REVEALED MARKOV STRATEGIES

ARTHUR DOLGOPOLOV

George Mason University

MIKHAIL FREER

University of Leuven (KU Leuven)

Abstract. Major problem with identification of strategies in the repeated games is the vastness of strategy space stemming from dependence on history (previous actions). Moreover, strategies can be of different complexity (depending on length of histories they use). In addition, there are two competing representations: a function of history and a finite automaton. We provide methodology for partial identification of strategies in repeated games for both representations. Moreover, we show that minimum complexity of a strategy that explains player's behavior can be efficiently found. In addition, we characterize a strict subset of finite automatons isomorphic to the set of history-dependent strategies. Finally, we illustrate the method using the experimental data on repeated prisoner's dilemma.

1 Introduction

Repeated games explain real life phenomena including price wars (Rotemberg and Saloner, 1986; Haltiwanger and Harrington Jr, 1991), models of time consistency (Chari and Kehoe, 1990) and risk sharing (Kocherlakota, 1996; Ligon et al., 2002).[1] In addition, repeated games have been widely used to build evolutionary models, e.g. (Fundenberg and Maskin, 1990; Friedman, 1991; Binmore and Samuelson, 1992; Kandori et al., 1993).

At the same time, empirical studies of repeated games are hindered by vast strategy space[2] – for a 4 period two-player game with two actions, the number of unique Markov strategies (mappings from histories to actions)[3] is 2^{85}. This complicates empirical (and

Date: April 15, 2018.

[1]See chapter 6 in Mailath and Samuelson (2006) for survey.

[2]Additionally, even when folk theorems do not apply and equilibrium predictions are clear, they can be very far from observed behavior, the latter fact giving rise to bounded rationality models of repeated games (see Kalai, 1990, for an early survey).

[3]Here we assume that strategy can depend on history of any depth from one to four. Hence, for a symmetric two player game with actions X and n periods, it is the number of possible mappings $(X \times X)^i \to X$ for $i \in 0..(n-1)$. The total number is therefore $\prod_{i=0}^{n-1} |X|^{(|X|^{2i})}$

experimental) studies of repeated games. Further problems arise from the fact that there is no unique way to model strategies in repeated games, and two different models co-exist. The first approach assumes that strategy is a function of a history, on which a person conditions her actions, subject to some maximal memory (Abreu, 1988; Aumann and Sorin, 1989; Sabourian, 1998; Barlo et al., 2009). The second assumes that a strategy is a Moore Automaton – a machine with a set of states (each mapped to an action) that transitions between states conditional on the observed profile of actions (finite automata in repeated games were suggested in Aumann (1981), studied in Rubinstein (1986) and the literature that followed; for a survey see e.g. Mailath and Samuelson (2006)).[4]

This paper provides methodology for partial identification of the strategies played in the repeated games. Moreover, we show that the minimum complexity of the strategy required to explain the observed behavior can be efficiently found. In addition we characterize the (strict) subset of Moore Automatons isomorphic to all the strategies of given finite memory. We illustrate the applicability of the methodology using experimental data on the repeated play in prisoner's dilemma. Our approach ignores the existence of the payoffs in the game as well as requirements for equilibrium behavior. Therefore, it should rather be perceived as a foundation for further identification of the strategies in repeated games that would rely on stricter assumptions.

There is a significant body of literature which investigates the behavioral patterns in repeated games. However, we can broadly divide it into two categories. First category uses the strategy method (Selten, 1967) to reveal a strategy of a player.[5] That is, players are asked which action they would take given observed history. The second category fits a set of "reasonable" strategies to explain the observed behavior.[6] However, both

[4]Automata were introduced to both represent and constrain strategies, by imposing preferences for smaller complexity in the supergame (Kalai, 1990; Neyman, 1985). We use automata for representation, following Osborne and Rubinstein (1994), thus avoiding the discussion of the supergame.

[5]Strategy method has a long history in repeated games and prisoner's dilemma in particular, going back at least to famous tournament by Axelrod and Hamilton (1981), for which strategies were solicited from other game-theorists. More recent work includes applying strategy method to two-stage voluntary contribution game (Muller et al., 2008), prisoner's dilemma with sanctioning (Falk et al., 2005; Bruttel and Kamecke, 2012), infinitely repeated prisoner's dilemma, cobweb markets (Sonnemans et al., 2004), dynamic asset pricing (Hommes et al., 2005). Linde et al. (2016) apply strategy method to a minority game with memory up to 5 periods. Here participants have the opportunity to run simulations against randomly chosen strategies of the previous round. Surprisingly, the paper results indicate absence of learning over rounds. Dal Bó and Fréchette (2015) elicit strategies of memory one in infinitely repeated prisoner's dilemma and compare reports to econometric estimation, showing that results are similar and that most strategies are of memory one. Additionally, for a survey and discussion of strategy method versus direct elicitation see Brandts and Charness (2011).

[6]Common econometric approaches include maximum likelihood (Fudenberg et al., 2012) and finite mixture modeling (Breitmoser, 2015). The latter paper combines data from four previous experimental studies on prisoner's dilemma and finds that most behavior is explained by Markovian strategies of

approaches suffer from the scalability issue – if the action space increases the strategy space becomes intractable (even fixing the maximum complexity). Applying strategy method would require asking too many questions to the subject and the set of candidate "reasonable" strategies becomes too large as well. At the same time the approach offered in the present paper does not suffer from this problem.[7] Finally, the method can be combined with fitting a set of "reasonable" strategies.

The remainder of the paper is structured as follows. Section 2 presents the main theoretical results of the paper. Section 3 shows the empirical application. Section 4 concludes. All proofs omitted in the text are collected in the Appendix.

2 Theoretical Framework

Let G denote the stage-game, which is completely characterized by the set of players $N = \{1, \dots, n\}$ and the set of actions $A = \times_{i \in N} A_i$, where A_i is the set of actions player i has. We assume that the same game is repeated for T periods. Let $a^t = (a_1, \dots, a_n) \in A$ be the profile of actions and let $\emptyset$ denote the absent (empty) profile of actions. The last part is an artificial construct which allows to control for the first periods of the game. Denote by $h^\tau(t) = (a^{t-\tau-1}, \dots, a^{t-1})$ the history of the length of τ which is present at the period t. Denote by H^τ the set of all possible histories of the length τ. We observe a finite experiment $D = \{(h^\tau(t), a^t)\}_{t=1}^T$ of T periods played. That is every period we know the history (action profiles taken in the previous τ periods) and the action taken by agent in the period t. Number of observations in this case is equal to the number of periods the game has been played. Two histories $h^\tau(t) = (a^{t-\tau-1}, \dots, a^{t-1})$ and $h^\tau(s) = (a^{s-\tau-1}, \dots, a^{s-1})$ are equal to each other if $a^{t-i} = a^{s-i}$ for every $i \in \{1, \dots, \tau+1\}$.

Let us elaborate a bit more on the structure of the data before proceeding further. Data is produced by repeated play of T_u periods with one partner and rematching to a new partner to play T_u more periods. This cycle goes for several rematchings and generates a data set of $T = T_u k$ observations, where k is amount of players with whom agent will be interacting during the experiment. The assumption of similar T_u's for different matchings could be relaxed with all the results being similar, however, we keep the assumption to avoid further abuse of notation. Finally, note that this structure requires a player to use the same strategy in different matchings. This assumption can also be relaxed, but would require cutting the data into smaller chunks and therefore decreasing the power of test.

The rest of the section is organized as follows. First we show the criteria for rationalization of the data with a Markov strategy. Next, we show the criteria for rationalization

memory one. Engle-Warnick and Slonim (2006) also fit a prespecified set of machines to explain trust game behavior finding that inferred strategies are almost always best responses of other inferred strategies.

