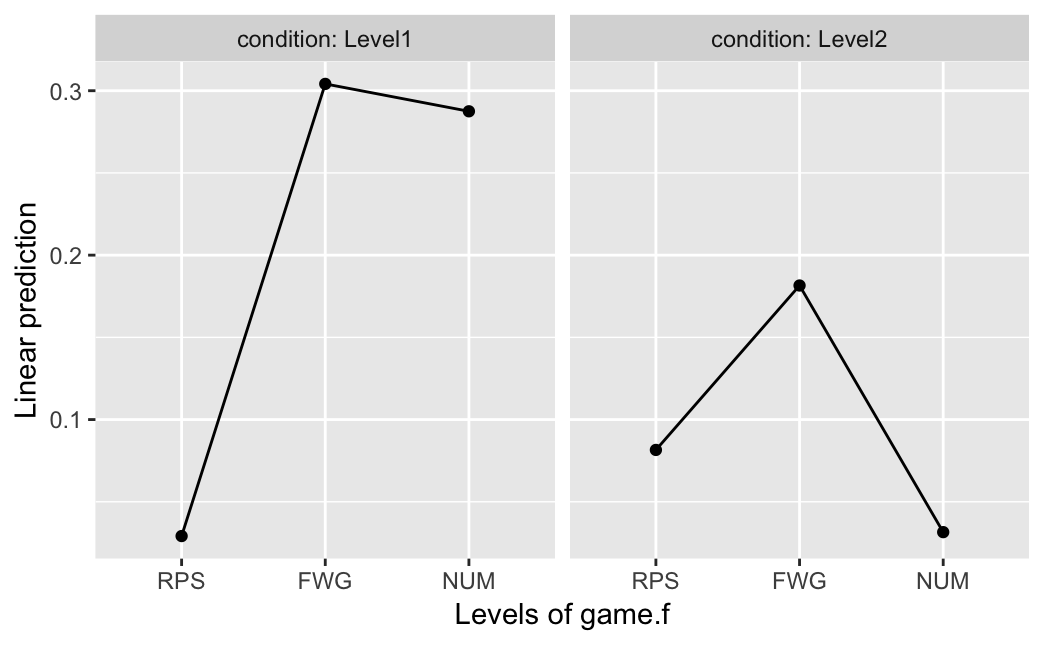


Figure 10: Marginal means of players scores by game and type of opponent

Table 2 shows the scores on rounds 2-6 for both type of opponents in each game. We also report the t-stats and p-values to test the null hypothesis that scores on these early rounds are equal to 0. Rejecting the null hypothesis means that prior learning has been used to beat the opponent in these early rounds. For level-1 facing players, there is evidence of learning transfer to both FWG and NUMBERS. For level-2 facing players, there is evidence for transfer to the similar game FWG (albeit scores are lower than for level-1 player) but not to the dissimilar game of NUMBERS.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Game | Condition | Score | SE | df | lower.CL | upper.CL | t-ratio | pvalue |
| RPS | Level-1 | 0.029 | 0.077 | 150 | -0.123 | 0.18 | 0.38 | 0.70480 |
| FWG | Level-1 | 0.304 | 0.077 | 150 | 0.152 | 0.46 | 3.96\*\*\* | 0.00012 |
| NUM | Level-1 | 0.287 | 0.077 | 150 | 0.136 | 0.44 | 3.74\*\*\* | 0.00026 |
| RPS | Level-2 | 0.082 | 0.073 | 148 | -0.063 | 0.23 | 1.11 | 0.26713 |
| FWG | Level-2 | 0.182 | 0.073 | 148 | 0.037 | 0.33 | 2.48\* | 0.01426 |
| NUM | Level-2 | 0.032 | 0.073 | 148 | -0.113 | 0.18 | 0.43 | 0.66724 |

*Table 2: Average-score means by game and type of opponent for rounds 2-6 with 95% CI.*

Our results when averaging across conditions (previous section) showed that there was indeed evidence for transfer to the more dissimilar game (NUMBERS). We can see from the table above that this transfer is exclusively driven by level-1 facing players, as scores of level-2 facing players are close to nil. Therefore, both groups of condition can generalise to the similar game, but only those facing the less sophisticated opponent are able to generalise to the less similar game.

