RECONSTRUCTING STRATEGIES IN DYNAMIC GAMES

ARTHUR DOLGOPOLOV

ICES, George Mason University

MIKHAIL FREER

Department of Economics, University of Essex

Abstract. The essential problem in the empirical analysis of the repeated games is to know what strategies are used by the players. We propose a simple algorithm to reconstruct strategies from the observed sequence of play. The algorithm accounts for the possibility of measurement and decision making errors. It stays agnostic about equilibrium restrictions. We apply the algorithm to both experimental and observational data. Using the experimental data, we provide conclusive evidence that players use strategies of memory no more than one period. Using the observational data, we support the hypothesis that Australian gas stations learn to tacitly collude for weekly price cycles using the day of the week as a coordination device.

1 Introduction

Repeated games are an integral part of economic science.[1] However, (empirical) analysis of the repeated games raises the foremost important question: what are the strategies used by players? Understanding which strategies real players use is vital for both accurate theoretical analysis of these games and policy-making. For instance, consider the repeated interaction within a market without perfect competition and a researcher/policy-maker who is interested in price dynamics. In order to consider the dynamic effects in the market, it is important to understand the dynamic response to a policy change. Therefore, a researcher needs to recover the strategies used by players on the market to study the possible counterfactual. Recovering the strategies of the

Date: October 11, 2021.

[1]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).[2] 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).

players is of crucial importance since commonly there is no unique way to escape the multiple equilibria problem.[3]

1.1 Contribution. This paper presents a simple and computationally efficient algorithm to reconstruct strategies from the observed sequence of play. The strategy prescribes an action to take in the current period, which depends on the "state of the world" (determined by the history of past interactions) and the actions of all other players (in the preceding period). The algorithm also accounts for the measurement and decision making errors and delivers a consistent estimator for the (partial) strategy. The algorithm only requires fixing the maximum allowed complexity (memory) of the strategy, ignores the equilibrium restrictions, and operates over the widest set of feasible strategies. However, given the (partial) strategies obtained, we can use ex-post procedures to test whether these strategies are consistent with rationality/equilibrium predictions. Besides, our algorithm allows for simple statistical tests to distinguish different models, in particular – to test what is the depth of the memory the strategy has, or what are the relevant state-determining parameters.

We illustrate the algorithm with applications to the experimental and observational data. The application to the observational data uses the daily gas-station prices in Australia for the seven years (the original data set has been collected by Byrne and De Roos, 2019). Using the data set, we reveal and confirm the fact that firms learn to coordinate the weekly price cycles on the day of the week. In addition, we show that throughout the period of observations, there is no need to consider the strategies of memory more than one.

For the experimental application, we conducted experiments with the infinitely repeated play of prisoner's dilemma/threshold provision of public goods (2 and 3 player

[3]An alternative is to make further assumptions to guarantee equilibrium selection. However, this approach is not universally applicable. There are four common refinement strategies: evolutionary arguments, complexity, efficiency, and subgame perfection. Some of these are too restrictive, the simple supporting strategies in the folk theorems are not claimed to correspond to actual behavior, and the equilibrium selection criteria are arbitrary and not well understood (Mailath and Samuelson, 2006). Importantly, neither do these refinements completely solve the problem of the multiplicity of equilibria. García and van Veelen (2016) show that an evolutionarily stable strategy may not exist; destabilizing invasions are always possible. They are not rare, and thus refinements depend on the evolutionary process itself even in a prisoner's dilemma. The folk theorems still apply for the refinements based on complexity, e.g., lexicographic preference for smaller complexity (Abreu and Rubinstein, 1988), for some games with two players and three actions (see the survey in Chatterjee and Sabourian (2009) and Bloise (1998)). Subgame perfection is known to not restrict the equilibria in three-person bargaining (See Chatterjee and Sabourian (2000, 2009), who point to the 1986 talk by Avner Shaked as the first proof). Finally, the Markov perfect equilibrium does not refine the equilibria enough in the asymmetric dynamic oligopolistic competition. It can introduce biases in counterfactuals if agents play other subgame-perfect equilibria — Salz and Vespa (2015) document this bias in the laboratory and simulations.

versions) and the Battle of Sexes. In both prisoner's dilemmas, we also found that subjects do not use the strategies of memory more than one. Moreover, the elicitation of strategies can help explain the decrease in cooperation that appears with the increasing number of players.[4] In particular, in 3-person prisoner's dilemma, all players use strategies that are initiated with defect (the first action at the beginning of the match), while in the 2-person prisoner's dilemma, there is a significant share of subjects who used the strategies initiated with cooperation (e.g., tit-for-tat or grim-trigger). Hence, while a simple reduction of the number of players who use always-defect strategy can solve the cooperation problem in the two-player version, it would not be as effective in the three-player version. Finally, in both games, we found evidence that subjects learn and converge to using pure strategies.[5] Our analysis for the Battle of Sexes shows significantly more noise, which suggests that players may be using mixed strategies. That said, there is no evidence that they play the Nash equilibrium mixed strategies. Only 2 out of 28 subjects could be mixing according to the mixed strategy Nash equilibrium.

Both illustrations justify the fact that our algorithm ignores the payoffs and equilibrium restrictions. In the application to gas pricing data, an estimate of the profits of a gas station requires an estimate of the costs of gas and the market demand. Hence, the payoffs are not observable in the pricing data itself. In the application to experimental data we can only enforce the monetary incentives, but the underlying preferences can be affected by social preferences or risk preferences.[6] Hence, even in the experimental test, we cannot guarantee that the true underlying payoffs of agents are known.

1.2 Related Literature There is a significant body of literature that investigates behavioral patterns in repeated games. However, closest to the paper are the studies concentrated on different methods for recovering the strategies. Two primary methods are commonly used for this purpose.[7] First is the strategy method; that is, a player is directly asked to list her complete strategy. The game is then played based on these reported strategies. The second approach tries to estimate the strategies directly either based on a mixed-type maximum likelihood estimator or other goodness of fit measures.

[4]This result is standard for the public good experiments (see e.g. Isaac and Walker, 1988; Isaac et al., 1994). Recall that prisoner's dilemma is essentially a threshold public good game, and we consider the cases of 2 and 3 players per group.

[5]However, they preserve some share of experimentation in the first period of every match, especially in the two-player version.

[6]Risk preferences play a role since we have to implement the infinitely repeated game via the probability of rematching.

[7]The paper is also related to the literature on estimating Hidden Markov Models (HMM). Even though HMM can be represented as an automaton, it is different from the model of the strategy we use. In particular, to adapt HMM to this setting, one would have to disentangle the state, and the input symbols since for canonical HMM state transitions depend only on the current state. More importantly, the HMM allows for stochastic transitions from each state to every other state. This idea corresponds to a mixed strategy rather than the pure strategy, which is the object of this study.

This paper is closer to the latter; however, we discuss both approaches to better explain the applicability and possible strengths and weaknesses of each.

The strategy method originated by Selten (1967) has been widely applied for experimental design. However, given the specifics of the repeated games, the strategy method can be implemented in different forms. Original advances have been taken by the seminal work of Axelrod and Hamilton (1981), who asked the game-theorists to submit the strategies (for infinitely repeated prisoner's dilemma) and Selten et al. (1997) who asked the experienced players to program their strategies in PASCAL (for finitely repeated Cournot duopoly game). Mengel and Peeters (2011) takes the strategy method to the voluntary contribution games, where strategy space is very rich. Dal Bó and Fréchette (2019); Romero and Rosokha (2018) employ a protocol in which they allow players to get familiarized with the game. Embrey et al. (2015) allows subjects to revisit the strategy at any moment of the game at small costs instead of extending the preparatory play. Cason and Mui (2019) allow players to choose out of the list of predetermined strategies.[8] However, the strategy method, even with controls for learning and possible decision-making errors, cannot be taken out of the lab.

The alternative approach requires estimation out of a preselected set of strategies. Several prominent results use the mixed-type likelihood estimators for this purpose (see e.g. Engle-Warnick and Slonim, 2006b; Dal Bó and Fréchette, 2011; Fudenberg et al., 2012; Breitmoser, 2015) as well as alternative goodness-of-fit measures and trade-offs (see e.g. Engle-Warnick and Slonim, 2006a; Camera et al., 2012). However, the common feature is that these methods require fixing the set of strategies ex-ante. Our algorithm does not require the selection of the set of strategies, but this comes at the cost of partial identification of pure strategies. Notably, the algorithm provided delivers maximum likelihood estimate over *all possible strategies*. The standard method would require an almost infinite data set and would also be computationally infeasible due to the size of the problem.[9] On the other hand, the (ideologically) closest approach is one

[8] In addition, the strategy method has been applied 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 the 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 indicates the absence of learning over rounds. Dal Bó and Fréchette (2019) elicit strategies of memory one in infinitely repeated prisoner's dilemma and compare the reports to econometric estimations, showing that the results are similar and that most strategies are of memory one. Additionally, for a survey and discussion of the strategy method versus direct elicitation, see Brandts and Charness (2011).

[9] Consider, for instance, a two-player, two-strategy stage game with players conditioning their actions on no more than five previous periods. The total amount of possible strategies is then $2^{28} > 10^8$. Even if the researcher obtains a sufficient amount of data, the mixed-type likelihood would be based on the optimization of a convex function with the 2^{28} parameters. Hence, the computational complexity

of Camera et al. (2012), who also estimates a pure strategy from noisy data. However, Camera et al. (2012) restricts the analysis to the set of simple strategies.[10]

Finally, our algorithm may be considered as more demanding in terms of data since we conduct the analysis at the individual level.[11] That is, while the standard mixed-type likelihood estimator would only bring restrictions on the total size of the data set, the algorithm we propose is demanding for the number of observations at the individual level. In particular, increasing the number of observations per subject would allow for better identification of the strategy used by the player.

1.3 Structure The remainder of the paper is organized as follows. Section 2 presents the theoretical framework and the algorithm itself. Section 3 shows the details of the application to the experimental data. Section 4 provides the details on the application to the observational data. Section 5 contains the concluding remarks. All proofs and main robustness checks are collected in the Appendix.

2 Theoretical Framework

This section consists of three parts. First, we present the basic notation for repeated games. Second, we consider the rationalization in the deterministic case (when players perfectly follow the strategy). Finally, we present a simple algorithm to reconstruct the strategy accounting for possible decision-making errors.

2.1 Basic Definitions We start by introducing the concept of a stage game. Let $N = \{1, \dots, n\}$ be the set of *players*. Let A_i be the set of *actions* for every $i \in N$. Let $A = \times_{i \in N} A_i$ be the set of *action profiles*, that is a vector of actions corresponding to each of the players. Denote by $\emptyset$ the empty action profile. That is an artificial construct that allows us to deal with the beginning of the game. A *stage game* $G = (N, A)$ is a tuple that consists of the sets of players and action profiles. Note that we disregard the payoffs.

Further, we define the repeated game; however, everything can be defined for any dynamic choice situation (with finite sets of histories and actions). A general definition would require further abuse of notation without providing extra insight. In particular, it would require a different set of players and actions in each of the stage games.

can become a problem. On the other hand, the algorithm we propose estimates the strategy with asymptotic complexity comparable to a single step of maximum likelihood evaluation.

[10]More precisely, they only consider automata with no more than two states. This restriction reduces the set of possible strategies to 32. In addition, they allow for mistakes to depend on the current state and the signal, while for the matter of illustration, we restrict the probability of mistake to be the same. It is possible to bring the results closer, and the same results would hold if we relaxed this assumption and allowed for heterogeneous probabilities of mistakes.

[11]While we recognize this limitation of the approach, it can be relaxed using Castillo and Freer (2018) "revealed differences" approach. Under weak assumptions, we can ensure that the recovered partial strategy is the consistent estimator of the true underlying strategy.

First, we need to define the histories and the relation between them. Denote the history of the length $\tau \in \mathbb{N}$, which is present at the period $t \in N$ by

$$h^{\tau}(t) = (a^{t-\tau}, \dots, a^{t-1}),$$

where $a^{t-i} \in A$ for every $1 \leq i \leq \tau$.[12] History is a sequence of action profiles observed between period $t-\tau$ and period $t-1$. Denote by H^{τ} the set of all possible histories of length τ.

Data set. We observe a finite sequence of play for T periods, with K rematches. That is a game with the previous opponent(s) finishes, and restarts with the new opponent(s), while the player restarts her strategy. Denote the length of k-th match by $T_k > 0$ such that $\sum_{k=1}^{K} T_k = T$. Let $D = \{(a^t)_{t=1}^{T}, (T_k)_{k=1}^{K}\}$ be a *data set*, that is a collection of action profiles at every period (a^t) and an identifier for the match at every period (T_k). This construction is sufficient to define histories of any length at any moment. There are two representations of a strategy used in the literature. Further, we introduce and discuss these definitions.

Strategies and Automata. A **strategy of memory** τ is a function that maps histories into actions:

$$s : H^{\tau} \to A_i$$

If this function is partial (determines actions only for a subset of histories), then this is a **partial strategy**. A data set is **rationalizable with a strategy of memory** τ if there is $s : H^{\tau} \to A_i$ such that $s(h^{\tau}(t)) = a_i^t$ for every $t = 1, \dots, T$. Hence, to rationalize the data, we need to find a strategy that can generate the observed sequence of actions.