[7]However, there is a caveat – the more complex the game becomes the less information can be reveled from similar amount of data.

of the data with a Moore automaton. Third, we show the connection between Markov strategies and Moore automatons. Finally, we introduce the (nonparametric) decision making errors to the models and provide consistent estimate for the model.

2.1 Markov Strategies A **Markov strategy** (MS) $s_i^\tau : H^\tau \to A_i$ is a function that maps history (of length τ) to an action. We say that a strategy is of memory τ if it depends on the histories of the length τ. An experiment is **rationalizable with Markov strategy of memory** τ if there is a strategy $s_i^\tau : H^\tau \to A_i$, such that $s_i^\tau(h^\tau(t)) = a^t$.

Definition 1. *An experiment* $D = \{(h^\tau(t), a^t)\}_{t=1}^T$ *satisfies* **Axiom of Revealed Markov Strategies of memory** τ **(ARMS-**τ**)** *if* $h^t = h^s$ *implies* $a^t = a^s$ *for every* $t, s \in \{1, \dots, T\}$.

ARMS-τ is obviously a necessary condition for the rationalizability of the experiment with MS, and it can also be shown to be sufficient via direct construction of the entire MS of memory τ from observed behavior.

Proposition 1. *An experiment is rationalizable with Markov strategy of memory* τ *if and only if it satisfies ARMS-*τ*.*

First of all, Proposition 1 shows that not every experiment is rationalizable with MS. Consider the experiment with prisoner's dilemma game given two players and two strategies C and D, and the following observed play:

$$\dots (C, C); (C, D); (D, C); (C, C); (C, D); (C, D); \dots$$

Let us test whether this experiment is rationalizable with history of memory 2. Consider first two observations, which form the history $h = ((C, C); (C, D))$ with $s(h) = D$. Next, consider the last three observations, with the same history h formed by the third and the fourth observations and with output being the action taken $s(h) = C$. This is a violation of ARMS-2, and therefore, experiment is not rationalizable with MS of memory 2. At the same time we can find an experiment that cannot be rationalized with experiment of any finite memory. Consider an experiment with two re-matchings which start with two different actions. Taking any length of the history this would be a violation of ARMS-τ, for any finite τ, because the history would be just a repetition of the $\emptyset$'s and therefore, would be similar for every τ.

Moreover, if the experiment is rationalizable, then the minimal complexity (length of history) of the strategy is well-defined and can be efficiently found. By computational efficiency we mean that it can be found in polynomial time. Moreover, we show that the minimal complexity lies between 1 and T_u and can be found as efficiently as a binary search over the possible τ's. Recall that the complexity of binary search is $\log_2(n)$, where n is the ordered length of the interval over which we search and equal to T_u in our case. In turn, each application of ARMS-τ takes $\tau(\tau-1)$ steps. This gives us the final complexity estimate as superposition of the algorithms.

Corollary 1. *Experiment is rationalizable with Markov Strategy if and only if it is rationalizable with Markov strategy of memory T_u. Moreover, if experiment is rationalizable with Markov strategy of memory T_u, then the minimal memory τ, such that experiment is rationalizable with Markov strategy of memory τ is well-defined and can be found as fast as $\mathcal{O}(T(T-1)(\log_2(T_u)+1))$.*

2.2 Moore Automaton Alternative way to model strategies in repeated games is through Moore Automata (Moore Machines). Moore automaton (MA) is completely characterized by the set of states, the input (signals / action profiles) and output (actions) alphabets, the output function (mapping states to actions) and the transition function (moving between states depending on a signal received). Let $\Omega = \{\omega_1, \dots, \omega_K\}$ be the set of states. Let A_i be the output alphabet and A_{-i} be the input alphabet. Let $\zeta : \Omega \to A_i$ be the output function and $\mathcal{T} : \Omega \times A_{-i} \to \Omega$ be the transitition function. Note that output function, does not have to be bijective – different states can deliver the same action.

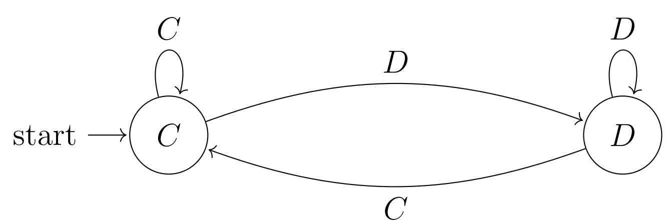


FIGURE 1. Moore Automaton for tit-for-tat strategy

Figure 1 presents the MA for the "tit-for-tat" strategy in prisoner's dilemma game. There are two actions at every stage game: C and D, which serve both as output and input symbols. Circles present the states of the MA with the letters inside the circles representing the output symbol corresponding to the state. Arrows represent the transitions between states with letters on them showing the signal (strategy of another player), upon receiving which the transition happens.

The task requires us to reconstruct the reverse of the ζ, which would map observations into the states. Denote this function by $\xi : \{1, \dots, T\} \to \Omega$. Assume we know the set of states, then we are only left to determine the output and transition functions. In order to reconstruct the ζ and $\mathcal{T}$ we need two conditions to be satisfied. First, is that every state is mapped to the unique action. Second, transition mapping also has to be a function – for every signal there is unique state to be mapped to.

Definition 2. *An experiment satisfies* **Axiom of Revealed Moore Automata with Ω states (ARMA-Ω)** *if and only if there is $\xi : \{1, \dots, T\} \to \Omega$ such that (1) if $\xi(t) = \xi(v)$, then $a_i^t = a_v^t$; and (2) if $\xi(t) = \xi(v)$ and $a_{-i}^t = a_{-i}^v$, then $\xi(t+1) = \xi(v+1)$.*

ARMA-Ω is a more complicated axiom than ARMS-τ. However, it can be easily restated as the set of linear inequalities. Denote by $x_{t,\omega} \in \{0,1\}$ the indicator variable, which shows whether $\xi(t) = \omega$.

$$\text{(MA)} \qquad \begin{cases} x_{t,\omega} + x_{v,\omega} \leq 2 - \mathbf{1}_{a_i^t \neq a_i^v} & \forall t, v \leq T;\ \forall \omega \in \Omega \\ x_{t,\omega} + x_{v,\omega} \leq 2 - \mathbf{1}_{x_{t+1,\omega'} \neq x_{v+1,\omega'}} & \forall t, v \leq T \\ & \text{ such that } a_{-i}^t = a_{-i}^v; \\ & \text{ and } \forall \omega, \omega' \in \Omega \\ \sum_{\omega \in \Omega} x_{t,\omega} = 1 & \forall t \leq T \end{cases}$$

The first two inequalities define the conditions (1) and (2) from ARMA-Ω and the last equality shows that every observation should be assigned a state. Finally an experiment is rationalizable with MA with Ω states if there is a MA which generates a sequence of states: ω^t such that $G(\omega^t) = a_i^t$ and $T(\omega^t, a_{-i}^t) = \omega^{t+1}$.

Proposition 2. *The following statements are equivalent:*

(1) An experiment is rationalizable with Moore automaton with Ω states;
(2) There are $x_{t,\omega}$ that satisfy system MA;
(3) ARMA-Ω is satisfied.

However, not every data set is rationalizable with some MA. For the intuition behind this, note that MAs with different initial states are different MAs. Hence, if the data set contains two different initial responses, then it cannot be rationalized with any MA. Moreover, Proposition 2 shows that ARMA-Ω is equivalent to the set of linear inequalities, hence, the minimal complexity of underlying MA is well defined and can be found as a result of minimization problem. This allows us to state the following immediate corollary.