## **Computational modelling of Human strategies**

To gain more insight into how participants played the games against the computer opponents, we compared multiple models of strategies the players may have been using in their decision making. The first is a model that assumes play is random, and each potential action is chosen with equal probability. Note that this corresponds to the Nash equilibrium strategy. Next, since our computer agents has a predetermined level-k strategy, we tested the explanatory power of various level-k models, with k ranging from 1 to 3. Finally, we also included in our model comparison a heuristic win-stay lose-shift (WS/LS) model, that has proven quite effective in explaining behaviour in repeated games (Worthy & Maddox, 2012; Steyvers et al. 2009). This model assumes that players repeat their previous actions if they win the last round and shift their actions towards what would have beaten the last round choice if they’d just lost. With regards to ties, assuming that players would also shift their actions in the same direction as losses results in a model that gives the exact same predictions as a level-1 model in this context of 3x3 games, and therefore, we choose to treat the player’s reaction to ties similarly as when a win happens, meaning players repeating their last action.

As participants may not employ a strategy consistently, we assume, for each model, that the strategy is followed with a probability . Otherwise (with probability 1-) the player choses randomly between all options. For each participant and model, this parameter is estimated by maximum likelihood (i.e., the parameter is estimated to maximise the probability of the observed responses according to the model). To allow that consistency differs between the games (e.g., as a result of learning) we estimated, in addition to a version which assumes is constant over games, also a version that assumed it may differ between the games. To compare the models, we used the BIC, a model selection criterion that penalizes the fit of a model by the number of parameters.

The results of the computational models confirm that while there is some heterogeneity in the strategies people chose, a significant number of players best respond to their opponent by adopting a level k+1 strategy when facing a level-k agent (k =1,2). More specifically, from Table 3 we can see that when facing level-1 agents, 75% (18/24) of players actions were best explained by a level-2 model. Likewise, the equivalent number of players whose actions are best fit by a level-3 strategy when facing level-2 agents is 53% (15/28).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Opponent | Level-2 | Level-3 | Random | WS/LS |
| Level-1 | 18 | 1 | 5 | 0 |
| Level-2 | 4 | 15 | 3 | 6 |

*Table 3: Distribution of best fitting player model by type of opponent*

In both cases, the models allowing the parameter to vary for each game were more successful in explaining the player’s behaviour. Suggesting that the likelihood that players choose actions consistent with the identified strategies is not constant over time but evolves dynamically.

To test that there are significant differences between the thetas estimated for each game, we run a repeated measures ANOVA with the estimated thetas as a dependent variable and game as a 3-level factor (RPS, FWG and NUMBERS). We run two ANOVAS, one for each type of opponent, using the thetas estimated for the best predictive model (level-2 model for level-1 agents and level-3 model for level-2 agents). Figure 13 shows the estimated average values of theta for each game along with 95% confidence intervals.

For level-1 facing players, there is a significant effect of the game factor *( F(1.98, 101.06) = 7.33, p = 0.001, ges = 0.05*) showing that thetas indeed vary between games. The average estimated theta was highest for FWG (*M = 0.48, SD = 0.36*) followed by NUMBERS (M = *0.30, SD = 0.34*) then RPS ( *M = 0.27 , SD = 0.29*).