The second concept is the automaton (or Moore machine). An **automaton** is a 4-tuple $\mu = (\Omega, \hat{\omega}, \mathcal{T}, G)$, where (1) $\Omega = \{\omega_0, \dots, \omega_{|\Omega|}\}$ is a set of states; (2) $\hat{\omega} \in \Omega$ is the initial state; (3) $\mathcal{T} : \Omega \times (A_{-i} \cup \{\emptyset\}) \to \Omega$ such that $T(w, \emptyset) = \hat{\omega}$ for every $\omega \in \Omega$ is a transition function; and (4) $G : \Omega \to A_i$ is the output function. Note that A_i and A_{-i} play the role of the output and input alphabet correspondingly. Two automata are said to be equivalent if they produce the same output symbol for any input sequence. Without loss of generality, we focus on the minimal automata because the minimal automaton is a unique characterization of a strategy.[13]

In order to define rationalizability with an automaton, we need to define the composition of the transition functions. Let $T^n(\omega, a_{-i}^t, a_{-i}^{t+1}, \dots, a_{-i}^{t+n-1})$ be the superposition

[12] Formally a history can be a more comprehensive set. That is, it only needs to be a vector of sets $A_i \times X$, where X is a finite set such that realizations of X are observed in every period.

[13] The uniqueness of the minimal automaton follows from the fact that every Mealy machine is unique (see, e.g. Linz, 2006) and that the spaces of Mealy and Moore (automata) machines are isomorphic.

of transition functions starting from the state $\omega \in \Omega$. More formally, we can introduce it via the following recursive definition.

$$T^1 = T(\omega, a_{-i}^t) \text{ and } T^{i+1} = (T^i(\omega, a_{-i}^t, \ldots, a_{-i}^{t+i-1}), a_{-i}^{t+i}).$$

Recall that a transition function maps states and action profiles back into the set of states. Hence, every superposition of the transition functions would return a well-defined state. This construct is an analog of the walk of length n along the graph representing the automaton. A data set is **rationalizable with an automaton** if there is $\mu = (\Omega, \hat{\omega}, \mathcal{T}, G)$ such that $G(T^t(\hat{\omega}, a_{-i}^1, \ldots, a_{-i}^t)) = a_i^{t+1}$ for every $t \in \{1, \ldots, T-1\}$.

From Kalai and Stanford (1988) we know that the space of strategies is equal to the space of automata. However, this result relies on the strategies of infinite memory. That is, a player may use a strategy that conditions on all past actions disregarding the length of play. Moreover, this result establishes equivalence in terms of representation, not in terms of rationalization. At the same time, this paper is concerned with extracting a strategy from a finite data set. Hence, we are concerned with equivalence in terms of the rationalization of the data set.

2.2 Deterministic case We start by stating the conditions for rationalizability. Next, we provide a connection between the representations in terms of rationalizability. Throughout this subsection, we consider the deterministic case. That is, we consider the data set without decision making or measurement errors. Finally, we extend the results to reconstructing the strategy in the stochastic case.

Lemma 1. *Let $D = \{(a^t)_{t=1}^T, (T_k)_{k=1}^K\}$ be a data set.*

- *It is rationalizable with a strategy of memory τ if and only if*

$$h^\tau(t) = h^\tau(v) \textit{ implies } a_i^t = a_i^v$$

for every $t, v \in T$.
- *It is rationalizable with an automaton with the set of states Ω if and only if there is $\xi : \{1, \ldots, T\} \to \Omega$ such that*
  (1) $\xi(1) = \xi(T_k + 1)$,
  (2) $\xi(t) = \xi(v)$ implies $a_i^t = a_i^v$, and
  (3) $\xi(t) = \xi(v)$ and $a_{-i}^t = a_{-i}^v$ implies $\xi(t+1) = \xi(v+1)$
  for every $k \in \{1, \ldots, K\}$ and $t, v \in \{1, \ldots, T\}$.

Lemma 1 provides the necessary and sufficient conditions for rationalization in terms of both strategies and automata. The conditions for rationalization with a strategy are particularly simple. They require only that the data contains the same output symbol for the same history at every occurrence of that history. At the same time, the conditions for rationalization with an automaton specify a mixed-integer program, which is more complicated but still proves to be computationally feasible.[14] Further, we establish a stronger result on the observational equivalence of the representations. This result also allows us to construct the automaton more efficiently. We start with one of the implications of the Lemma 1.[15]

Remark 1. *A data set* $D = \{(a^t)_{t=1}^T, (T_k)_{k=1}^K\}$ *is rationalizable with a strategy if and only if it is rationalizable with a strategy of memory* $\bar{T} = \max_{k \leq K} T_k$.

Remark 1 shows that the infinite memory strategies have no empirical content and also provide the explicit bound on the maximal memory of the strategy that would have empirical content given the data set. This result also allows us to construct an efficient algorithm to find the minimal complexity of the strategy that rationalizes the data set. That is, we do not need to test the rationalizability with memory more than $\bar{T}$ and less than 1. Using these bounds, one can use, for instance, a binary search tree to find the minimal memory necessary to rationalize the data set. Since the memory can only be an integer, the algorithm converges in finite time. Finally, using this result, we can establish the observational equivalence between the automata and strategies of finite memory.

Proposition 1. *A data set is rationalizable with a strategy of finite memory if and only if it is rationalizable with an automaton.*

Proposition 1 shows that the space of strategies of finite memory is observationally equivalent to the space of automata. Kalai and Stanford (1988) showed that, in general, the space of automata is a strict super-set of the space of the strategies of finite memory. our result shows that there is a particular class of automata that cannot be distinguished from the automata corresponding to the finite memory strategies. This set of automata equivalent to the strategies of finite memory is the set of automata without counters (see Schützenberger, 1965; McNaughton and Papert, 1971). A counter is a non-trivial sequence of states $q_0, q_1 ... q_n$ and a sequence of action profiles (a "word") w, s.t. $q_0 \xrightarrow{w} q_1 \xrightarrow{w} q_2 \xrightarrow{w} ... \xrightarrow{w} q_n \xrightarrow{w} q_0$, where each arrow is a walk with a sequence of play w. In other words, a counter is a set of several states connected through a chain of transitions, so that the resulting sequence of player's and opponent's actions is the same in all these chains between all states, forming a periodic cycle.[16]

[14]Conditions (1)-(3) determine the inequalities, necessary for the mixed-integer program. The total number of states can be used as the objective function since we only seek rationalization with a minimal automaton.

[15]Note that if $K = 1$, then there is no empirical content beyond any of rationalizations. That can also be seen from the proof of Remark 1.

[16]Counters are usually defined for finite-state acceptors (automata that accept or reject input strings, but produce no other output), while in economics an automaton usually means a finite-state transducer (an output for every state). All representation results, however, are general and apply to semiautomata – automata with no output whatsoever. We can maintain the link between the two by constructing

This result states the limitations of the approach, stemming from the fact that infinite memory strategies cannot be well-tested with finite data. On the positive side, this result guides the researcher to include possible relevant counters since the algorithm cannot efficiently recover them from the data. We illustrate this point in the empirical application presented in Section 4. Below we refine this result and link the complexity of automata (number of states) to the complexity of strategy (depth of memory). This refinement will allow for a characterization of the set of automata equivalent to the strategy of memory τ. Also, this result improves the efficiency of the algorithm for finding the automaton that rationalizes the data, since the minimal depth of memory would provide an upper bound on the number of states that the corresponding automaton may have.

Definition 1. *An automaton μ is τ* **transition invariant** *if*

$$G(T^\tau(\omega, a_{-i}^1, \ldots, a_{-i}^\tau)) = G(T^\tau(\omega', a_{-i}^1, \ldots, a_{-i}^\tau))$$

for all $\omega, \omega' \in \Omega$ and $a_{-i}^1, \ldots, a_{-i}^\tau \in A_{-i} \cup \{\emptyset\}$ such that $G(T^t(\omega, a_{-i}^1, \ldots, a_{-i}^t)) = G(T^t(\omega', a_{-i}^1, \ldots, a_{-i}^t))$ for all $t < \tau$.

Transition invariance states that if, starting from any two states with the same output symbol, we were to feed the automaton with the same sequence of actions of length τ, the output symbols should be the same. Note that once an automaton is τ transition invariant, then it is transition invariant for every $\tau' \geq \tau$.

Figure 1 illustrates an automaton, which is 2 transition invariant but not 1 transition invariant. Consider two sequences of actions starting from w_2 and w_3, and assume that $a_{-i} = C$ in both cases. Given that $G(w_2) = G(w_3) = D$, the 1 transition implies that the output symbols should be the same. At the same time, we know that $D = G(T(w_2, C)) \neq G(T(w_3, C)) = C$, which immediately contradicts the condition. One can easily verify that the automaton is 2 transition invariant.

Proposition 2. *A data set is rationalizable with a strategy of memory τ if and only if it is rationalizable with a τ transition invariant automaton.*[17]

Proposition 2 shows that τ transition invariant automata are observationally equivalent to rationalization with a strategy of memory τ. The proof of this proposition is constructive in both ways and allows us to construct the minimal automaton corresponding to a given strategy.

a finite state acceptor from the Moore machine - assume that all states are final and add the output symbol of every sate to all transitions to that state. This automaton would accept only the sequences of play that could be produced by the strategy of the Moore machine, and this allows for the application of Schützenberger (1965); McNaughton and Papert (1971) result directly.

[17]As an immediate observation that comes from this proposition, the set of τ transition invariant automata (for every finite τ) is equal to the set of automata without counters.

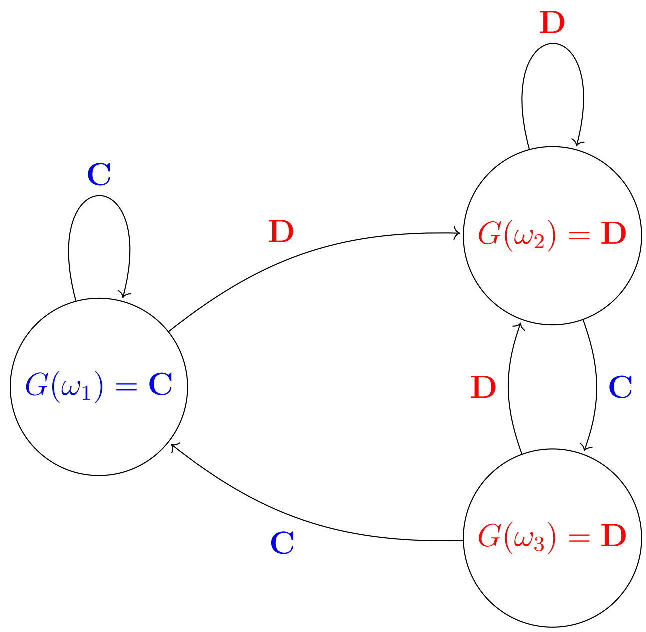


FIGURE 1. $\tau = 2$ transition invariant automaton

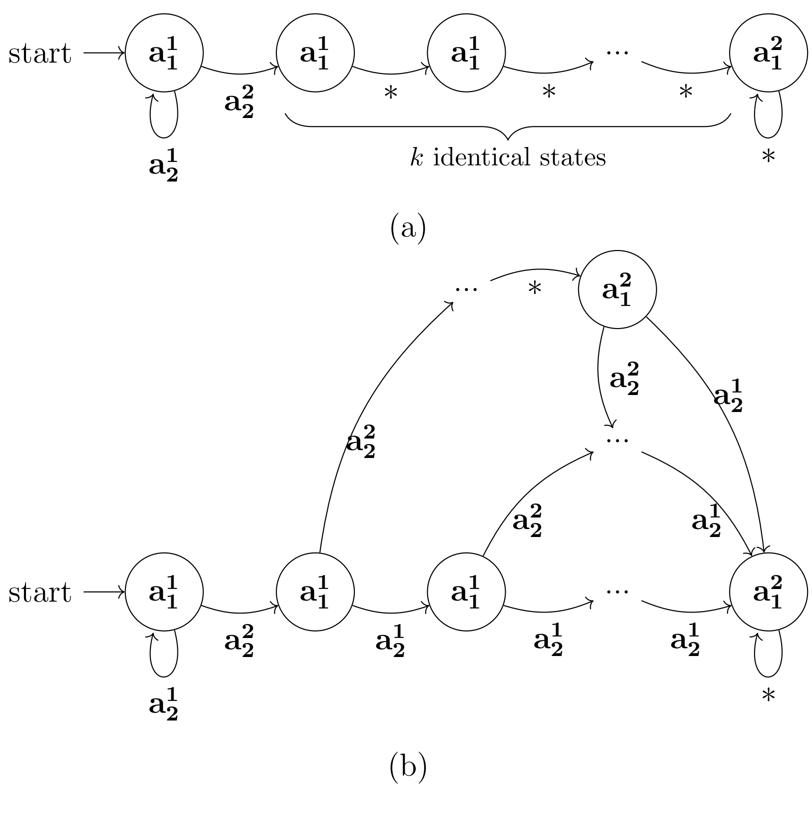


FIGURE 2. Moore machines with arbitrarily large but finite memory $\tau = k + 1$

Figure 2 presents two automata, which both are $\tau = k + 1$ transition invariant, but have a very different number of states. We use $*$ as a symbol for any symbol. In particular, the automaton on the panel (a) has just $k + 1$ states, and the one on the panel (b) has $1 + k + \frac{1}{2}(k-2)(k-1)$ states. However, we still can provide the bound on the number of states of τ transition invariant automaton. This bound can be directly inferred from the proof of Proposition 2.