Corollary 2. *If an experiment is rationalizable with some Moore automaton, then the minimal complexity (number of states) of the underlying Moore automaton can be found as*

$$\sum y_\omega \to \min$$

subject to MA, $y_\omega \in \{0,1\}$ and

$$y_t \geq \sum_{t=1}^{T} x_{t,\omega}$$

2.3 Connection between representations First we define the subset of Moore automatons that is isomorphic to the set of all possible MS of given complexity. Next, we show that MS and MA are observationally equivalent. Finally, we present the (tight) bound on the complexity of MA given complexity of MS and present the algorithm to construct minimal MA given MS.

Let $T^n(\omega, a^t, a^{t+1}, \dots, a^t)$ be the superposition of transition functions starting from state $\omega \in \Omega$. More formally, we can introduce it via the following recursive definition.

$$T^1 = T(\omega, a^t) \text{ and } T^{i+1} = (T^i(\omega, a^{t+i}, \dots, a^t), a^{t+i+1})$$

Recall that transition function maps states and action profiles back into the set of states. Hence, every superposition of the transition functions would return a state of the MA. This is an analog of the walk of length n along the graph representing the MA.

Definition 3. *Moore Automaton $(\Omega, G, \mathcal{T})$ is τ-transition invariant (τ-TI) if*

$$G(T^\tau(\omega, a^{t+\tau}, \dots, a^t)) = G(T^\tau(\omega', a^{t+\tau}, \dots, a^t))$$

for all $\omega, \omega' \in \Omega$ and $a^t, \dots, a^{t+\tau} \in A \cup \{\emptyset\}$.

Transition invariance implies that starting from every state of the MA the same state (or states with the same output symbol) should be reachable in τ steps, if those steps are similar. Consider 1-TI, then conditioning on the same previous profile of actions observed (regardless of the current state) we should obtain the same output. Hence, every output symbol can be reached in one step and depends only on the profile of actions in the previous period.

Proposition 3. *Every Markov Strategy of memory τ can be represented by a τ-TI Moore automaton and vice versa.*

Proposition 3 shows the equivalence between the class of MS of memory τ and a subset of MA. In particular this implies that every MA that satisfies τ-TI for some finite τ can be converted into MS with finite memory. However, not every MA can be converted to MS with finite memory. To illustrate this consider again an example game of two players with two strategies C and D (e.g. prisoners' dilemma).

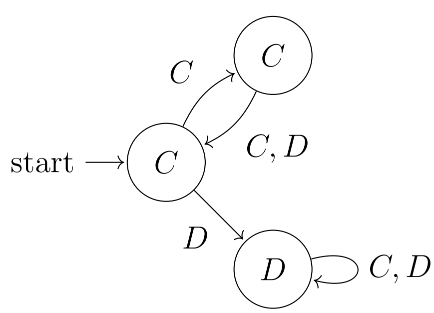


FIGURE 2. Moore Automaton that cannot be represented as a Markov Strategy of finite memory

Figure 2 shows the MA that cannot be represented as MS with finite memory. This automaton would imply that player uses strategy C unless her counterpart uses D in odd period. Otherwise player keeps on using D for the rest of the game. Hence, in order to determine the correct action it is necessary to know whether the period of observation is odd or even,[8] which in turn requires knowing the entire history of observations. Therefore, this MA does not satisfy τ-TI for any finite τ and cannot be represented by MS of finite memory.

However, finite data set generated by MA from Figure 2 would be consistent with ARMS-$(T_u + 1)$. Hence, there would be the MS of finite memory which can rationalize the observed behavior. This fact is a result of imperfect (finite) observability problem, and can be further generalized given the results from Propositions 1 and 2.

Corollary 3. *A finite experiment can be rationalized by Moore automaton if and only if it can be rationalized by Markov strategy of finite memory*

Corollary 3 shows the observational equivalence between the Markov strategies and Moore automatons. Further we show that the corresponding minimal MA can be reconstructed from MS using a simple linear program. Recall that ARMS-τ allows to partially identify the Markov strategy s_i and, therefore, no symbol would be attached to some histories. Hence, the partially identified strategy is a mapping from the set of histories to the set of actions that includes "any action" symbol – $s_i : H_\tau \to A_i \cup \{*\}$, where $*$ is "any action" symbol. Let x be a binary matrix of the size of $|H_\tau| \times |\bar{\Omega}|$, with $\bar{\Omega}$ denoting the upper bound on the number of states. Let y be a binary vector of length $|\bar{\Omega}|$. Denote by $\mathbf{1}_{f(x)}$ the indicator function that returns 1 if the logical expression is correct and 0 otherwise.

[8]Such automaton is called an automaton with a counter. The intuition behind the counter is that for some values of the inputs the automaton cycles, producing the same sequence of action profiles indefinitely. This allows the automaton to "keep track of time", for example of odd and even states, and at the same time precludes the outside observer from learning the current state of MA from finite histories. Formally, a counter is such non-trivial cyclical sequence of states $\{q_0, q_1, ..., q_k = q_0\}$ that it results in a history (of length k) consisting only of some (non-trivially) repeating subsequence. For example, for the automaton in Figure 2 the history for the cycle is $\{(C, C), (C, C)\}$, that is (C, C) repeating twice. Moreover, using this connection we can infer categorical similarity of the Proposition 3 and the McNaughton-Papert-Schutzenberger theorem (MPS theorem) in formal grammar theory and draw the conjecture about further results. MPS theorem shows that the set of automatons without counters is isomorphic to the set of star-free regular expressions. Hence, if the set of τ-TI MA is isomorphic to the set of MA without counters, then it is isomorphic to the set of all star-free regular expressions. This would be a refinement of the result from Kalai and Stanford (1988) which shows that all MS (including those of infinite memory) is isomorphic to the space of all possible expressions in a formal grammar.

$$
(\text{MS}\to\text{MA}) \qquad \begin{cases} \sum\limits_{\omega\in\bar{\Omega}} y_\omega \to \min & \\ x_{\omega,t} + x_{\omega,v} \le 2 - \mathbf{1}_{s_i(t)\neq s_i(v)} & \forall t,v \in H^\tau, \forall \omega \in \bar{\Omega}; \\ x_{\omega,t} + x_{\omega,v} \le 2 - \mathbf{1}_{x_{\omega,t+(s_i(t),a_{-i})}\neq x_{\omega,v+(s_i(v),a_{-i})}} & \forall t,v \in H^\tau, \forall \omega \in \bar{\Omega}; \\ & \text{and } \forall a_{-i} \in A_{-i}; \\ \sum\limits_{\omega\in\Omega} x_{\omega,t} = 1 & \forall t \in H^\tau; \\ y_\omega \ge x_{\omega,t} & \forall t \in H^\tau, \forall \omega \in \bar{\Omega}. \end{cases}
$$

Condition MS→MA shows the linear optimization problem that constructs the minimal automaton from a partially identified MS of memory τ. We use the indicator function that may seem to be a nonlinear operator for better illustration of the condition. However, each of indicator functions can be linearized by assigning numerical values to the strategies and outcomes and using the absolute values of differences between these numerical values. Condition MS→MA is very similar to the condition MA, hence it would recover the minimal automaton (this can be proven using the same logic as in Proposition 2). Hence, we are left to provide the upper bound on $|\bar{\Omega}|$. This bound can be directly inferred from the proof of Proposition 3. However, we assume that $|A_i| \le |A_{-i}|$, which is a technical assumption to avoid further abuse of notation.