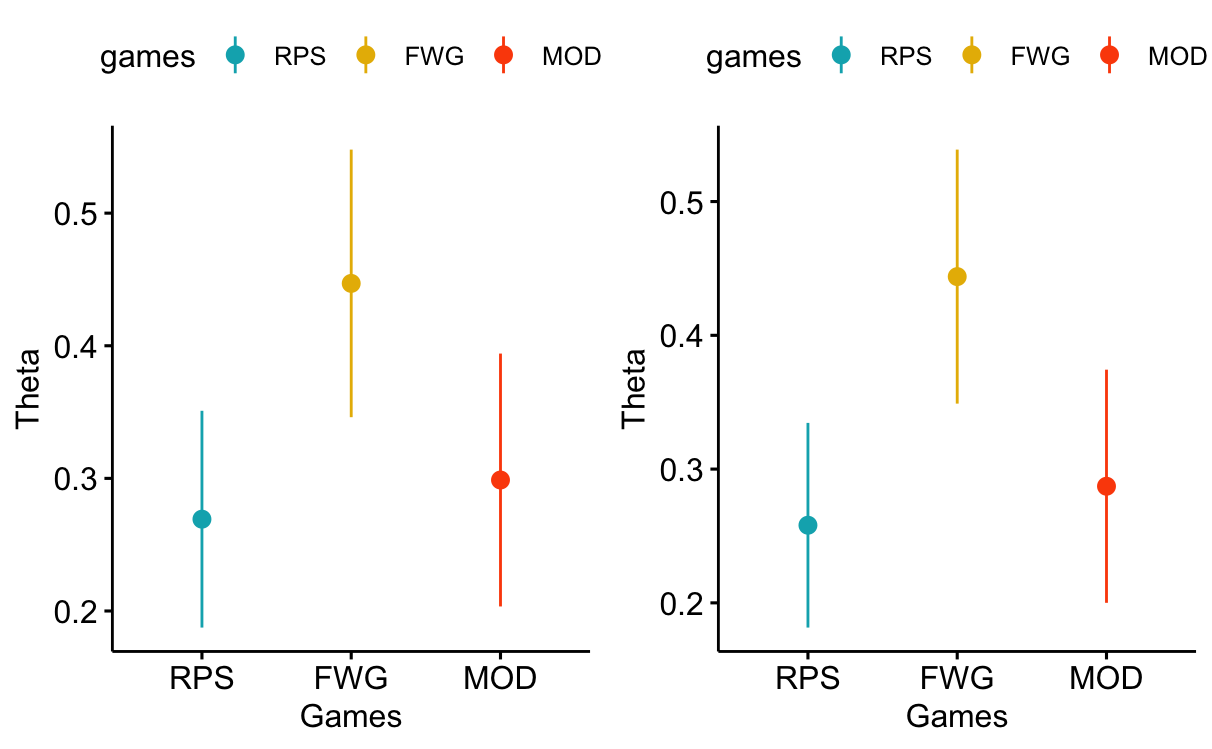


Figure 11: Mean estimated theta by game for levels 1 & 2 facing participants with 95% C.I

For level-2 facing players, the game factor is also significant ( *F(1.98, 101.06) = 8.99, p = 0.0003, ges = 0.07*). The average estimated theta was highest for FWG (*M = 0.44 , SD = 0.34*) followed by NUMBERS (*M = 0.28 , SD = 0.31*) then RPS ( *M = 0.26 , SD = 0.27*).

Although the models do not explicitly analyse how participants might learn about their opponent and change strategy accordingly, we can get some insight into such learning by plotting, in figure 12, the likelihood (the probability of the chosen action according to each model) as a function of game and round.