Remark 2. *If $|A_i| \leq |A_{-i}|$ and $\mu = (\Omega, \hat{\omega}, \mathcal{T}, G)$ is a minimal τ transition invariant automaton, then $|\Omega| \leq |A_{-i}|^{\tau} + 1$.*

2.3 Stochastic case Given Propositions 1 and 2 we can focus on the estimation of the strategy. We employ a simple model of random decision making. We assume that at every period, the player takes action prescribed by the strategy ($s(h) \in A_i$) with probability P and randomly chooses any other action with probability $1 - P$ (see Figure 3). We also assume that $P > .5$ to guarantee partial identification of the strategy.

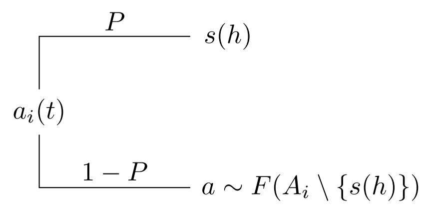


FIGURE 3. Data generating process

Before we proceed further, let us briefly discuss the assumptions behind the model. The first assumption is that the probability of using a strategy is independent of history. Second, we do not assume any distribution over the non-strategy actions (the random error) and keep it unknown. However, this comes at a price of assuming that the support of this error distribution does not include $s(h)$. While the first assumption can be relaxed without changing the results below, relaxing the second assumption would cause a non-identification problem. Instead, the second assumption can be substituted with the assumption about the parametric family of distributions over A_i that determines the random component. Although results for the estimates of $P(h)$ would depend on this assumption, neither the estimate of the strategy nor its consistency depends on these assumptions. Hence, if one is only concerned with recovering strategies, these assumptions can be disregarded under the assumption that $P(h) > .5$ for every history $h \in H^{\tau}$.[18]

Consider estimating a multi-dimensional parameter $\theta = \{P, \sigma\}$, where σ corresponds to the partial strategy, and P is the probability of using the partial strategy. Denote by $\hat{H}^{\tau} \subseteq H^{\tau}$ the set of observed histories of memory τ. Let $\sigma : \hat{H}^{\tau} \to \Delta A_i$ be a partial strategy. That is σ represents the probability of playing every action $a \in A_i$

[18]The result with formal definitions and a discussion are presented in Supplementary Materials. We also relax the assumption that errors are identically distributed. However, we need to require that not only the total amount of observations but the number of observations for every observed history converges to infinity.

after any given history $h \in \hat{H}^\tau$. However, we assume that the action is drawn from this distribution only once, and at every point, with history h, the subject will take the same decision. This assumption is a technical one to maintain a correspondence between the solution and the pure strategy. Hence, we do not capture the idea of a mixed or behavioral strategy. As mentioned above, once we allow for mixing in this framework, it becomes impossible to distinguish between the mixing and the mistakes. We denote an element of $\hat{\sigma}$ as $\hat{\sigma}_a(h) \in [0, 1]$, that is the probability that the agent will take action $a \in A_i$ whenever he encounters history h.

Denote by $\hat{\sigma}_a(h)$ the observed frequency of choosing action $a \in A_i$ conditionally on observing a history $h \in \hat{H}^\tau$. Denote by

$$\hat{\sigma}(h) = \operatorname*{argmax}_{a \in A_i} \hat{\sigma}_a(h)$$

the most frequently observed action given history $h \in \hat{H}^\tau$. This way, we can define an estimate of the partial strategy, $\hat{\sigma}$. Denote by

$$\hat{p}(\hat{\sigma}|h) = \max_{a \in A_i} \hat{\sigma}_a(h)$$

the frequency of observing most frequent action given history h.[19] Denote by $\hat{q}(h)$ the frequency of observing history $h \in \hat{H}^\tau$. Let

$$\hat{P} = \sum_{h \in \hat{H}^\tau} \hat{q}(h)\hat{p}(\hat{\sigma}|h)$$

that is the aggregated frequency of observing the chosen partial strategy, which in turn is an estimate of P.

Proposition 3. *Let $P > .5$, then $\hat{\theta} = \{\hat{P}, \hat{\sigma}\}$ is a consistent (and asymptotically efficient) estimator of θ with*

- *$\hat{\sigma}$ capturing the partial strategy, and*
- *$\hat{P}$ capturing the probability of using the partial strategy.*

Proposition 3 shows that the probability of incorrectly recovering the strategy converges to zero as the number of observations converges to infinity. It also shows that the estimator recovers the true value of P in the limit as well. Before we proceed further, let us make several important remarks about the nature of the estimator.

Remark 3. *$\hat{\theta} = \{\hat{P}, \hat{\sigma}\}$ is a regular maximum likelihood estimator.*

Remark 3 implies that there is a likelihood function that delivers the estimate. This observation is important because it provides us with a useful tool for empirical analysis. Recall that since we need to fix the memory of the strategy for the estimate, it may be necessary to compare different models (different memory depths). Hence, Remark 3 allows us to use the (standard) likelihood ratio tests for this purpose.

[19] If we allow for heterogeneous errors, then $\hat{p}(\hat{\sigma}|h)$ is a consistent estimate of the probability of error given history $h \in \hat{H}^\tau$.

Remark 4. *If $\hat{H}^\tau = H^\tau$, then the minimal automaton $\hat{\mu}$ obtained from $\hat{\sigma}$ is a consistent estimate of the true underlying automaton (μ).*

Remark 4 shows that if we observe all histories, then it is possible (in the limit) to recover the true underlying automaton. Moreover, if we observe all histories, then we can use the Moore algorithm to minimize the automaton, which is even more efficient than the mixed-integer programming.[20] We can still relax this assumption that requires to observe all the histories that can be reached from the initial state. Note that history also takes into account the actions of the player herself. Hence, some histories may not be reachable (in deterministic case) given a strategy player uses. However, the relevant part of automaton consists only of these histories. Assume that a player uses the unconditional defect strategy in prisoner's dilemma, then a set of reachable histories consists of $\{\emptyset, (D, C), (D, D)\}$. As long as they are observed, the corresponding automaton is *well-defined* and is a consistent estimate of the underlying automaton.[21]

If instead, we only observe a strict subset of the reachable histories, then we have to make decisions for the unobserved elements in the transition function to complete the automaton. Otherwise, the automaton is not well-defined, and the result would not generally have to be the underlying automaton. Note that this problem is not unique to our algorithm. For instance, the standard mixed-type likelihood would fall in the same identification problem if some of the relevant histories are not observed. Assume we only observe cooperative outcomes in a prisoner's dilemma. Then, the mixed type likelihood would not be able to identify what share of players is using unconditional cooperation and what share is using a grim-trigger, since, for this set of histories, the strategies are observationally equivalent. It is at least necessary to observe the history at which the counterpart of the player defects (by mistake), to say whether the player is using unconditional cooperation or not. Hence, the algorithm we propose still provides a well-identified partial strategy that can be used for counterfactual simulations and policy analysis.

In the experiments presented in Section 3, we on average observe 96.1% (78.4%) of histories of memory 1 and 67.5% (33.2%) of histories of memory 2 in the 2-player (3-player) PD. For Battle of Sexes we observe all histories of memory 1 and 84.1% of histories of memory 2.

3 Experimental Application

We conducted experiments at George Mason University to illustrate our test with three different games/treatments: (1) 2-person Prisoner's Dilemma (2-PD); (2) 3-person Prisoner's Dilemma (3-PD); and (3) Battle of Sexes (BoS) games. Matrix form games

[20] While mixed-integer programming can be NP in the worst case, the Moore algorithm is not just polynomial; it is quadratic in the worst case.

[21] See Supplementary Materials for more details.

are presented in Figure 4. There is a total of 103 participants with 36, 39, 28 participants for 2-PD, 3-PD, and BoS treatments correspondingly. The average payment for the subject was 18.75 US dollars.

Player 1 \ Player 2	A	B
A	40,40	10,60
B	60,10	25,25

(a) Prisoner's Dilemma (2-PD)

Player 1 \ Player 2	A	B
A	100,10	0,0
B	0,0	10,60

(b) Battle of Sexes (BoS)

	Player 3: A		Player 3: B	
Player 1 \ Player 2	A	B	A	B
A	40,40,40	10,60,10	10,10,60	10,50,50
B	60,10,10	50,50,10	50,10,50	25,25,25

(c) Three players Prisoner's Dilemma (3-PD)

FIGURE 4. Payoff matrices

To ensure the indefinitely repeated nature of the interaction, we use the following protocol. The entire experimental session lasts for 160 periods. After the end of every stage game with .8 probability subjects stay in the same match and with .2 probability groups are reshuffled (each subject gets a new random match). Rematching is done at the session level; that is, all groups in the same session are getting reshuffled. At the beginning of each period, subjects are told, whether they are matched to the same person as in the previous round or they got a new partner.[22] The average duration of the match is 5.08 rounds. Note that possible violation of the infinitely repetitive nature of the game comes from the last match since the total length of the game is limited (although the current round number was not shown). Hence, a natural robustness check is to remove the last match since it can be treated as a finitely repeated part of the

[22]We do not provide the subjects with the history of interactions, which could bias down the memory of the strategies.

experiment. All results presented below are robust to the elimination of the last match (see Appendix B.1.2).

Memory		2-PD	3-PD	BoS
$\tau = 1$	Mean	.903	.935	.771
	min Q_{1-3} max	.76 .84 .89 .97 1	.73 .89 .96 .99 1	.60 .70 .77 .85 .96
$\tau = 2$	Mean	.915	.944	.833
	min Q_{1-3} max	.77 .88 .92 .97 1	.76 .92 .96 .99 1	.72 .78 .84 .89 .96
$\tau = 3$	Mean	.924	.949	.859
	min Q_{1-3} max	.81 .89 .92 .98 1	.77 .93 .97 .99 1	.77 .82 .87 .89 .96
$\tau = 4$	Mean	.928	.952	.878
	min Q_{1-3} max	.83 .89 .93 .98 1	.79 .93 .97 .99 1	.79 .84 .88 .92 .97

TABLE 1. Mean and quartiles for $\hat{P}$ by game and memory depth

3.1 Complexity of strategies Table 1 presents the mean estimated $\hat{P}$ and the quartiles of the distribution of individual estimates. Each column corresponds to one treatment. Each row-block corresponds to the memory from $\tau = 1$ to $\tau = 4$. Recall that the higher is $\hat{P}$, the better the reconstructed strategy explains the data. Levels of $\hat{P}$ are comparable between these treatments. Note also that $\hat{P}$ increases at a decreasing pace. Estimates for the BoS are below the estimates for 2-PD ad 3-PD. Note that the deeper memory we allow, the more permissive is the test. Once we increase τ, there are fewer observations per history, while the observations for the same history are the source of variation in $\hat{P}$. Hence, to draw the further inference, we use the fact that estimate for $\hat{P}$ is a regular maximum likelihood estimator. Moreover, the model with a shorter memory is nested within the model with deeper memory. Hence, we can use the likelihood ratio test to compare the performance of the models.

However, the likelihood ratio test is known to be conservative (under-rejecting the null hypothesis). Therefore, selecting the standard thresholds for the p-value may be misleading. To construct a meaningful p-values, we use the following procedure. Consider an example of comparing the theories of whether a subject used a strategy of memory two or memory one. Consider a strategy of memory two (which cannot be represented as a strategy of memory one) with the $\hat{P}$ equal to the estimate for the memory of $\tau = 2$. Using the observed sequence of actions made by other players with whom this subject played in the session, we generate a data set of such strategies and the corresponding maximum likelihood. We can then compute the log-likelihood ratio statistic (asymptotically distributed as χ^2) and the corresponding p-value. Repeating this for each possible realization of each possible strategy of memory two, we obtain the distribution of p-values. Using the corresponding percentile of this distribution, we can obtain the threshold for hypothesis testing. We refer to this threshold as the power-corrected threshold.

	2-PD				3-PD				BoS			
	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$
$\tau = 1$		89%	100%	100%		100%	100%	100%		57%	93%	100%
$\tau = 2$			100%	100%			100%	100%			100%	100%
$\tau = 3$				100%				100%				100%

TABLE 2. Share of subjects for whom the restricted model (smaller τ) can not be rejected in favor of the unrestricted (larger τ) model for different depths of memory.