Corollary 4. *If a finite experiment is rationalizable with Markov strategy of memory τ, then the minimal Moore automaton that represents the strategy is defined by conditions MS→MA with $|\bar{\Omega}| \le |A_{-i}|^\tau + 1$.*

2.4 Introducing Decision Making Error We employ a quantal response model according to which player with probability P chooses the actions prescribed by MS and with probability $(1-P)$ chooses any other action (see Figure 3 for illustration). We do not make any assumption about the distribution.[9] In addition for reasoning below we assume that $P > .5$, that is the identifying assumption.

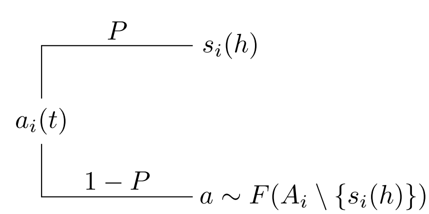


Figure 3. Quantal Response Process

[9]Formally this is a quite strong assumption that says that if player makes a random action, then she cannot choose the action prescribed by the strategy used. However, this assumption is required for identification and one can easily check that relaxing it would immediately destroy the possibility of identification for P.

Complication in this case is that we do not know ex-ante what action is prescribed by the MS. Denote by $\hat{p}(s_i(h)|h)$ the observed frequency of selecting the most frequently observed action given history h. Denote by $\hat{q}(h)$ the observed frequency of history h appearing in the data – fraction of times h is observed over T. Further we show that frequency of the most frequently chosen strategy (given the history) is the consistent estimator of P.[10]

Proposition 4. *If $P > .5$, then $\hat{p}(s_i(h)|h)$ is a consistent estimator of P, for every history h.*

Proposition 4 shows that frequency of the most frequently chosen strategy given history h is a consistent estimator of P. Hence, for every history we obtain a consistent estimator of P. Next, we need to aggregate this collection of estimators into a number. Let $\hat{P} = \sum_{h \in H} \hat{q}(h)\hat{p}(s_i(h)|h)$. Note that $\hat{P}$ (if we keep the weights constant) would also be a consistent estimate of P. This assumption about keeping the weights constant is dictated by the complexity of the data generating process. In particular, the data generating process depends on the strategies used by other players, which are not observed. Moreover, given that we can estimate the asymptotic variance of $\hat{p}(s_i(h)|h)$ by $\frac{\hat{p}(s_i(h)|h)(1-\hat{p}(s_i(h)|h))}{T\hat{q}(h)}$ the proposed estimate of P discounts the contribution of the estimates driven from a very small amount of observations and weights higher the more precise ones. More precisely, we can estimate the asymptotic variance of $\hat{P}$ as follows.

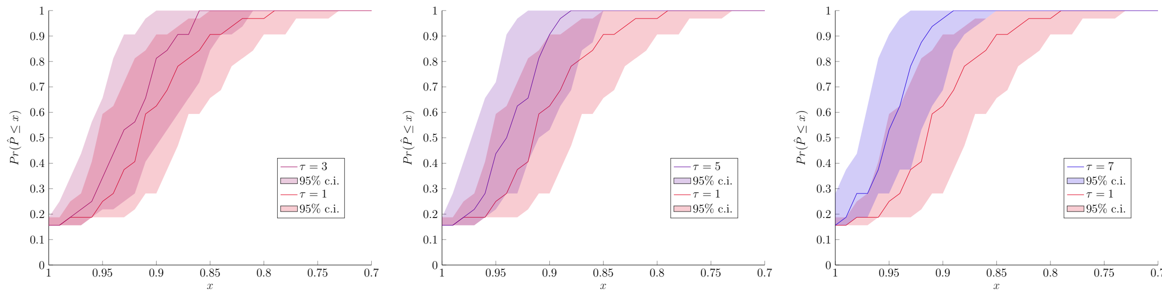
$$\text{Var}(\hat{P}) = \sum_h \hat{q}^2(h) \frac{\hat{p}(s_i(h)|h)(1 - \hat{p}(s_i(h)|h))}{T\hat{q}(h)}$$

Note that given the equivalence and transformation of representation results here we concentrate exclusively on the estimator for the quantal model for the MS. Formally, one can do the same for the MA, but this would be closely linked to the vast literature on estimation of the Markov chains (see for instance Ephraim and Merhav, 2002).

3 Empirical Illustration

We use the data from Embrey et al. (2015), who conduct a finite horizon repeated prisoner's dilemma with 2-by-2 factorial design: two sets of payoffs in the stage game, and two time horizons (4 and 8 rounds). What follows is the analysis for 8 round treatment with 32 subjects, each playing 20 games and making 160 choices in total. Experiment was conducted with undergraduate students at NYU, subjects had access to complete history of play and were randomly rematched after every game of 8 rounds.

[10] Recall that estimator is said to be consistent if as the number of data points increases indefinitely, the resulting sequence of estimates converges in probability to the true value of the parameter. This means that the distributions of the estimates become more and more concentrated near the true value of the parameter being estimated, so that the probability of the estimator being arbitrarily close to the true value converges to one.



(a) $\tau = 3$ (b) $\tau = 5$ (c) $\tau = 7$

FIGURE 4. Comparing cumulative distribution functions of $\hat{P}$ estimates of $\tau = 1$ with other memory lengths

Figure 4 presents the estimates of $\hat{P}$ for Markov strategies of different memory. Solid lines present the cumulative distribution functions of $\hat{P}$ and the shadowed areas present the 95% confidence intervals. Confidence interval of the cumulative distribution function for $\hat{P}$ for $\tau \in \{3, 5, 7\}$ intersects with the corresponding confidence interval for $\tau = 1$. Hence, there is no significant gain in explanatory power from increasing the memory size. There is no control for the power of test, however, the presented test is rather conservative and controlling for the power of test would only make the difference smaller than on Figure 4. The higher is the chosen τ, the lower is the power of test. Hence, controlling for the power of test would only make the $\hat{P}$ estimates for $\tau > 1$ closer to the one of $\tau = 1$.

Figure 5 presents the distribution of error for $\tau \in \{1, \ldots, 8\}$. Frequency of mistakes is higher during the first match and during the first period of all the matches. This trend can be explained by the fact of learning and experimenting with the strategy to be used. In addition, there is a trend for systematic deviation from the chosen strategy in the last round (for matches 1-7), round 7 (for matches 8-15) and round 6 (for matches 16-20). To explain this let us recall that subjects are playing the finitely repeated prisoners dilemma with the unique equilibrium being (D, D). Hence, it is important to see the direction of the deviations. Figure 6 presents the distribution of the deviations by type of act chosen. First column shows the distribution of errors such that subject chooses D, while MS predicts C and the second column shows the distribution of errors such that subject chooses C while MS predicts D. First type of deviation almost completely absorbs the deviations in the later periods. Moreover, the deviations we observe are more towards defecting – 70% for $\tau = 1$, from 62% (for $\tau = 6$) to 83% (for $\tau = 4$) overall out of all the observed deviations. In particular, if we consider the periods in which the majority of deviations is observed the relative frequency of deviations towards defecting is between 84 % and 85 % for all $\tau \leq 4$.