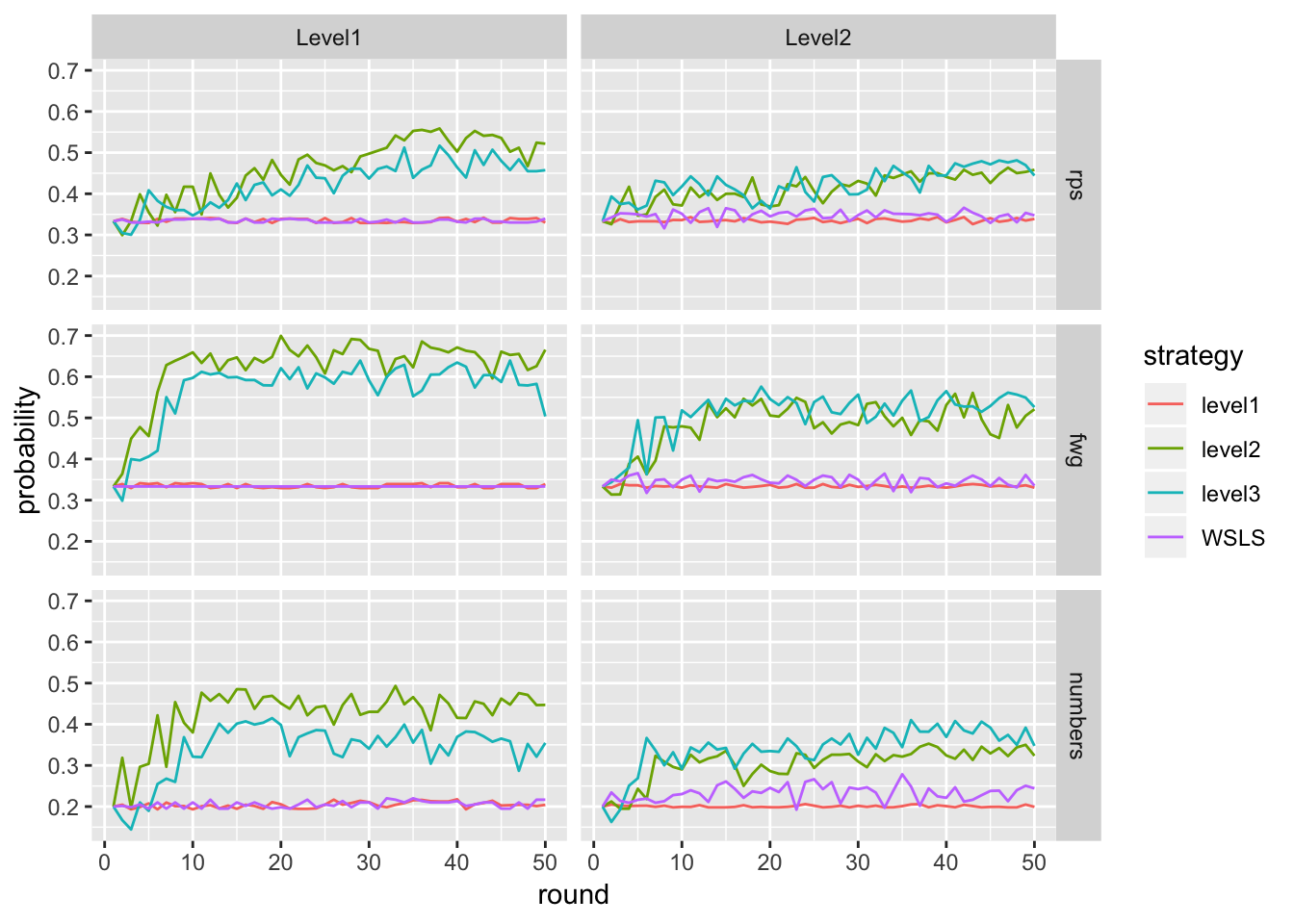


Figure 12: Probability of player’s actions being consistent with the models per game & round.

Initially, the models all have equally low probability, but as play continues, human participants played actions predicted by a level k+1 strategy more frequently and consistently over time. The slope of increase in the average probability conveys important information about the speed of convergence towards the level-k strategy. We can see that the likelihood of both level-2 and level-3 increases steadily but slowly as rounds go by in the RPS game. However, the likelihood of these two models increases rapidly in the FWG game, indicating individuals are exploring less and are playing level-k+1 consistent choices more rapidly. Furthermore, for level-1 facing players, by the end of play in each round, it seems clear that the level-2 playing rule is best at explaining the observed pattern of play in each of the 3 games. However, this is not the case for level-2 facing agents for which it is difficult to distinguish whether level-2 or level-3 strategy was best at explaining the observed actions as the two curves are very close to each other.

The results of the computational modelling are consistent with a qualitative analysis of the feedback received from participants at the end of the experiment. Some examples of comments are attached in the appendix. We could distinguish multiple themes from the written comments. One group clearly wasn’t able to learn the opponent strategy and admitted that they were not sure what the opponent was doing. A second group tried to articulate the opponent strategy, but their explanations were either not clear or not consistent with what the agent was programmed to do. A small number of players however was able to describe the strategy correctly, reflecting their ability to learn how the agent was playing and best respond to that.

# **Discussion**

In this study, we investigated human learning transfer across games by making human participants play against computer agents with rule-based level-k strategies. We were interested in uncovering evidence for transfer and exploring whether it is modulated by the degree of similarity between games and the sophistication of the agent.

The results of our online experiment show that the majority of participants learn to adapt to the opponent strategy over multiple interactions and generalise this learning to very similar games. We found that using results on very early rounds is a better proxy for measuring transfer as it is not tainted by any within game learning. Using this approach, we showed that transfer to the more dissimilar game was modulated by the degree of sophistication of the agent, with evidence for transfer when players face the less sophisticated agent but not the more sophisticated one. The question is therefore: what did the players learn exactly about the opponent strategy? What form did this learning take?