Table 2 presents the results of the likelihood ratio test using the power-corrected p-value of 10%. That is the share of subjects for whom the restricted model (smaller τ) can not be rejected in favor of the unrestricted (larger τ) model.[23] For 2-PD, we can reject the $\tau = 1$ model only for 11% of the sample. For 3-PD, this hypothesis cannot be rejected for any of the subjects. For BoS, the hypothesis that subjects use a strategy of memory one can be rejected for 43% percent of the sample. For 7% of the sample, the hypothesis that subjects use a strategy of memory two can also be rejected. Hence, while for 2-PD and 3-PD treatments, we get conclusive evidence that subjects do not use the strategy of memory more than one, the same conclusion cannot be made for the BoS.

3.2 Learning Since the share of random decisions is nontrivial, it is important to look at the distribution of errors. Dynamic patterns of errors can shed some light on the sources of errors. For instance, it can come from learning, experimentation (in order to get out of the "bad" equilibrium), or simply using a mixed strategy. Recall that mixed strategies cannot be successfully estimated using our method. Hence, this exercise can help us verify the method. [24]

Figure 5 presents the heat-map for the distribution of errors for the "best" estimated complexity of the strategy. The warmer is the color (from deep blue to sunny yellow); the higher is the frequency of mistakes. Different matches (from first to last) lie on

[23] We only consider τ up to 4 because it is the maximum level for which we can get meaningful numbers. The test statistic (log-likelihood ratio) is asymptotically distributed as χ^2. The number of degrees of freedom is determined by the difference in the dimensionality of the parameter space. That is the difference in the number of constraints we impose on the variables. This number is equal to the difference between the number of free parameters (histories). The number of histories grows exponentially; hence, for comparing the memory 4 and 5 strategies, we need to use χ^2 with about 1000 degrees of freedom. Hence, the hypothesis becomes non-rejectable.

[24] Recall that another assumption we make is i.i.d errors. This assumption is necessary for the log-likelihood test. Evidence of learning thus justifies additional robustness checks reported in Supplementary Materials.

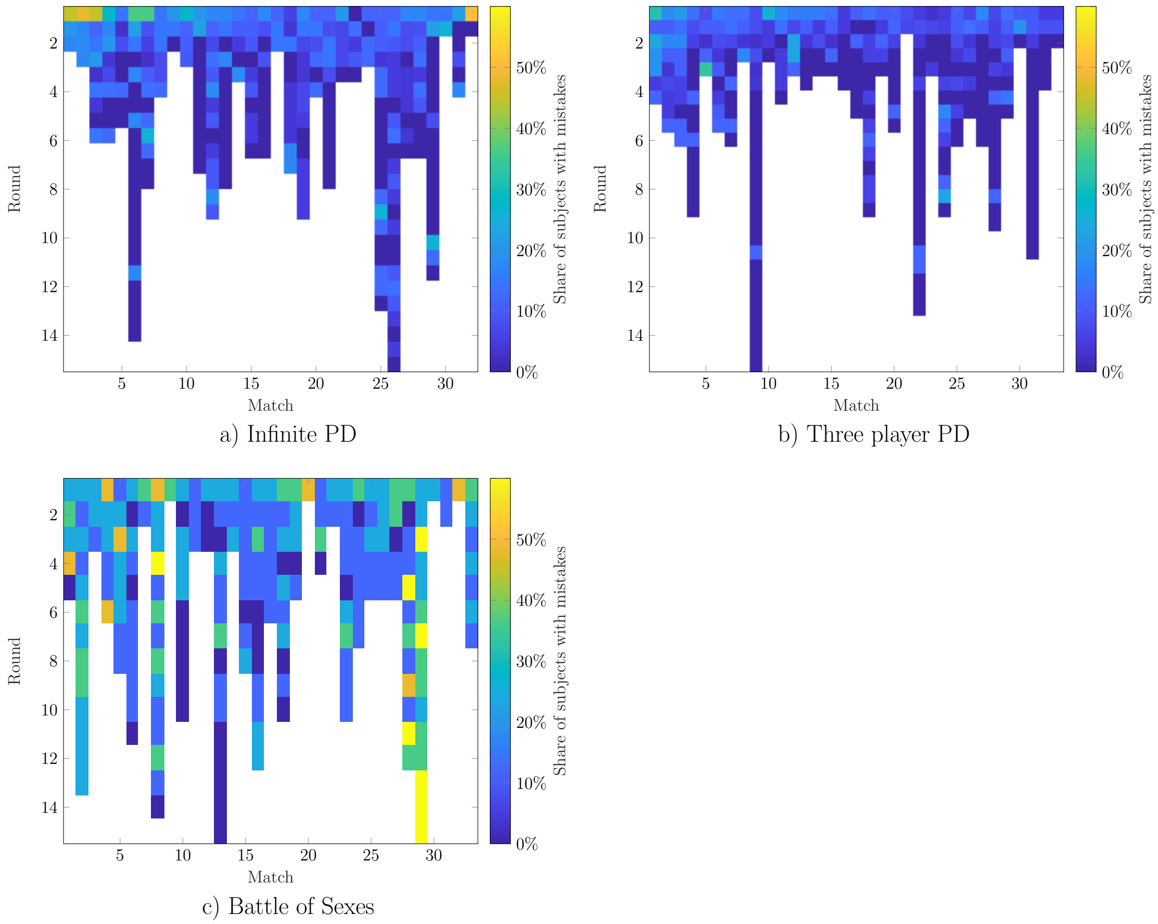


FIGURE 5. Frequency of deviations from strategy by round and match

the horizontal axis. Rounds within the match lie on the vertical axis. Moreover, the maximum length differs across matches. The white areas represent the parts for which there is no data in a given match.

In 2-PD and 3-PD games, the errors are concentrated at the beginning of the game (top left corner). That is, the subjects are experimenting with the strategy in the first periods of the first matches. The frequency of mistakes decreases with both the match number and the period number. Hence, subjects stabilize the strategy over time, and early mistakes can be attributed to learning. We also check the direction of the deviations, that is — what is the frequency of using cooperate when strategy prescribes using defect and vice versa. The frequency of both directions of the deviations is comparable at the beginning of the game for 2-PD. However, the later it comes (later matches and later periods), cooperating when strategy prescribes defecting becomes more frequent. In BoS, errors are more widespread across the experiment. There is no explicit learning between or within each match. Finally, the presence of errors is larger compared to the frequency of errors in the PD treatments.

The mean errors across matches exhibit a trend only in 3-PD treatment according to the Mann-Kendall test and χ^2-tests do not show a significant difference between the first

	2-PD	3-PD	BoS
Mann-Kendall test (p-value)	0.18	0.009	0.31
χ^2-test, 1 match (p-value)	0.84	0.92	0.97
χ^2-test, 5 matches (p-value)	0.99	0.97	0.69
Individual tests			
χ^2-test, 5 matches (num. passed at 0.05)	13/36	11/39	5/28
odds ratio, mean	1.28	1.26	1.24
odds ratio, min	1.11	0.70	0.60
odds ratio, max	1.71	1.60	2.03

TABLE 3. Persistence and dynamics of errors

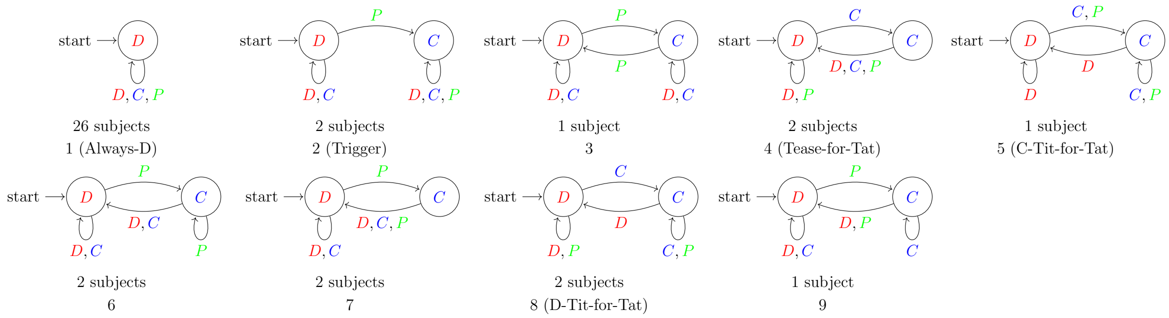
and the last matches. Individual tests of the first 5 rounds against the last 5 rounds are significant for about a third of subjects in PD treatments and 17% in BoS. The mean odds ratio among these subjects is reported in Table 3. Taking action prescribed by the strategy is x times more likely in the last 5 rounds than in the first 5 rounds, where x is the odds ratio.

3.3 Individual strategies Finally, we take a closer look at the strategies used by the players. We restrict our analysis to the 2-PD and 3-PD treatment since there is sufficient evidence that subjects do not persistently use the same strategy throughout the game in BoS treatment. Analysis for BoS is presented in Appendix B.1.3.

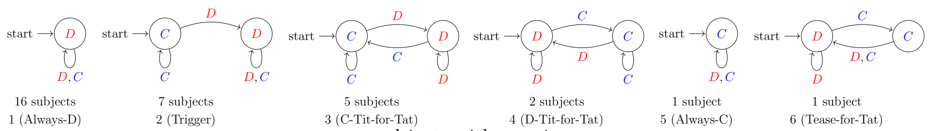
Figure 6 presents the strategies used in both 2-PD and 3-PD treatments. We denote by P the partial cooperation profile, that is (C, D) or (D, C) since we only consider symmetric strategies. That is, we do not distinguish which of the agents has taken each of the actions (symmetric strategies). The reason behind this is that symmetric strategies perform at least as good as asymmetric strategies (see Appendix B.1.1 for more details). The top panel presents strategies used by subjects in 2-PD, and bottom one presents the strategies used by subjects in 3-PD.

In 3-PD and 2-PD treatments, there is a significant share of subjects who unconditionally defect 16 (44%) and 26 (67%) subjects. The share of tit-for-tat strategies also smaller in the 3-PD – 3 subjects (8%) in 3-PD vs. 7 subjects (19%) in 2-PD.[25] Another interesting strategy used is tease-for-tat. Tease-for-tat prescribes the subject to switch to the cooperation upon observing cooperation, but to stay in the cooperative state for one period only, regardless of the actions of the other players. This strategy can be used to implement "taking turns" split of the rewards with tit-for-tatters and is also resilient against the defectors. Among the $\tau = 2$ strategies, the most interesting, 2C-2D

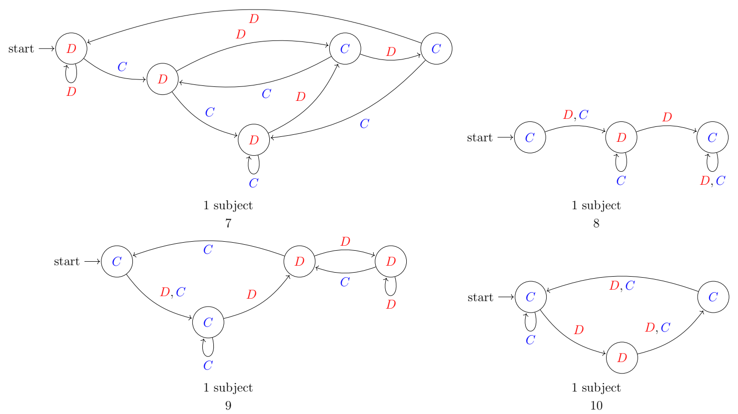
[25] We say that a 3-PD strategy is tit-for-tat as if it prescribes cooperating after both opponents cooperate and defecting after both opponents defect. Moreover, if there is a switch from defection to cooperation at partial cooperation, then there should also be a switch at full cooperation.



(a) 3-person PD.



subjects with $\tau = 1$



subjects with $\tau = 2$

(b) 2-person PD.

FIGURE 6. Revealed strategies for Prisoner's Dilemmas

strategy, has two main differences from standard tit-for-tat: (1) subject defects until she observes cooperation twice in a row, and (2) she has a buffer cooperation state (which is also initial). Having a buffer cooperation state is important to overcome the possible mismatch with tit-for-taters, which can appear due to noise or the fact that counterpart is using D-tit-for-tat. Finally, note that while there is almost an equal split between C and D initial state in 2-PD, in 3-PD *all subjects* use strategies with initial state having the action D. This feature of individual strategy helps explain the significant decrease in cooperation with the increased number of subjects.

To compare the performance of strategies, we conduct an evolutionary tournament and random pairwise comparisons of the strategies. We follow the Moran process protocol (Moran, 1958; Nowak et al., 2004) without mutation. The population starts with an observed population of automata and evolves with individuals chosen for reproduction proportionally to their round-robin payoff score and chosen for death uniformly at random. We use both noisy and pure versions of the strategies to show to which extent noise affects the results of the tournament. We summarize the results here and provide detailed results in Appendix B.1.4.

	2-PD		3-PD	
Strategy	Noisy	Pure	Noisy	Pure
Defect	79.7%	69.9%	91.2%	–
Grim Trigger	5.5%	7.3%	–	–
Tit-for-tat	10.1%	13.6%	2.8%	–
Tease-for-tat	2.0%	3.8%	3.3%	–

Table 4. Probability of winning the evolutionary tournament by the type of strategy.