Hence, deviation in the later rounds can be explained by the subjects doing the backward induction. Therefore, we observe that over the course of the experiment

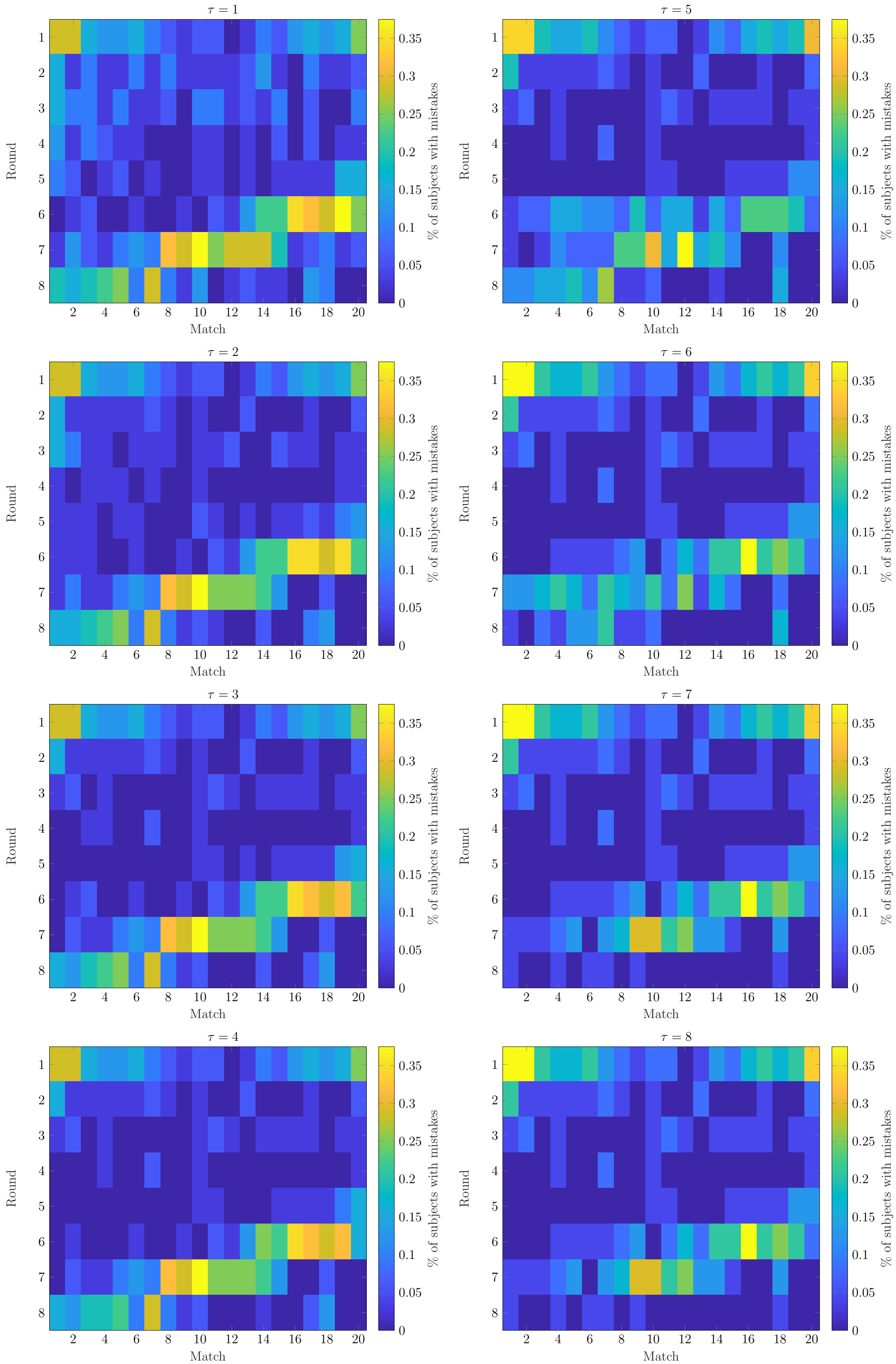


FIGURE 5. Distribution of errors for different τ

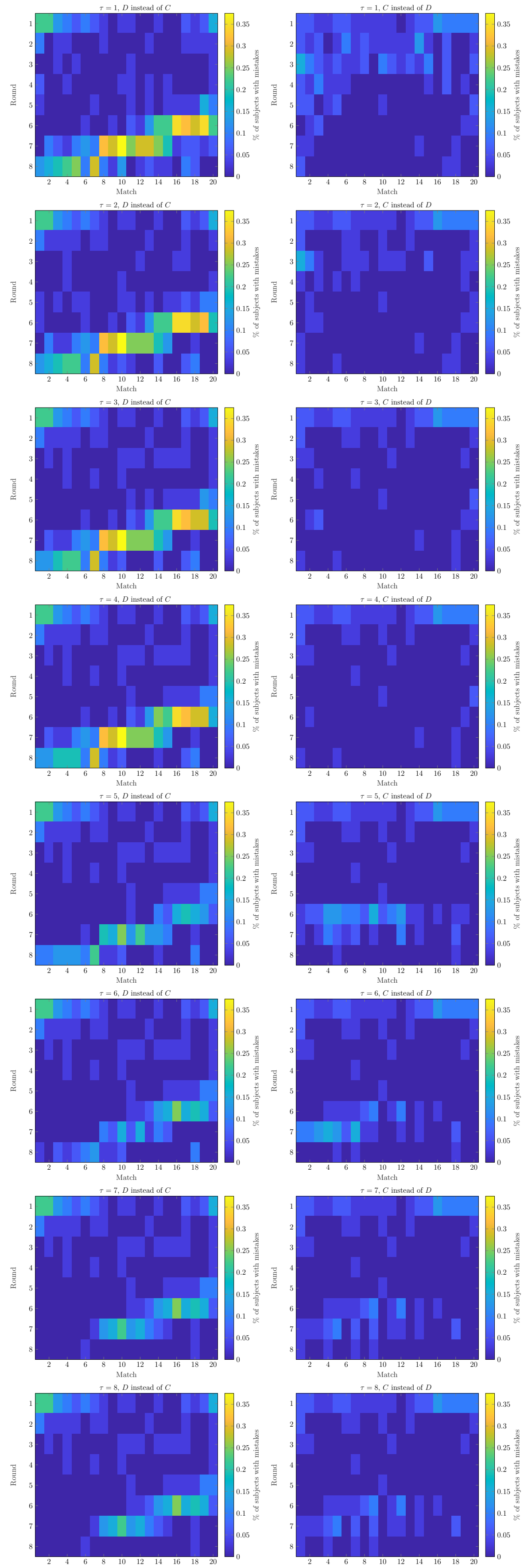


FIGURE 6. Distribution of errors for different τ

subjects do additional step of backward induction and get closer to the subgame perfect equilibrium.

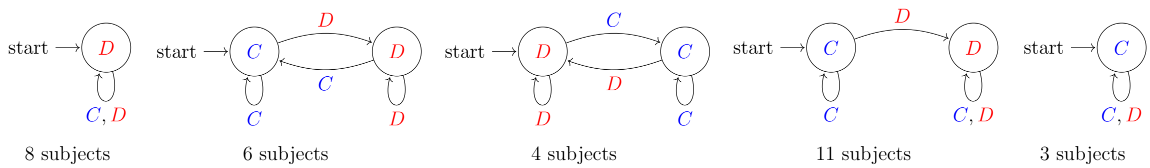


FIGURE 7. Revealed Strategies $\tau = 1$.

Figure 7 presents the revealed strategies: always defect, tit-for-tat (with two different initial states), trigger strategies and always cooperate. The most popular (11/32 subjects) strategy is the trigger strategy. Second most popular strategy is the tit-for-tat (10/32 subjects), but there are two instances of this strategy. Six subjects follow the standard tit-for-tat strategy, which starts with C. Four other subjects start tit-for-tat with D. Note that the latter strategy was not used in the prior studies. In order to distinguish whether these strategies are different we compare the frequency of choosing the initial state. Those subjects who start with C choose it with frequency of 90.83% and those who start with D choose it with frequency of 80%. If we take each subject with either of the versions of tit-for-tat as a single observation, the difference in mean probability of cooperating at first round is significant between the two groups according to Wilcoxon rank sum test at 1% level.