One possibility is that participants learn patterns of play that depend on their most recent actions or those of the opponent. For instance, such a mapping might sound like: “When I typically played scissors in the last round, I needed to play rock in this round to win”. Cognitively, this is not taxing as there are only 3 transitions to memorise and could explain the within-game learning in RPS. This type of associative learning would be consistent with a reinforcement-learning type model of behaviour (Erev & Roth, 1998) in which participants are more likely to repeat plays that yielded positive rewards in the past. It can lead to the appearance of learning as participants uncover winning plays without learning the rules according to which the opponent is plying.

We believe this explanation is unlikely for multiple reasons. First, it would not be consistent with evidence for transfer of learning through early performance in FWG. A model of naïve reinforcement-learning is not capable of generalisation of action choices from one game to the other. Indeed, the associations would need to be re-discovered for the new “weapons” or a mapping established between weapons in the first and second games. This would not be achievable with any degree of certainty within the first 5 rounds and cannot explain the very early above chance performance in FWG.

There are other possible explanations for learning the opponent’s strategy, such as the ability to memorise the best response to all possible states in the previous round. This would be akin to learning the complete structure of the environment and would be similar to having a model-based reinforcement learning approach to solving the game (Kurayev & Sutton, 1997). This is not a satisfactory model however: Besides the fact that holding the whole structure of the environment in working memory is quite difficult and would be very taxing cognitively, this approach would not circumvent the issue of explaining the significantly positive early scores on rounds 2 to 6 since it is even less unlikely that players would learn the new structure of the environment over a few rounds.

A third possibility is that players learn very simple spatial heuristics after repeated play, and realise they have a winning pattern. Spatial heuristics are simple rules that are based on the spatial disposition of the choices. For instance, cycling through the 3 options clockwise (choose the option to the right of the previous choice) or contraclockwise, can in some situations end up being a winning strategy. For instance, if rock / paper / scissors are displayed in this order, and the player’s last choice was scissors, then a level-1 opponent will choose rock (it thinks player will repeat scissors) and the winning choice for the player is paper (the option to the left). Then in the following round, following the same logic, the winning move is again the one to the left (rock). Therefore, once this cycle is established, it can be a winning strategy to always choose the option to the left against a level-1 agent.

These heuristics can certainly explain some within-game learning. However, the order of display of the options on the screen was purportedly altered between the first and second games, so any successful spatial heuristic in RPS would no longer work in FWG, and players would need re-discover the new heuristic. This is inconsistent with observed results where most players scored higher in the first half of FWG than in the second half of RPS. It is also not consistent with players scoring significantly above chance performance in rounds 2-6 of the second game. Likewise, for NUMBERS, since the number of actions is different, this approach would not work and wouldn’t be consistent with evidence of transfer.

We have seen that the above strategies (associative mapping of actions, model based and model free reinforcement learning and spatial heuristics) cannot explain transfer of learning between games. Another important point is that they are not consistent with significant score differences between those facing level-2 and level-1 opponent. After all, if players were using some type of associative learning or spatial heuristics, then their scores should not depend on the degree of strategic sophistication of the opponent since their approaches would render this variable irrelevant.

These strategies can however result in a certain type of learningcalled *“implicit”,* a process that yields knowledge that is generally not verbalizable nor is it accessible to cognition (Holyoak and Spellman, 1993). A key aspect of this learning is that it relies on associative processes, circumvents cognition, and therefore is unable to produce mental representations of what is being learned (Mandler, 2004). Consequently, this type of learning cannot be operated on cognitively or transferred to another domain (Holyoak and Spellman, 1993). This would explain how some participants can learn within games yet fail to transfer learning between games.

In our view, the more likely explanation of the learning transfer is that it is driven by a group of participants that are able to build a mental representation of what the strategy of the opponent is. A successful mental representation would take the perspective of the opponent or endow it with intentionality in order to detect its strategy when the opponent is playing based on a level-k reasoning model. For instance, the player may think “My opponent is always trying to be one step ahead of me, therefore, I will be one step ahead of where it thinks I will be”.

This mental representation would facilitate the use of theory of mind abilities and thus enable the players to learn opponent strategies when they are based on human-like reasoning models such as level-k or cognitive hierarchy. This type of learning would be deemed “explicit” in the psychology literature as a process through which knowledge consists of cognitive representations of concepts and rules, as well as the relationship between them. It involves the evaluation of explicit hypotheses and results in better problem-solving skills (Mandler, 2004). Since it is less context dependent, this type of learning is generalizable to new situations.