Table 4 summarizes the results of the evolutionary tournaments by the main groups of strategies for both PD treatments. The evolutionary tournament is stochastic. Hence, instead of a deterministic winner, there is a distribution of winners (last survivor). First, the unconditional defect is the most likely winner given the population of strategies we have, and the presence of noise only increases the chances of defectors to be winners. In 2-PD, the second most-likely winner is tit-for-tat (mainly D-TFT for noisy and C-TFT for the pure tournament). In 3-PD, the second most-likely winner is tease-for-tat since the equilibrium is much less cooperative in that case. Pure strategy tournament for 3-PD does not deliver any information given that all strategies start with the defect and only switch states upon observing at least partial cooperation. Hence, all strategies would receive equal payoff (corresponding to (D, D, D)). Hence, the probability that a strategy is a winner is determined by the original share of each strategy and the realization of random selection for death. Note it is an important difference since while in the 2-PD evolutionary process can result in cooperative equilibrium, in the 3-PD evolutionary process would always deliver the non-cooperative equilibrium. Finally, note that among the $\tau = 2$ strategies, only 7 has a non-trivial probability of winning: 1.9% for noisy and 3.8% for pure tournaments.

We also conduct pairwise tournaments. In this case, we focus on the less standard strategies, since the results for standard strategies in 2-PD are well-known (see, e.g. Axelrod and Hamilton, 1981). First, we discuss the performance of the $\tau = 2$ strategies in 2-PD. Strategies 8, 9, and 10 manage to maintain cooperation with tit-for-tats and grim-triggers but are particularly vulnerable in front of defectors. Strategy 7 is more risk-averse (since starts with defect) and is well-protected from the defectors, but only manage to maintain partial cooperation with tit-for-tats, which initialize with C. In 3-PD without noise, all strategies end-up with the equal (non-cooperative) payoff. When we introduce noise, the changes are minor, and the unconditional defect strategy is still dominant and performs better than any other strategy against any counterpart.

4 Empirical Application

We use the data from Australian gas stations collected by Byrne and De Roos (2019) from Perth county in (western) Australia. The data we use comes from a price transparency program called Fuelwatch. The program (Introduced in January 2001) requires all gasoline retailers to submit their station prices for the following day, every day before 2 pm. Prices submitted have to match the prices posted by the gas station the next day (24 hrs starting from 6 am of the next day). Prices are published 2:30 pm each day. Very high compliance with the program is enforced in part through possible fines and prosecution, and prices are posted online for all individual stations.

There are six major companies in the region owning 661 stations. Four of them being major oil companies, and the other two correspond to the stations owned by supermarket chains. The biggest of them is British Petroleum (BP), which controls about 22% of the market. Small independent stations account for 32% of the market. The market shares stay stable between 2005 and 2015. Finally, let us note that the system is no particularly efficient among consumers. Only about 10-15% of consumers use the Fuelwatch website.

We use the data collected between January 2008 and December 2014. We consider four time blocks: (I) before April 2009; (II) April 2009 – January 2010; (III) January 2010 – July 2012 (IV) after July 2012 following the original paper. Moreover, these periods are interesting given the hypothesis, which can be derived from the analysis in the original paper. Period (I) is the restart of the game after the recession with no clear pattern being present; in period (II) the aggregate data looks consistent with the tacit collusion where BP initiated cycle with price jump on Wednesday (market leader collusion); in period (III) there is a transition to the different agreement on tacit collusion; finally in period (IV) data looks consistent with the weakly price cycles starting on Thursday.

We discretize the set of actions and reduce it to the three possible actions: (1) increase, (2) decrease, and (3) no change. The price has changed is the difference between the old and new price is at least 6 cents per liter. That is, if the price increased or decreased by 8 cents per liter, then the actions would be an increase or a decrease correspondingly. If the price increased or decreased by 3 cents per liter, then there was no change. We conduct robustness checks for different magnitudes. We conduct robustness checks for other versions of discretization (0, 1, and 3 cents per liter being a bandwidth) and find consistent results.

We present results for the four different determinants of the history: (1) market trend, (2) BP, (3) day of the week, and (4) all of before.[26] Note that there are several

[26] Unless we take the day of the week into account, it will play the role of a counter. Our method, being a finite sample method, does not make a distinction between a counter and a set of finite memory sequences because they are observationally equivalent. Nonetheless, a counter may appear in a revealed automaton since it is the most concise representation of the corresponding strategy (and it often does in

stations owned by BP; therefore, whenever we talk about BP or market trend, we mean the aggregate behavior of all BP stations or all stations at the market. Otherwise, we would still be able to estimate the strategies but would not be able to perform the corresponding likelihood ratio test.[27]

Moreover, for simplicity, we present only results for the strategies of memory one. We present more extensive analysis with other combinations of the factors (not only all together) and different depths of memory in the Supplementary Materials.

Memory		I.	II.	III.	IV.
Market	Mean	.953	.88	.874	.857
	min Q_{1-3} max	.92 .94 .95 .96 1	.78 .85 .85 .90 1	.66 .85 .86 .90 1	.60 .83 .86 .87 1
BP	Mean	.954	.891	.872	.857
	min Q_{1-3} max	.92 .94 .95 .96 1	.78 .85 .87 .91 1	.65 .85 .86 .90 1	.61 .83 .86 .87 1
Day	Mean	.954	.919	.941	.96
	min Q_{1-3} max	.92 .94 .95 .96 1	.80 .89 .91 .93 1	.69 .92 .95 .97 1	.73 .95 .98 1 1
All	Mean	.956	.924	.944	.96
	min Q_{1-3} max	.92 .95 .95 .96 1	.80 .90 .92 .93 1	.71 .92 .95 .97 1	.73 .95 .98 1 1

TABLE 5. Mean and quartiles for $\hat{P}$ by time blocks and history determinants

Table 5 presents the mean estimated $\hat{P}$ with quartiles of the empirical distribution of individual-level $\hat{P}$. Each column corresponds to the time block. Each row-block corresponds to the different history determinants. Recall that the higher is $\hat{P}$, the better reconstructed strategy explains the data. In the period I all models explain the data equally well with marginal differences. For the latter periods, one can see that the $\hat{P}$ is higher for the All and Day of the week models comparing to the BP and Market models. In period II, the $\hat{P}$ the All model explains data better than the day of the week model. In period III, the $\hat{P}$ for the All model is still higher but is less sizeable. This observation can serve as the evidence that some of the firms still follow the market leader and did not yet learn to coordinate on the day of the week. Finally, in period IV,

the experimental data). For observational data, we endogenize counters to avoid imposing unreasonable state complexity on subjects to count time (and compare it with other specifications). Moreover, the original paper hypothesized that firms learn to coordinate using the day of the week. Hence, this counter seems to be natural and interesting to consider.

[27]Recall that the log-likelihood ratio is distributed according to a weighted χ^2 distribution with the degrees of freedom determined by the difference in the number of histories. If we condition on all possible actions in the market versus the day of the week, then just given the degrees of freedom of the asymptotic distribution it is not possible to reject the day of the week model in favor of a less restrictive one due to the properties of the asymptotic distribution.

the All model and the day of the week model explain data equally well. Similarly to the experimental data, we have different theories that can be more or less permissive. We control for this fact to draw the further inference.

	I.				II.				III.				IV.			
	Market	BP	Day	All	Market	BP	Day	All	Market	BP	Day	All	Market	BP	Day	All
Market		0%	0%	0%		0%	0%	0%		21%	1%	0%		14%	0%	0%
BP	3%		0%	0%	29%		2%	0%	5%		0%	0%	2%		0%	0%
Day	0%	0%		0%	63%	50%		0%	74%	78%		0%	79%	79%		0%
All	1%	0%	2%		69%	59%	8%		76%	81%	9%		79%	79%	0%	

TABLE 6. Share of subjects for whom the row model explains data better than the column model by time blocks

Table 6 presents the results for likelihood ratio tests. Unlike in the case, different depths of the memory the theories are not necessary nested. In order to perform the likelihood ratio test for non-nested theories, we use the result of Vuong (1989).[28] In time block I all theories explain data equally well. Note that this fact and the quite high $\hat{P}$ are the consequence of the fact that there is almost no fluctuation in prices during this period that does not match the market price of gas. During period II, the models that explain the data best are the All and Day of the week models, which is consistent with the results from Table 5. However, there is still a nontrivial share of firms (41% and 50%) for whom the market leader model (BP) explain data as well as the All and day of the week models. In periods III and IV for about 80% of subjects All and day of the week model explain the data better than the alternative models. If in time blocks III, there are still 9% of the firms for whom model with all factors as history determinants explain the data better than the day of the week model, there are no firms like that in period IV. This observation is consistent with the hypothesis from Byrne and De Roos (2019). That is, firms learn to coordinate purely on the day of the week over time. Even in the time block II, there is already half of the firms conditioning their action on the day of the week. This finding requires adjustment of the original hypotheses from Byrne and De Roos (2019).

5 Concluding Remarks

The paper presents an efficient algorithm to reconstruct strategies in dynamic games. The algorithm accounts for possible noise in the data and reconstructs the underlying

[28]Note that the test is originally symmetric; hence, it can only test whether theories explain the behavior equally well. To draw a more informative inference, we compare the values of the likelihoods for each of the theories and select the theory with higher likelihood as the one that explains the data better.

automaton using a simple mixed-integer program. The algorithm allows for individual-level heterogeneity of strategies but only valid for the pure strategies used by the players. The algorithm ignores possible equilibrium restrictions, although they can be easily embedded into the programming problem by adding the inequalities corresponding to equilibrium restrictions. However, this would require either further assumptions or increasing the size of the program. Even though it is clear how to embed equilibrium restrictions, the computational efficiency of the resulting program can be a significant concern (especially for large games). Hence, optimizing such an approach taking into account some game specifics, opens a fruitful venue for future research.

To illustrate the possible usage of the algorithm, we take it to both experimental and observational data. Note that both data sets fit as good illustrations for the cases when one may prefer to ignore equilibrium restrictions. Reason in the experimental data can be that, for instance, preferences are not perfectly observed (due to social or risk preferences). In the application to observational data, we use is the daily price data for the gas stations. Hence, we cannot compute the payoffs of the player without estimating the demand and cost functions. However, these estimates would require further assumptions and/or further data collection.

The experimental application allows us to show that in the repeated prisoner's dilemma, subjects are not using strategies of memory more than one. We also recover strategies that have not gotten attention in previous studies. Besides, we discover that the main driver behind the negative relation between the level of cooperation and the number of players in the group is a structural change in the strategies players use and not just an increasing number of unconditional defectors. In particular, there are still several subjects who use strategies that similar to tit-for-tat but start with defection. This observation has a significant effect on possible counterfactual analysis.[29] In particular, if we reduce the share of defectors in the population for the 2-player version, then tit-for-tat strategies perform better in the evolutionary tournaments and guarantee that the outcome is cooperative. While in the 3-player version, the tease-for-tat strategy becomes the favorite, but tease-for-tat delivers mutual defection as the outcome.

Finally, the application to the observational data allows us to confirm that gas stations learn to collude the weekly price cycles on the day of the week over time. This finding can be a worrying signal for the policymakers, who originally introduced the transparent pricing scheme to help consumers, but in practice, simplified collusion by introducing perfect public monitoring to the game. Even more worrying is that similar policies can be used on the labor market, to promote equality. In particular, making wages public can allow firms to collude on the wages for their employees since the underlying structural shift in the game can be similar.

[29]For more details, see Supplementary Materials.

References

Dilip Abreu and Ariel Rubinstein. The structure of nash equilibrium in repeated games with finite automata. *Econometrica: Journal of the Econometric Society*, pages 1259–1281, 1988.

Robert Axelrod and William Donald Hamilton. The evolution of cooperation. *science*, 211(4489):1390–1396, 1981.

Kenneth G Binmore and Larry Samuelson. Evolutionary stability in repeated games played by finite automata. *Journal of economic theory*, 57(2):278–305, 1992.

Gaetano Bloise. *Strategic complexity and equilibrium in repeated games.* PhD thesis, University of Cambridge, 1998.

Jordi Brandts and Gary Charness. The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14(3):375–398, 2011.

Yves Breitmoser. Cooperation, but no reciprocity: Individual strategies in the repeated prisoner's dilemma. *American Economic Review*, 105(9):2882–2910, 2015.

Lisa Bruttel and Ulrich Kamecke. Infinity in the lab. how do people play repeated games? *Theory and Decision*, 72(2):205–219, 2012.

David P Byrne and Nicolas De Roos. Learning to coordinate: A study in retail gasoline. *American Economic Review*, 109(2):591–619, 2019.

Gabriele Camera, Marco Casari, and Maria Bigoni. Cooperative strategies in anonymous economies: An experiment. *Games and Economic Behavior*, 75(2):570–586, 2012.

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Timothy N Cason and Vai-Lam Mui. Individual versus group choices of repeated game strategies: A strategy method approach. *Games and Economic Behavior*, 114:128–145, 2019.

Marco Castillo and Mikhail Freer. Revealed differences. *Journal of Economic Behavior & Organization*, 145:202–217, 2018.