Note that tit-for-tat which starts with D would deliver cooperation only to the unconditional cooperating strategy and therefore, it can be perceived as the more pessimistic version of the usual tit-for-tat strategy. The two remaining strategies are simple: always cooperate and always defect, with the latter being more frequent (used by 3 and 8 subjects respectively).

4 CONCLUDING REMARKS

The paper constructs the methodology for revealing strategies (including decision making error) used in the repeated games. Moreover, we show the equivalence of the representations (MA and MS) and provide efficient algorithms to map representations into each other. Additionally, we illustrate the methodology by taking it to the data on repeated play of prisoner's dilemma. We find that observed behavior can be explained well by the strategies with of memory one and there is no significant additional explanatory power brought by increasing the memory. Finally, we reveal that subjects are using the following Markov strategies: always defect, always cooperate, trigger and tit-for-tat. However, tit-for-tat is used in two different instances: cooperation as initial state and defection as initial state. The latter strategy was not considered in the previous studies.

Recall that the method proposed is a foundation for further nonparametric research in repeated games. In particular, based on the revealed strategies one can easily check whether the observed behavior is a best-response to the actions of other players (or population of other players). Additionally, the method proposed can be easily modified to reveal strategies that satisfy equilibrium conditions. That is, instead of revealing strategies unconditionally, it would constrain the set of possible strategies to be best-responses to each other.

Appendix A: Proofs

Proof of Proposition 1

Proof. (⇒). On the contrary assume that ARMS-τ is violated. Then, there is a history h^τ, such that $s_i^\tau(h^\tau) = \{a^t, a^s\}$. This contradicts the fact, that s_i^τ is a function (cannot be set-valued).

(⇐). Let us prove this by constructing a Markov strategy of memory τ, that rationalizes the experiment. Denote by $\bar{H}^\tau$ the set of histories observed in the experiment. Then, for every $h^\tau \in \bar{H}^\tau$ let $s_i^\tau(h^\tau) = \{a^t : h^\tau(t) = h^\tau\}$ and for every $h^\tau \in H^\tau \setminus \bar{H}^\tau$ let $s_i^\tau(h^\tau) = a_i^{t-1}$.

That is for every unobserved history we are mapping the strategy into the last action taken. For example, consider an experiment $((C,C);(C,D);(D,C))$ and $\tau = 1$. Then $\bar{H}^1 = \{\emptyset, (C,D), (C,C)\}$ and $H^1 \setminus \bar{H}^1 = \{(D,D),(D,C)\}$. Strategy $\{\emptyset, (C,C) \to C;$ $(C,D),(D,C),(D,D) \to D\}$ (grim-trigger) rationalizes this experiment.

ARMS-τ implies that the constructed mapping is a function. Assume on the contrary that there is a set-valued $s_i^\tau(h^\tau)$, then $h^\tau \in \bar{H}^\tau$ (due to construction). This would imply that there are two histories such that the corresponding actions are different. That is a direct violation of ARMS-τ. Since constructed Markov strategy is a function, then the fact that it rationalizes the experiment follows from the construction. □

Proof of Corollary 1 Before we start the proof let us explain the idea behind it. One can think of rationalization as the inifnitely dimensional binary vector, where 1 at the τ-th place means that data set is rationalizable with the MS of memory τ and 0 at τ-th place means that data set is not rationalizable with MS of memory τ. Searching for minimal τ means that finding the first entry of 1 in this sequence.

The proof proceeds in two steps. First we show that rationalization of the data with MS of memory τ is well-ordered with respect to τ. That is if 1 enters at the τ-th place, then there are no zeros to the right from it. Second we show that it has the exact upper bound – there is an ex-ante known number such that if zero enters at its place, then the whole sequence consists of zeros only.

These two properties are sufficient to apply the binary search tree to find the first unit entry, because knowing whether 1 or 0 enters at the τ-th place allows to make inference about the elements to the right or to the left from this place. In particular, if 1 enters at the τ-th place, then every element on the right should be 1; if 0 enters at the τ-th place, then every element to the left from this entry should be 0. Moreover, the second part is required in order to set up the initial condition and guarantee that the search interval is bounded. Both of these together guarantee that we can conduct a search over the (almost) fully balanced binary search tree (rooted at $[T_u/2]+1$) with complexity of $\mathcal{O}(\log_2(T_u)+1)$. For example, the minimal τ for a game with $T_u = 6$ rounds can be found with a binary search in Figure 8. Every square node in the tree represents one run of the corresponding ARMS-τ, and since the height is three, search requires only three runs at the worst.

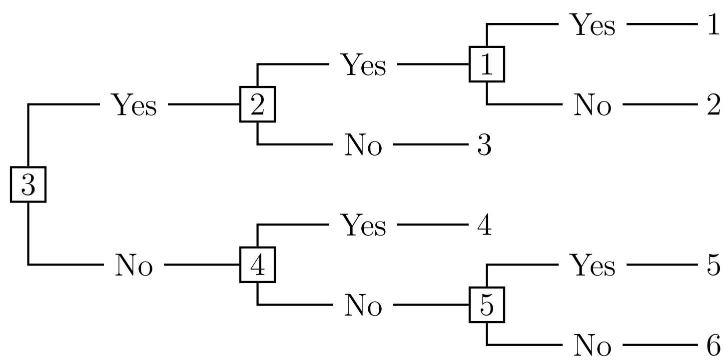


Figure 8. Binary Search Tree

Lemma 1. *If experiment is rationalizable with Markov strategy of memory τ, then it is rationalizable with Markov strategy of memory τ' for every $\tau' \geq \tau$.*

Proof. To prove this corollary it suffices to show that if ARMS-τ is satisfied, then ARMS-$(\tau+1)$ is satisfied as well. On the contrary assume that ARMS-$(\tau+1)$ is violated. That is there are histories $h^{\tau+1}(t) = h^{\tau+1}(s)$, such that $a^t \neq a^s$. Take shorter histories $h^\tau(t)$ and $h^\tau(s)$ which start from $t-\tau-1$ and $s-\tau-1$ respectively. Then, $h^\tau(t) = h^\tau(s)$ and $a^t \neq a^s$ implies violation of ARMS-τ. □

Lemma 1 shows that the vector of outcomes of testing ARMS is well-ordered. Formally, we show that only if 1 enters at the τ-th place then every entry on the right should also be equal to 1. Second claim follows immediately, since if 0 enters at the τ-th place, 1 enters at the τ'-th place and $\tau' < \tau$, then we immediately get contradiction to Lemma 1.

Lemma 2. *An experiment is rationalizable with Markov strategy of some memory if and only if it is rationalizable with Markov strategy of memory T_u.*

Proof. In order to prove this we show that ARMS-(T_u+1) is violated. Take histories $h^{T_u}(t) = h^{T_u}(s)$ such that $a^t \neq a^s$. Then, $h^{T_r+1}(t) = (\emptyset, h^{T_u+1}(t))$ and $h^{T_u}(s) = (\emptyset, h^{T_u}(s))$ – $h^{T_u+1}(t) = h^{T_u+1}(s)$. That is a violation of ARMS-(T_u+1). □

Proof of Proposition 2 We prove the proposition by constructing a cycle of implications.

Proof. **(1)⇒(2)** Given a Moore Automaton, let us directly construct the solution for the MA. Let $x_{t,\omega} = 1$ if and only if at the time of observation t the Moore Automaton was in the state ω. The last inequality is then automatically satisfied, because at every point in time automaton can be only in one state. The first equation is also satisfied, because $G : \Omega \to A_i$ is a function – there is a unique action corresponding to every state. The second equation is also satisfied, because $T(\omega, a)$ is a function, that is if in two time periods automaton was in the same state and the same signal was received, then transition function would map these two observations to the same state.