For instance, if a player believes: “The computer is trying to beat whatever I played most recently” such a formulation is robust to changes in the actions description (whether “Scissors” or “Fire” were played in the last round, the strategy can tell me what to do next). It is also an insight that can be applied straight away at the beginning of the new game without needing to re-discover correct mappings to old games, which would be consistent with the early scores being significantly above chance. This type of learning would account for evidence for between games transfer, even in the very early rounds of a new game environment.

How does our explanation of explicitly modelling the opponent tie-in with the effect of similarity and condition? Our analysis showed that for FWG, evidence for transfer is strong for both types of players. FWG is very similar to RPS both strategically as well as in some aspects of description (number of actions) and opponent modelling would easily explain transfer. For Numbers, at first glance, the transfer to this dissimilar game seems to be modulated by the sophistication of the opponent, with evidence for transfer for level-1 facing players and no transfer for level-2 facing players. Shouldn’t a player modelling the opponent be able to transfer the knowledge to the dissimilar game as well?

We believe they should, and our results do not contradict this. For players facing the sophisticated opponent, we suggest that the lack of transfer for the dissimilar game might simply be the consequence of the difficulty of learning the strategy in the first place. Our computational modelling work shows that almost 50% of level-2 facing players did not play according to the optimal level-3 strategy. We believe that these players, by having more difficulty in representing the opponent’s strategy were unable to learn the playing rules or model the opponent strategy and thus, they could not generalise what they haven’t learned.

While our results are encouraging, there is room for further exploration on this topic so as to improve the experimental design and the computational modelling. As an area for future research, and to test our explanation of lack of transfer to the dissimilar game, it would be useful to distinguish the participants who managed to represent the opponent strategy correctly, from those who didn’t. This is likely to be an arduous task, but computational modelling can be used to identify whether the players are effectively best responding to the strategy or playing according to some non-optimal heuristic. Coupled with qualitative analysis of self-reports from the participants, this approach might yield interesting results in the categorisation of players. Our assumption is that the players whose actions are consistent with the optimal response (level k+1 for level-k opponent) would exhibit transfer for both similar and dissimilar games. The other players are likely to exhibit no transfer to dissimilar games, and might be able, due to their implicit learning to show tentative evidence of transfer in similar games.

Moreover, games allowing an easier identification of the participant’s degree of strategic sophistication should be preferred. An example of this is the 11-20 money request game (Arad & Rubinstein, 2012) which, by design, is superior to 3-action games such as RPS in identifying the player’s level as each choice can be mapped to a unique degree of iterated reasoning.

On the methodology, there are two aspects that we believe can be improved. First, we believe the incentives used to encourage participants to beat the opponent rather than play randomly were in some cases insufficient, as was seen by the very short playing time taken by some participants, which sometimes averaged to less than 2 seconds per game. Thinking about better incentives should help increase engagement with the experiment and yield more interesting responses. Finally, randomly changing the order of the second and third games for participants may be necessary. Since the first two games are very similar, players might have been “conditioned” to the simple structure of the environment after 100 rounds. Some participants expressed boredom or fatigue after such a long spell. It is possible that this has contributed to players not engaging with the third game.

# **Conclusion**

To sum up, this study shows that players can adapt to agents reasoning based strategies, such as level-k and best respond to it. Building a mental representation of the opponent’s strategy is the most likely explanation of learning transfer across games. However, our results also show significant individual heterogeneity in overall scores, as well as in the ability to learn the opponent strategy, and transfer that learning across situations. The success of opponent modelling also seems modulated by the complexity of the opponent’s strategy. Finally, successful transfer may in addition be modulated by the degree of similarity between the strategic structure of the state-action spaces of the games. Further study of the topic should help shed some light on the source of heterogeneity between participants, and answer questions regarding the form that such representation of the opponent takes.

**Acknowledgement**

I would like to thank my supervisor, Dr Maarten Speekenbrink for the invaluable help he provided throughout the life of this project. I have learned a great deal about programming experiments and computationally modelling human behaviour thanks to his support, and this work would simply not have been possible without his continued guidance.

1. ges stands for generalised eta squared, a measure of effect size of the factor in the ANOVA. [↑](#footnote-ref-1)
2. Because for the presence of within subject effects (game and condition), the usual way to plot standard errors is misleading, we choose therefore to omit them from our interaction plots. [↑](#footnote-ref-2)