KC Chanda. A note on the consistency and maxima of the roots of likelihood equations. *Biometrika*, 41(1/2):56–61, 1954.

Varadarajan V Chari and Patrick J Kehoe. Sustainable plans. *Journal of political economy*, 98(4):783–802, 1990.

Kalyan Chatterjee and Hamid Sabourian. Multiperson bargaining and strategic complexity. *Econometrica*, 68(6):1491–1509, 2000.

Kalyan Chatterjee and Hamid Sabourian. Game theory and strategic complexity. *Encyclopedia of Complexity and Systems Science*, pages 4098–4114, 2009.

Pedro Dal Bó and Guillaume R Fréchette. The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, 101(1):411–29, 2011.

Pedro Dal Bó and Guillaume R. Fréchette. Strategy choice in the infinitely repeated prisoner's dilemma. *American Economic Review*, 109 (11):3929–52, November 2019. doi: 10.1257/aer.20181480. URL `http://www.aeaweb.org/articles?id=10.1257/aer.20181480`.

Matthew Embrey, Guillaume R Fréchette, and Sevgi Yuksel. Cooperation in the finitely repeated prisoner's dilemma. 2015.

Jim Engle-Warnick and Robert L Slonim. Inferring repeated-game strategies from actions: evidence from trust game experiments. *Economic theory*, 28(3):603–632, 2006a.

Jim Engle-Warnick and Robert L Slonim. Learning to trust in indefinitely repeated games. *Games and Economic Behavior*, 54(1):95–114, 2006b.

Armin Falk, Ernst Fehr, and Urs Fischbacher. Driving forces behind informal sanctions. *Econometrica*, 73(6):2017–2030, 2005.

Daniel Friedman. Evolutionary games in economics. *Econometrica: Journal of the Econometric Society*, pages 637–666, 1991.

Drew Fudenberg, David G Rand, and Anna Dreber. Slow to anger and fast to forgive: Cooperation in an uncertain world. *The American Economic Review*, 102(2):720–749, 2012.

Drew Fundenberg and Eric Maskin. Evolution and cooperation in noisy repeated games. *The American Economic Review*, 80(2):274–279, 1990.

Julián García and Matthijs van Veelen. In and out of equilibrium i: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory*, 161: 161–189, 2016.

John Haltiwanger and Joseph E Harrington Jr. The impact of cyclical demand movements on collusive behavior. *The RAND Journal of Economics*, pages 89–106, 1991.

Cars Hommes, Joep Sonnemans, Jan Tuinstra, and Henk van de Velden. A strategy experiment in dynamic asset pricing. *Journal of Economic Dynamics and Control*, 29(4):823–843, 2005.

R Mark Isaac and James M Walker. Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, 103(1): 179–199, 1988.

R Mark Isaac, James M Walker, and Arlington W Williams. Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups. *Journal of public Economics*, 54(1):1–36, 1994.

Ehud Kalai and William Stanford. Finite rationality and interpersonal complexity in repeated games. *Econometrica: Journal of the Econometric Society*, pages 397–410, 1988.

Michihiro Kandori, George J Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica: Journal of the Econometric Society*, pages 29–56, 1993.

Narayana R Kocherlakota. Implications of efficient risk sharing without commitment. *The Review of Economic Studies*, 63(4):595–609, 1996.

Erich L Lehmann and George Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.

Ethan Ligon, Jonathan P Thomas, and Tim Worrall. Informal insurance arrangements with limited commitment: Theory and evidence from village economies. *The Review of Economic Studies*, 69(1):209–244, 2002.

Jona Linde, Daniel Gietl, Joep Sonnemans, and Jan Tuinstra. Information and learning in the minority game: A strategy experiment. 2016.

Peter Linz. *An introduction to formal languages and automata*. Jones & Bartlett Learning, 2006.

George J Mailath and Larry Samuelson. *Repeated games and reputations: long-run relationships*. Oxford university press, 2006.

Henry B Mann and Abraham Wald. On stochastic limit and order relationships. *The Annals of Mathematical Statistics*, 14(3):217–226, 1943.

Robert McNaughton and Seymour A Papert. *Counter-Free Automata (MIT research monograph no. 65)*. The MIT Press, 1971.

Friederike Mengel and Ronald Peeters. Strategic behavior in repeated voluntary contribution experiments. *Journal of Public Economics*, 95(1-2):143–148, 2011.

Patrick Alfred Pierce Moran. Random processes in genetics. In *Mathematical proceedings of the cambridge philosophical society*, volume 54, pages 60–71. Cambridge University Press, 1958.

Laurent Muller, Martin Sefton, Richard Steinberg, and Lise Vesterlund. Strategic behavior and learning in repeated voluntary contribution experiments. *Journal of Economic Behavior & Organization*, 67(3-4):782–793, 2008.

Martin A Nowak, Akira Sasaki, Christine Taylor, and Drew Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646, 2004.

Kamta Rai and John Van Ryzin. A note on a multivariate version of rolle's theorem and uniqueness of maximum likelihood roots. *Communications in Statistics-Theory and Methods*, 11(13):1505–1510, 1982.

Julian Romero and Yaroslav Rosokha. Constructing strategies in the indefinitely repeated prisoner's dilemma game. *European Economic Review*, 104:185–219, 2018.

Julio J Rotemberg and Garth Saloner. A supergame-theoretic model of price wars during booms. *The American Economic Review*, 76(3):390–407, 1986.

Tobias Salz and Emanuel Vespa. Estimating dynamic games of oligopolistic competition: An experimental investigation. *Available at SSRN 2621270*, 2015.

Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and control*, 8(2):190–194, 1965.

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

Reinhard Selten, Michael Mitzkewitz, and Gerald R Uhlich. Duopoly strategies programmed by experienced players. *Econometrica: Journal of the Econometric Society*, pages 517–555, 1997.

Joep Sonnemans, Cars Hommes, Jan Tuinstra, and Henk van de Velden. The instability of a heterogeneous cobweb economy: a strategy experiment on expectation formation. *Journal of Economic Behavior & Organization*, 54(4):453–481, 2004.

Quang H Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333, 1989.

A Proofs

Before we proceed with the proofs let us make a small remark about the automata. Formally we allow for disconnected machines. That is not all the states can be reached from the initial state given the sequence of signals. Denote by

$$\hat{\Omega}(\mu) = \left\{\omega : \omega = T^t(\omega, a_{-i}^1, \dots, a_{-i}^t)\ \forall t > 0\ \forall a_{-i}^1, \dots, a_{-i}^t \in A_{-i} \cup \{\emptyset\}\right\}$$

the maximal connected component of the Moore machine. Denote by $\hat{\mu}$ the maximal component of the automaton connected to the initial state μ. Denote by $\mu \sim \mu'$ the observationally equivalent automaton i.e. the automata such that $\hat{\mu} = \hat{\mu}'$. That is any machine such that the maximal components connected to the initial state are the same.

Observation 1. *A data set is rationalizable with μ if and if it is rationalizable with any $\mu' \sim \mu$.*

Recall the definition of rationalizability. A data set is *rationalizable with an automaton* if there is μ such that $G(T^t(\hat{\omega}, a_{-i}^s, \dots, a_{-i}^{s+t-1})) = a_i^t$ for every $t \in \{1, \dots, T\}$. That is, the observed action can only be generated as an outcome symbol from the state connected to the initial state. Hence, we can restrict our attention to the partitioning of the Moore machines with respect to $\sim$ relation.

A.1 Proof of Lemma 1

We present proofs for both parts separately.

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

($\Leftarrow$). Let us prove this by constructing the 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 $s(h^\tau) = \{a^t : h^\tau(t) = h^\tau\}$. First, since we require

$$h^\tau(t) = h^\tau(s) \text{ implies } a^t = a^s$$

we can guarantee that $|s(h)| = 1$, i.e. $s : \hat{H} \to A_i$ is a well-defined function that rationalizes the data by construction. However, at this point s is only a partial strategy and it remains to complete the strategy. Note that completion of the strategy wouldn't affect the rationalization, since the only remaining histories are the unobserved histories. Hence, for every $h \in H^\tau \setminus \hat{H}^\tau$ we can assign an arbitrary action to $s(h) = a_i$.

□

Proof for automaton rationalization. ($\Rightarrow$). Suppose that the data set is rationalizable, but one of the conditions is violated. Consider these potential violations one-by-one.

(1) Suppose that $\xi(1) \neq \xi(T_k + 1)$. Recall that $t = 1$ and $t = T_k + 1$ corresponds to the rematching. That is an automaton is supposed to have two different initial

states, which is a contradiction to the definition of the automaton, since the initial state is unique $\hat{\omega} \in \Omega$.

(2) Assume that there is $\xi(t) = \xi(v) = \omega \in \Omega$ such that $a_i^t \neq a_i^v$. Recall that since the data set is rationalizable, $G(\omega) = a^t$, where $\omega = T^t(\hat{\omega}, a_i^1, \dots, a_i^{t-1}) = T^v(\hat{\omega}, a_i^1, \dots, a_i^{v-1})$. It is a contradiction since G is a function, hence, $G(\omega)$ is a singleton.

(3) Assume that $\xi(t) = \xi(v) = \omega \in \Omega$ and $a_{-i}^t = a_{-i}^v$ and $\xi(t+1) \neq \xi(v+1)$. Hence, $T(\omega, a_{-i}^t) = \xi(t+1) \neq \xi(v+1) = T(\omega, a_{-i}^v)$. However, it is a contradiction, since $a_{-i}^t = a_{-i}^v$ and therefore $T(\omega, a_{-i}^t) = T(\omega, a_{-i}^v)$.

($\Leftarrow$). Suppose there is a function that satisfies the conditions and let us construct the automaton. Recall that an automaton is a 4-tuple $\mu = (\Omega, \hat{\omega}, T, G)$. Hence, we are left to define the initial state, the output function, and the transition function. We start by constructing the partial automaton that rationalizes the data set. Next, we show that any extension of partial automaton that rationalizes the data also rationalizes the data set.

– Let $\hat{\omega} = \xi(1) \cup \bigcup_{k \leq K} \xi(T_k + 1)$. Given that $\xi(1) = \xi(T_k + 1)$ for every $k \leq K$, hence, $\hat{\omega}$ is well defined and $|\hat{\omega}| = 1$.
– Let $G(\omega) = \bigcup_{t:\xi(t)=\omega} a^t$. Since $|G(\omega)| \leq 1$ since $\xi(t) = \omega = \xi(v)$ implies that $a^t = a^v$, the the output function is singleton whenever defined.
– Let $T(\omega, a_{-i}) = \bigcup_{t:\xi(t)=\omega; a^t = a_{-i}} \xi(t+1)$. Condition (3) implies that $|T(\omega, a_{-i})| \leq 1$ and is well-defined whenever the corresponding history is observed.

Hence, we have constructed a partial automaton. We now show that it rationalizes the data. On the contrary, assume that there is $t \in T$ such that $G(T^t(\hat{\omega}, a_{-i}^1, \dots, a_{-i}^t)) \neq a_i^t$. However, recall that by construction of the automaton, $T^t(\hat{\omega}, a_{-i}^1, \dots, a_{-i}^t) = \xi(t)$ and $a_i^t = G(\xi(t))$. Hence, the output symbol coincides with the one prescribed by the automaton.

Finally, we need to complete the automaton, since the functions G and T are partial. Note that any completion of this automaton would still rationalize the data, since for an automaton $\tilde{\mu}$ we know that for every $t \in T$ such that $\tilde{G}(\tilde{T}^t(\hat{\omega}, a_{-i}^1, \dots, a_{-i}^t)) = G(T^t(\hat{\omega}, a_{-i}^1, \dots, a_{-i}^t)) = a_i^t$. For every $\omega \in \Omega$ and $a_i \in A_i$ let $\tilde{T}(\omega, a_i) = \omega$ if $T(\omega, a_i) = \emptyset$; and $\tilde{T}(\omega, a_i) = T(\omega, a_i)$ otherwise. For every $\omega \in \Omega$ let $\tilde{G}(\omega) = a_i \in A_i$ if $G(\omega) = \emptyset$; and $\tilde{G}(\omega) = G(\omega)$ otherwise.

□

A.2 Proof of Remark 1

Proof. Note that if the data is rationalizable with a strategy of memory $\bar{T}$, then it is tautologically rationalizable with a strategy of finite memory. Hence, we are left to prove that if a data set is rationalizable with a strategy of finite memory, then it is rationalizable with a strategy of memory $\bar{T}$. Before we proceed with the proof of the Remark itself, let us provide supplementary results: *If a data set is rationalizable with a strategy of memory τ, it is rationalizable with a strategy of memory τ' for every $\tau' \geq \tau$.*

Recall that according to Lemma 1 we know that a data set is rationalizable with a strategy of memory τ if and only if for every $t, v \in T$ such that

$$h^\tau(t) = h^\tau(v) \text{ implies } a_i^t = a_i^v.$$

Note that it enough to prove that the data set is rationalizable with $\tau + 1$. Assume on the contrary that the data set is not rationalizable with $\tau + 1$. That is there are $t, v \in T$ such that $h^{\tau+1}(t) = h^{\tau+1}(v)$ and $a_i^t \neq a_i^v$. However, construction of the history implies that $h^\tau(t) = h^\tau(v)$. Since the data set is rationalizable with a strategy of memory τ, then $a_i^t = a_i^v$ that is a contradiction.