(2)⇒(3) Let us construct ξ out of the solution for the MA. Let $\xi(t) = \bigcup_{x_{i,\omega}=1} \omega$ and last equation guarantees, that ξ would be a singleton. Moreover, the first equation guarantees, that $G(\xi)$ is a singleton, because $a_i^t = a_i^v$ if $\xi(t) = \xi(v)$. Moreover, if $\xi(t) = \xi(v)$ and $a^t = a^v$, then $\xi(t+1) = \xi(v+1)$, due to the second equation. Therefore, the transition matrix is well defined.

(3)⇒(1) Further we finish the construction of the Moore Automaton based on the ξ. Both output and transition mappings are so far functions by construction. Hence, the only part to be completed is the transition function. Let $T(\omega, a) = \omega$ for every $a \in A$ such that $a \neq a^t$ for every $t \in \{1, \dots, T\}$. □

Proof of Proposition 3

Proof. **(⇒)** Further we construct the Moore automaton that represents the Markov strategy of memory τ. Let us start from initial state ω_0, and let $G(\omega_0) = s(\emptyset)$. Further we introduce a new state of the world for every possible history $h^\tau \in H^\tau$. Denote the states by $\omega(h^\tau)$. Let $G(\omega(h^\tau)) = s(h^\tau)$. Denote by $h^\tau + a = (h^{t-\tau-2}, \dots, h^{t-1}, a)$ – that is an operation of adding a^t to the queue (line) of the length τ.[11] Since we are adding element to the end of the queue, it pushes out the element in the beginning of the queue. Next, let $T(\omega(h^\tau, a)) = \omega(h^\tau + a)$ for every $\omega \in \Omega$ and $a \in A$.

We constructed a Moore Automaton, since it has fully defined set of states and well defined transition function. Hence, we are left to show that it satisfies τ-TI. By construction $T(\omega(h^\tau), a) = \omega(h^\tau + a)$. Denote by $\tilde{h}^\tau = (a^t, \dots, a^{t+\tau})$ for some sequence of signals.. Hence, applying this τ times, $T^\tau(\omega(h^\tau), a^{t+n}, \dots, a^t) = \omega(h^\tau + a^t + \dots + a^{t+\tau}) = \omega(\tilde{h}^\tau)$

[11]Readers familiar with computer science literature, can see that we are actually using the formal definition of the structure called queue.

for every $\omega(h^\tau) \in \Omega$.[12]

(⇐) Let $s_i(h^\tau) = \bigcup_{\omega\in\Omega} G(T^\tau(\omega, h^\tau))$. Hence, for every possible history there is an action assigned and $s_i : H^\tau \to A_i$ is a function, since MA satisfies τ-TI. That is $|\bigcup_{\omega\in\Omega} G(T^\tau(\omega, h^\tau))| = 1$. □

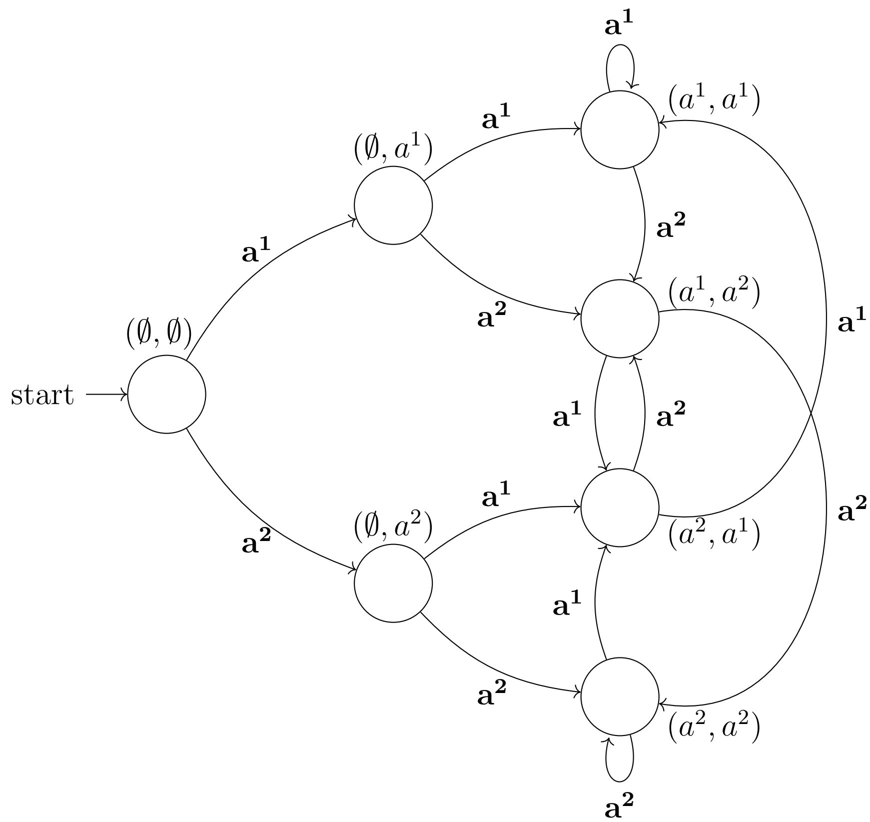


FIGURE 9. Construction of Automaton from Markov Strategy of $\tau = 2$

For example, the automaton for $\tau = 2$ and $|A| = 2$ would require creating 7 states at first as in Figure 9. Once the actions are assigned to states based on the Markov strategy, the automaton may be found to contain equivalent states and can be reduced.

Proof of Corollary 3 Proposition 3 together with Lemma 2 implies that every data set that is rationalizable by Markov strategy is rationalizable by Moore automaton. Hence, we are left to prove the reverse: everything that can be rationalized by MA can be rationalized by MS.

[12]The last equality follows from the rule of summation we imposed. Recall that in the queue of the length τ every next element kicks off the previous one. Hence, adding τ elements, would crowd out all of the original elements and the queue would consist only of τ consecutive signals.

Proof. Assume that the experiment is rationalizable with Moore automaton and not with MS of finite memory. Recall that according to Lemma 2 an experiment is rationalizable with MS of finite memory if and only if it is consistent with ARMS-T_u. Hence, if data is not consistent with MS of finite memory, there should be a violation of ARMS-T_u.

If $T_u = T$, then it is immediately rationalizable with MS of memory T_u. Hence, we assume that there are at least two matchings to get a violation of ARMS-T_u. That is, these matchings contain the same history of length T_u resulting in different actions taken by the player.

Recall that MA has the unique initial state $\omega(\emptyset)$, and we assume that the automaton is restarted with every rematching. Violation of ARMS-T_u implies that there are $h^{T_u}(k) = h^{T_u}(l)$ that result in different actions taken by the player. At the same time $\omega^{T_u}(k) = T^{T_u}(\omega(\emptyset), h^{T_u}(k))$ and $\omega^{T_u}(l) = T^{T_u}(\omega(\emptyset), h^{T_u}(l))$. Hence, $\omega^{T_u}(k) = \omega^{T_u}(l)$ and therefore the actions taken by the player should be the same (because of the uniqueness of the output symbol). Hence, there cannot be a violation of ARMS-T_u and the data is rationalizable with MS of memory T_u. □

Proof of Corollary 4

Proof. Consider construction of the MA out of MS from the proof of Proposition 3. It starts from the initial state. Let us start from the "incomplete" histories, i.e. histories that contain at least one $\emptyset$ symbol. At every note there are at most $|A_{-i}|$ edges to go out of the state to the new state. Hence, it would generate at most $|A_{-i}|^{\tau-1}$ states, the first element of the history was taken by the initial state. Finally, at every state of the complete history, there are at most $|A_{-i}|$ outgoing edges, which can appear, because the first action profile of the history was fixed by the initial state. At this point, there is a state which corresponds to every achievable history. Hence, at every currently existing state the new signal should forward to one of the existing states, because the output symbol should correspond to the exact history leading to creation of this state.