The claim above already guarantees that if a data set is rationalizable with a strategy of memory $\tau \leq \bar{T}$, then it is rationalizable with a strategy of memory $\bar{T}$. To complete the proof it is enough to show that if data set is not rationalizable with a strategy of memory $\bar{T}$, then it is not rationalizable with a strategy of memory $\tau \geq \bar{T}$. Recall that $\bar{T}$ is the longest match, hence, the only way to obtain such a history is to add empty histories to the beginning. That is $h^\tau = (\emptyset, \ldots, \emptyset, h^{\bar{T}})$, so the empty histories are always equal to each other. Hence, if there are $t, v \in T$ such that $h^{\bar{T}}(t) = h^{\bar{T}}(v)$, then $a_i^t = a_i^v$, then simply by adding enough of empty histories to the beginning. This fact would guarantee that the data set cannot be rationalized with memory τ.

□

A.3 Proof of Proposition 2

Proof. ($\Rightarrow$) Take a strategy that rationalizes the data set. Further, we construct the automaton that represents the strategy of memory τ. Let us start from an initial state $\hat{\omega}$, and let $G(\hat{\omega}) = s(\emptyset)$. 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 $h^\tau + a = (h^{t-\tau-2}, \ldots, h^{t-1}, a)$ – that is an operation of adding a^t to the queue (line) of the length τ.[30] Since we are adding an element to the end of the queue, it pushes out the element in the beginning of the queue. Next, let $T(\omega(h^\tau, a_{-i})) = \omega(h^\tau + (s(h^\tau, a_{-i})))$ for every $\omega \in \Omega$ and $a_{-i} \in A_{-i}$. We constructed an automaton since it has fully defined set of states and well-defined transition function. That is for every state there is a transition to one of the states given any potential actions of other players. Figure 7 illustrates how the algorithm works for the strategy of memory two in the stage game with two players and two actions for each. For simplicity assume that the revealed strategy for player

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

1 has the same action a_1 for all histories. In general different actions would require different mappings between the states. The automaton can then be minimized. This automaton in particular can be minimized into an automaton with a single state.

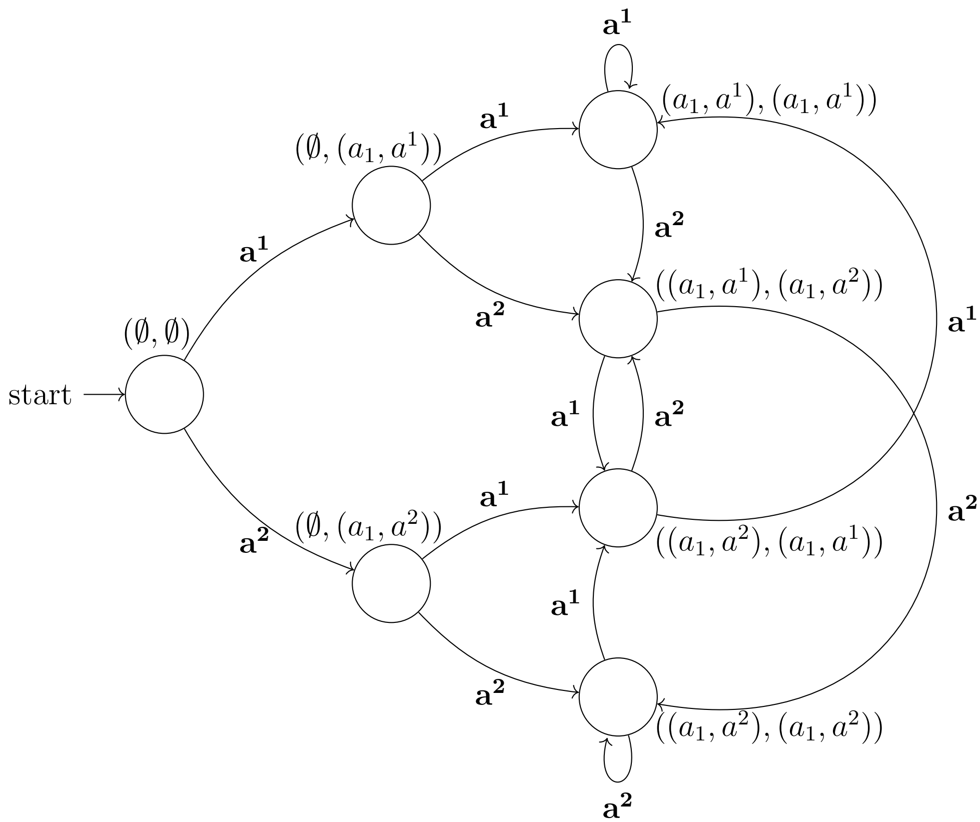


FIGURE 7. Construction of automaton from strategy of $\tau = 2$

We are left to show that it satisfies τ transition invariance. By construction $T(\omega(h^\tau), a_{-i}) = \omega(h^\tau + (s(h^\tau), a_{-i}))$. Denote by $\tilde{h}^\tau = (a^t, \ldots, a^{t+\tau})$ for some sequence of signals. Hence, if

$$G(T^t(\omega, a^1_{-i}, \ldots, a^t_{-i})) = G(T^t(\omega', a^1_{-i}, \ldots, a^t_{-i}))$$

for every $t < \tau$, implies that

$$T^\tau(\omega(h^\tau), a^1_{-i}, \ldots, a^\tau_{-i}) = \omega(h^\tau + a^1_{-i} + \ldots + a^\tau_{-i}) = \omega(\tilde{h}^\tau)$$

for every $\omega(h^\tau) \in \Omega$.[31] Since output function returns the unique symbol, then τ transition invariance is trivially satisfied.

[31]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.

($\Leftarrow$) Take an automaton μ that rationalizes the data. Denote by $\emptyset^\tau$ the initial history, i.e. the history that contains τ empty symbols. Let

$$s(\emptyset^\tau) = G(\hat{\omega})$$

and

$$s(h^\tau) = \bigcup_{\omega \in \Omega} G(T^\tau(\omega, (a^1_{-i}, \ldots, a^\tau_{-i}))$$

where $h^\tau = ((G(\omega), a^1_{-i}), \ldots, (G(T^{\tau-1}(\omega, (a^1_{-i}, \ldots, a^{\tau-1}_{-i})), a^\tau_{-i}))$. Note that there are some histories not yet covered by the automaton. Denote the histories corresponding to the $\hat{\Omega}$ by $\hat{H}^\tau$.

Observation 2. *$\hat{H}^\tau$ includes all observed histories.*

Intuition behind the Observation 2 is that the observed histories are composed of the Since an automaton is τ transition invariant, then $|s(h^\tau)| = 1$ for every $h^\tau \in \hat{H}^\tau$. Recall that by construction $s(h^\tau(t)) = G(T^\tau(\omega, (a^1_{-i}, \ldots, a^\tau_{-i})) = a^t_i$, since Moore machine rationalizes the data. Hence, the constructed partial strategy also rationalizes the data. So we are left to extend the strategy such that it rationalizes the data. Take every $h^\tau \in H^\tau \setminus \hat{H}^\tau$ and assign an arbitrary action to $s(h^\tau)$. Hence, the completed strategy rationalizes the data set since we only assigned actions to the unobserved histories (see Observation 2). $\square$

We mentioned the fact that this part provides the algorithm to convert a strategy into an automaton. The algorithm we use to construct an automaton from a strategy is less straight-forward. Moreover, it does not have to return the minimal machine. However, the Moore automaton can be minimized using the *mixed-integer programming* problem. Further, we provide the direct mixed-integer program to convert (partially identified) strategy into automaton.

To start with the construction of the minimal automaton we need to define the set of potential states. Proof above immediately implies that τ transitive invariant machine can be described by at most $|\bar{\Omega}| \leq |A_{-i}|^\tau + 1$ states.[32] Hence, define by $\bar{\Omega}$ the set of possible states. Let x be a binary matrix of the size of $|H^\tau| \times |\bar{\Omega}|$. Every cell in the matrix $x_{\omega,t}$ is equal to one if the state ω corresponds to the response to the history $t \in H^\tau$. Let y be a binary vector of length $|\bar{\Omega}|$ that shows whether the state ω is active. This variable is important, since $\bar{\Omega}$ represents the potential set of states, i.e. they do not have to be employed by the actual machine. Recall that 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. Denote by $\mathbf{1}_{f(x)}$ the indicator function that returns 1 if the logical expression is correct and 0 otherwise. Hence, we can formulate the conversion problem as the following minimization problem.

[32]This bound is legit under the assumption that $|A_i| \leq |A_{-i}|$ we keep this assumption to avoid further abuse of notation.

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

Condition (S→A) 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 (S→A) would recover the minimal automaton since it is essentially merging the algorithm from the proof of Proposition 2 and the minimization of the Moore machine. Let us provide some brief intuition behind the algorithm.

We start with the conditions that correspond to y_ω. The inequality that $y_\omega \geq x_{\omega,t}$ essentially controls that the state is active if it corresponds to at least one of the histories. And as the objective, we want to find the minimal machine, i.e. we try to minimize the number of active states that corresponds to the minimization of the $\sum_{\omega\in\bar{\Omega}} y_\omega$. Next, we describe the conditions for $x_{w,t}$. First, we need to make sure that every history is presented at least by one state of the machine, i.e. $\sum\limits_{\omega\in\Omega} x_{\omega,t} = 1$. Finally, one state can represent multiple histories we just need to ensure that they deliver the same output symbol (first inequality) and transition these histories goes to the equivalent states (second inequality).

A.4 Proof of Proposition 1 Note that part (2) already implies one side of the implication. In particular, if a data set is rationalizable with a strategy of finite memory, then it is rationalizable with a Moore machine. Hence, we are left to show the reverse, i.e. if a data set is rationalizable with an automaton, then it is rationalizable with a strategy of finite memory.

Proof of (⇐). Denote by

$$\bar{T} = \max_{k\leq K}\{T_k\}$$

the maximal length of a match. Consider a data set rationalizable with Moore machine, and let us show that the data set is rationalizable with a strategy of memory $\bar{T}$.

Recall that a data set is rationalizable with an automaton if $s(h^{\bar{T}}(t)) = s(h^{\bar{T}}(t'))$ if $h^{\bar{T}}(t) = h^{\bar{T}}(t')$ for every $t, t' \in T$. Assume on the contrary that the data is not rationalizable with a strategy of memory $\bar{T}$. Hence, there are two similar histories such that the corresponding actions are different.

Recall that automaton has the unique initial state $\hat{\omega}$, and we assume that the machine is restarted with every rematching. Consider the two histories for which we observe the violation. To show that this creates a contradiction with rationalizability with an automaton we need to use a little abuse of notation. In particular, we can refer to every state of the maximal connected part of the automaton as to transition from the initial state using a history of $h^{\bar{T}}$. This requires formal enriching of the alphabet with $\emptyset$ and requirement, that once we observe $\emptyset_{-i}$ as input symbol, then any state is mapped into itself.

Hence, a violation of rationalization with an automaton of memory $\bar{T}$ implies that

$$a^t = G(T^{\bar{T}}(\hat{\omega}, h^{\bar{T}}(t)) \neq G(T^{\bar{T}}(\hat{\omega}, h^{\bar{T}}(t')) = a^{t'}$$

given $h^{\bar{T}}(t) = h^{\bar{T}}(t')$. That is an immediate contradiction of the rationalizability with a Moore machine, because at least in one instance the output symbol of the Moore machine does not match the observed action.

□

A.5 Proof of Remark 2

Proof. Consider the construction of the automaton from strategies from the proof of Proposition 2. 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 node, 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 the creation of this state.

□

A.6 Proof of Proposition 3 Before we proceed with the proofs let us show that our estimate can be a solution to the maximum likelihood problem. We will assume that $P \in [0.5, 1[$ here for simplicity, because for $P = 1$ a deterministic test suffices. This is not a restrictive assumption, and a similar proof for the general case of $P \in [0.5, 1]$ is in the online appendix. Constrained likelihood function can then be expressed as

$$\mathcal{L}(X, \theta) = \prod_{h \in H} \sum_{a \in A_i} (\sigma_a(h) P^{q_a(h)} (1-P)^{q(h)-q_a(h)}),$$

or as log-likelihood $l = \log \mathcal{L}(X, \theta)$:

$$l = \sum_{h \in H} \log \sum_{a \in A_i} (\sigma_a(h) P^{q_a(h)} (1-P)^{q(h)-q_a(h)}),$$

with restrictions that $\sum_a \sigma_a(h) = 1$ for all $h \in H$, $\sigma_a(h) \in [0,1]$ for all $a \in A_i, h \in H$, and $1 > P > 0.5$. Here $q_a(h)$ - frequency of action a after history h in observations, $q(h) = \sum q_a(h)$ - frequency of history h in observations, and $\hat{H}$ - the set of all observed histories. Note that this unbounded case is a zero-probability event if the model is true.