□

Proof of Proposition 4 Formally further we show that the estimator satisfies even stronger property than consistency – it is unbiased in the limit.

$$\lim_{n\to\infty} E(\hat{p}(s(h)|h)) = Pr(s(h)|h) = P$$

The second part of the equality follows from the construction of the quantal response model.

Proof. Denote by $\hat{p}_n(a)$ the expected probability that action a is the most frequently observed action in the sample of the size n. Denote by $\hat{q}_n(a|h)$ the probability with which the strategy a is observed in the sample of the size n. Then, we can directly rewrite the expectation of the estimate as follows.

$$E(\hat{p}(s(h)|h)) = \sum_{a \in A} \hat{p}_n(a)\hat{q}_n(a|h) \leq \sum_{a \neq s(h)} \hat{p}_n(a) + q_n(s(h)|h)\hat{p}_n(s(h))$$

Note that we substited $\hat{q}_n(s(h)|h)$ with $q_n(s(h)|h)$ since if we consider a Bernoulli random variable of observing $s(h)$ or not, then $\hat{q}_n(s(h)|h)$ is an unbiased estimator of $q_n(s(h)|h)$.

$$\lim_{n\to\infty} E(\hat{p}(s(h)|h)) = \lim_{n\to\infty} \left(\sum_{a \in A} \hat{p}_n(a) \right) + \hat{q}_n(s(h)|h) \lim_{n\to\infty} \hat{p}_n(s(h))$$

Recall that $\hat{q}_n(s(h)|h)$ is a consistent estimator of $q_n(s(h)|h)$ and $\hat{p}_n(s(h)) \geq Pr(\hat{q}_n(s(h)|h) > .5)$. Hence, $\lim_{n\to\infty} \hat{p}_n(s(h)) = 1$. At the same time $0 \leq \sum_{a \neq s(h)} \hat{p}_n(a) \leq Pr(\hat{q}_n(s(h)|h) \leq .5)$.

Hence, $\lim_{n\to\infty} \left(\sum_{a \in A} \hat{p}_n(a) \right) = 0$. This completes the proof. □

References

Abreu, D. (1988): "On the theory of infinitely repeated games with discounting," *Econometrica: Journal of the Econometric Society*, 383–396.

Aumann, R. J. (1981): *Essays in game theory and mathematical economics in honor of Oskar Morgenstern*, vol. 4, Bibliographisches Institut.

Aumann, R. J. and S. Sorin (1989): "Cooperation and bounded recall," *Games and Economic Behavior*, 1, 5–39.

Axelrod, R. and W. D. Hamilton (1981): "The evolution of cooperation," *science*, 211, 1390–1396.

Barlo, M., G. Carmona, and H. Sabourian (2009): "Repeated games with one-memory," *Journal of Economic Theory*, 144, 312–336.

Binmore, K. G. and L. Samuelson (1992): "Evolutionary stability in repeated games played by finite automata," *Journal of economic theory*, 57, 278–305.

Brandts, J. and G. Charness (2011): "The strategy versus the direct-response method: a first survey of experimental comparisons," *Experimental Economics*, 14, 375–398.

Breitmoser, Y. (2015): "Cooperation, but no reciprocity: Individual strategies in the repeated Prisoner's Dilemma," *American Economic Review*, 105, 2882–2910.

Bruttel, L. and U. Kamecke (2012): "Infinity in the lab. How do people play repeated games?" *Theory and Decision*, 72, 205–219.

Chari, V. V. and P. J. Kehoe (1990): "Sustainable plans," *Journal of political economy*, 98, 783–802.

Dal Bó, P. and G. R. Fréchette (2015): "Strategy choice in the infinitely repeated prisoners' dilemma," .

Embrey, M., G. R. Fréchette, and S. Yuksel (2015): "Cooperation in the finitely repeated prisoner's dilemma," .

Engle-Warnick, J. and R. L. Slonim (2006): "Inferring repeated-game strategies from actions: evidence from trust game experiments," *Economic theory*, 28, 603–632.

Ephraim, Y. and N. Merhav (2002): "Hidden markov processes," *IEEE Transactions on information theory*, 48, 1518–1569.

Falk, A., E. Fehr, and U. Fischbacher (2005): "Driving forces behind informal sanctions," *Econometrica*, 73, 2017–2030.

Friedman, D. (1991): "Evolutionary games in economics," *Econometrica: Journal of the Econometric Society*, 637–666.

Fudenberg, D., D. G. Rand, and A. Dreber (2012): "Slow to anger and fast to forgive: Cooperation in an uncertain world," *The American Economic Review*, 102, 720–749.

Fundenberg, D. and E. Maskin (1990): "Evolution and cooperation in noisy repeated games," *The American Economic Review*, 80, 274–279.

Haltiwanger, J. and J. E. Harrington Jr (1991): "The impact of cyclical demand movements on collusive behavior," *The RAND Journal of Economics*, 89–106.

Hommes, C., J. Sonnemans, J. Tuinstra, and H. van de Velden (2005): "A strategy experiment in dynamic asset pricing," *Journal of Economic Dynamics and Control*, 29, 823–843.

Kalai, E. (1990): "Bounded rationality and strategic complexity in repeated games," in *Game theory and applications*, Elsevier, 131–157.

Kalai, E. and W. Stanford (1988): "Finite rationality and interpersonal complexity in repeated games," *Econometrica: Journal of the Econometric Society*, 397–410.

Kandori, M., G. J. Mailath, and R. Rob (1993): "Learning, mutation, and long run equilibria in games," *Econometrica: Journal of the Econometric Society*, 29–56.

Kocherlakota, N. R. (1996): "Implications of efficient risk sharing without commitment," *The Review of Economic Studies*, 63, 595–609.

Ligon, E., J. P. Thomas, and T. Worrall (2002): "Informal insurance arrangements with limited commitment: Theory and evidence from village economies," *The Review of Economic Studies*, 69, 209–244.

Linde, J., D. Gietl, J. Sonnemans, and J. Tuinstra (2016): "Information and Learning in the Minority Game: A Strategy Experiment," .

Mailath, G. J. and L. Samuelson (2006): *Repeated games and reputations: long-run relationships*, Oxford university press.

Muller, L., M. Sefton, R. Steinberg, and L. Vesterlund (2008): "Strategic behavior and learning in repeated voluntary contribution experiments," *Journal of Economic Behavior & Organization*, 67, 782–793.

Neyman, A. (1985): "Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma," *Economics letters*, 19, 227–229.

Osborne, M. J. and A. Rubinstein (1994): *A course in game theory*, MIT press.

Rotemberg, J. J. and G. Saloner (1986): "A supergame-theoretic model of price wars during booms," *The American Economic Review*, 76, 390–407.

Rubinstein, A. (1986): "Finite automata play the repeated prisoner's dilemma," *Journal of economic theory*, 39, 83–96.

Sabourian, H. (1998): "Repeated games with M-period bounded memory (pure strategies)," *Journal of Mathematical Economics*, 30, 1–35.

Selten, R. (1967): "Die strategiemethode zur erforschung des eingeschräncht rationalen verhaltens in rahmen eines oligopolexperiments', in beitraägge zur experimentellen wirtschaftforschung, ed. by h. sauermann, tübingen: Jcb mohr," .

Sonnemans, J., C. Hommes, J. Tuinstra, and H. van de Velden (2004): "The instability of a heterogeneous cobweb economy: a strategy experiment on expectation formation," *Journal of Economic Behavior & Organization*, 54, 453–481.