It is easy to see that it is maximized in the space of admissible parameters $\hat{P} \in]0.5, 1]$ and $\sum_{a \in A_i} \hat{\sigma}_a(h) = 1$ when two conditions hold. First, the $\hat{\sigma}_a(h) = 1$ for a, such that $q_a(h)$ is one of the highest for h. That is the strategy is the corresponding pure strategy (corner solution). And, second, $\hat{P} = 1 - \frac{k}{T}$, where $k = \sum_h (q(h) - max_a(q_a(h)))$ is the total number of deviations from the strategy. For the first condition, notice that first derivative of log-likelihood in any $\hat{\sigma}_a(h)$ for any fixed value of $1 > P > 0.5$ and any $h \in H$ is

$$\frac{\hat{P}^{q_a(h)} (1-\hat{P})^{q(h)-q_a(h)}}{\sum_{a \in A_i} \hat{\sigma}_a(h) \hat{P}^{q_a(h)} (1-\hat{P})^{q(h)-q_a(h)}}$$

That is the denominator is the same for all $a \in A_i$, and the numerator is the highest for a that has the largest $q_a(h)$. This leads to a corner solution with corresponding $\hat{\sigma}_a(h) = 1$. Hence the estimate is not Type 2, i.e. it is not an interior solution of likelihood maximization, so it does not satisfy the usual set of regularity conditions for MLE to have desirable properties (see e.g. p. 472 and p. 516 in Casella and Berger, 2002).

Hence, we need to regularize likelihood using change of variables, such that the likelihood for pseudo-model would satisfy the regularity conditions. In particular we introduce pseudo-parameters $\bar{\sigma}_a(h)$ and $\bar{P}$ such that

$$\hat{\sigma}_a(h) = (\sin \bar{\sigma}_a(h))^2$$

for $\bar{\sigma}_a(h) \in]\varepsilon, \pi + \varepsilon[$ for some $\varepsilon \in]0, \pi/2[$, and

$$\hat{P} \to 0.5 + \epsilon + (\frac{1}{2} - \epsilon) \frac{\bar{P}^2}{1 + \bar{P}^2},$$

for any small positive $\epsilon < \frac{1}{2}$.

Given the substitution of variables, we can see that the solution for the likelihood of the pseudo-model, the solution for $\bar{\sigma}_a(h)$ are such that:

$$\bar{\sigma}_a(h) \in \{\pi/2; \pi\} \text{ and } \sum_{a \in A_i} (\sin(\bar{\sigma}_a(h))^2 = 1 \text{ for every } h \in \hat{H},$$

hence, the solution is interior for $\bar{\sigma}_a(h)$ and is interior for $\bar{P}$.

Recall that our primary objective here is to prove the consistency of $\hat{\sigma}$ and $\hat{P}$. For this purpose we show that corresponding transformed parameters are consistent. Note that $\hat{\sigma}$ and $\hat{P}$ are continuous transformations of $\bar{\sigma}$ and $\bar{P}$ correspondingly. Hence, the continuous mapping theorem (Mann and Wald, 1943) implies that if $\bar{\sigma}$ and $\bar{P}$ are consistent then their continuous transformations ($\hat{\sigma}$ ad $\hat{P}$) are consistent as well. Hence, it suffices show that transformed estimators are consistent. In order to do so, we show that the transformed likelihood is *regular*. The regularity of likelihood implies that $\bar{\sigma}$ and $\bar{P}$ are consistent, asymptotically efficient and asymptotically normal. The last two properties provide additional leverage for the further analysis we use in the paper. The regularity conditions are summarized in the following Lemma.

Lemma 2. *Transformed likelihood model satisfies the regularity conditions from Lehmann and Casella (2006):*

(A0) *The distributions of observations are identifiable, that is if $\theta \neq \theta'$ then $f(x|\theta) \neq f(x|\theta')$;*
(A1) *The distributions of observations have common support;*
(A2) *The observations X are i.i.d with density $f(x_i|\theta)$;*
(A3) *There exists an open subset of parameter space containing the true parameter point θ^0 such that for almost all x, the density $f(x|\theta)$ admits all third derivatives $\frac{\partial^3}{\partial\theta_j\theta_k\theta_l} f(x|\theta)$ for all θ;*
(B) *The first and second logarithmic derivatives of f satisfy:*

$$E[\frac{\partial}{\partial\theta_j} \log f(X|\theta)|\theta] = 0$$

for $j = 1, ...s$ and

$$I_{jk}(\theta) = E[\frac{\partial}{\partial\theta_j} \log f(X|\theta) \frac{\partial}{\partial\theta_k} \log f(X|\theta)|\theta] = E[-\frac{\partial^2}{\partial\theta_j\partial\theta_k} \log f(X|\theta)|\theta],$$

where I is Fisher information matrix;
(C) *$I_{jk}(\theta)$ are finite, $I(\theta)$ is positive definite for all θ;*
(D) *There are functions M_{jkl} such that $|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} \log f(x|\theta)| \leq M_{jkl}(x)$ for all θ, where $m_{jkl} = E[M_{jkl}(X)|\theta = \theta^0] < \infty$ for all j, k, l.*[33]

Proof. We prove the properties in order they listed in the Lemma.

A0-A2: are immediate from the assumption of the model. Given that original likelihood is identifiable (recall that we restrict our attention to observable histories only) and we take the bijective transformation of parameters we can guarantee that the transformed parameters are also identifiable. Further given the data generating process we explicitly assume that X are i.i.d. with density $f(x_i|\theta)$ and common support.

A3: First, let us show that the true parameters lie in the interior of the set of parameters. Solutions of the original likelihood (under the null hypothesis) are $\hat{\sigma}_a(h) \in \{0, 1\}$

[33]Condition (A) is (i), (C) is (iii) and (D) is part of (ii) in Chanda (1954) and Rai and Van Ryzin (1982). The remainder of (ii) also holds, i.e. the first two derivatives of $\mathcal{L}$ are bounded.

for every $a \in A_i$, $h \in \hat{H}$ and $\hat{P} \in].5, 1[$. Hence, the transformed parameters would also line in the interior of the corresponding interval. In particular, $\bar{\sigma}_a(h) \in \{\pi/2, \pi\} \in]\varepsilon; \pi + \varepsilon[$ for every $a \in A_i$, $h \in \hat{H}$ and $\bar{P} > 0$.

Next, we need to illustrate that likelihood is at least three-times differentiable. Note that in its original form, the likelihood is polynomial, hence, it is infinitely differentiable. Moreover, since our transformations are at least three times differentiable, hence, the transformed likelihood is also at least three-times differentiable.

B: The argument is standard. Since $1 = \int f(x|\theta)dx$ holds for every θ, we can take derivative to write

$$0 = \frac{\partial}{\partial \theta_j} \int f(x|\theta)dx = \int \frac{\frac{\partial f(x|\theta)}{\partial \theta_j}}{f(x|\theta)} f(x|\theta)dx = E[\frac{\partial}{\partial \theta_j} \log f(X|\theta)|\theta_j],$$

where switching integral and derivative (Leibniz integral rule) is simply a derivative of sum, since $f(x|\theta)$ is discrete. For continuous distributions, this holds under additional mild regularity conditions. Differentiating this expression using the product rule yields the following:

$$0 = \frac{\partial}{\partial \theta_j} \int \log f(x|\theta) f(x|\theta)dx = \int \frac{\partial^2}{\partial \theta_j \theta_k} \log f(x|\theta) f(x|\theta)dx + \int \frac{\partial}{\partial \theta_j} \log f(x|\theta) \frac{\frac{\partial}{\partial \theta_k} f(x|\theta)}{f(x|\theta)} f(x|\theta)dx$$

or

$$0 = E[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f(X|\theta)|\theta] + E[\frac{\partial}{\partial \theta_j} \log f(X|\theta) \frac{\partial}{\partial \theta_k} \log f(X|\theta)|\theta].$$

This yields the second part of condition (B).

C: First, we explain that all derivatives exist and finite. Both derivatives are defined for all admissible values of $\bar{P}$, $\bar{\sigma}_a(h)$, $q_a(h)$, q_h. The derivatives are also bounded in terms of $\hat{P}$ since $\frac{\partial \hat{P}}{\partial \bar{P}}$ and $\frac{\partial \sigma_a(h)}{\partial \bar{\sigma}_a(h)}$ are also bounded by construction.

Next, we show that I is positively definite. Since $\frac{\partial}{\partial \theta_j} \log f(X|\theta) > 0$ for any consistent data and any j, $I(\theta)$ is positive definite. The quadratic form in question can be written as $u^T I u$, and from (B) we obtain:

$$\sum_{i,j=1}^{m} u_i E[\frac{\partial}{\partial \theta_i} \log f(X|\theta) \frac{\partial}{\partial \theta_j} \log f(X|\theta)|\theta] u_j,$$

where m is the number of rows in the information matrix (the total number of variables). By linearity of expectation:

$$u^T I u = E[||\frac{\partial}{\partial \theta_i} \log f(X|\theta) u||^2|\theta],$$

which is strictly positive for any non-zero u given that the score vector $\frac{\partial}{\partial \theta_j} \log f(X|\theta)$ is strictly positive. Hence, the matrix I is positively definite.

D: For the last condition, notice that for the $|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}\log f(x|\theta)|$ is bounded for every $P \in [.5, 1[$. For the same argument $I_{jk}(\theta)$ is finite. For explicit matrix I see the derivation for variance below. $\square$

Proposition 4 (Lehmann and Casella (2006))**.** *If a likelihood satisfies A0-D regularity conditions, then the corresponding estimates are consistent, asymptotically efficient and asymptotically normal.*

Proposition 4 concludes the proof of Proposition 3 given the arguments we listed above. Moreover, the asymptotic normality of the estimate allows us to compare models using the likelihood ratio tests.

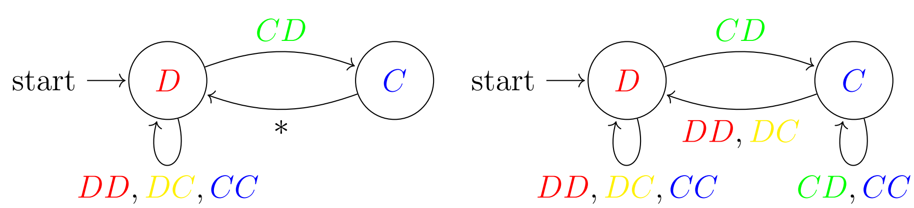
A.7 Proof of Remark 3. This remark directly follows from the proof above.

A.8 Proof of Remark 4. Recall that every automaton has a unique equivalent minimal automaton. Moreover, since the estimate of the strategy is consistent, then there is a unique (up to equivalence and observational equivalence) automaton. Hence, the uniqueness of equivalent automaton guarantees its consistency.

B Additional Analysis

B.1 Experimental Application.

B.1.1 Asymmetric Strategies in 3-PD. For one subject symmetric strategies can be rejected in favor of asymmetric strategies at 1% level for three-player PD. For one other subject – at 15% level. Their automata are illustrated below. For all other subjects symmetric strategies can not be rejected at any level above machine precision.



B.1.2 Removing Last Period. Recall that according to the experimental protocol we have the deterministic ending after 160 periods. Hence, in the last match (even though subjects do not see the round count) subjects may deviate from the indefinite nature of the game. Therefore, to ensure that the last match and finite nature does not affect the results.

Table 7 presents the results for the likelihood ratio test for different depths of memory by the experimental treatment. Note that once we remove the last period, the results become even more prominent for PD treatments. That is, even for more subjects strategy of memory one is sufficient to explain the observed play. At the same time, even the results for BoS treatment look more consistent with strategies of memory one.

	2-PD				3-PD				BoS			
	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$
$\tau = 1$		92%	100%	100%		100%	100%	100%		71%	100%	100%
$\tau = 2$			100%	100%			100%	100%			100%	100%
$\tau = 3$				100%				100%				100%

TABLE 7. Share of subjects for whom restricted model (smaller τ) can not be rejected in favor of unrestricted (larger τ) model for different depths of memory.

However, the difference between PD treatments and BoS treatment is still sizeable. All likelihood tests for experimental data are also robust to estimating at fixed 10% level and 90% 95% 99% percentile power-corrected thresholds.

B.1.3 Revealed strategies in Battle of Sexes Figure 8 presents the strategies for the BoS game. Strategies mainly differ in the way they enforce the potential counterpart to share the surplus. However, there are still five subjects who always choose one action (that is an analog for unconditional cooperation/defection in PD). Two subjects (8 and 9) enforce the split of the surplus via strict taking turns. Eight subjects use a match-switch strategy (3 and 5). That can be treated as an analog for tit-for-tat strategies in PD. Among the strategies of the higher memory interesting one is BAB (10), which combines both match-switch and unconditional taking-turns strategy.

B.1.4 Evolutionary Tournaments and Counterfactual Analysis. Table 8 presents detailed results for evolutionary tournaments. Moreover, the same table reports the potential counterfactual varying the probability of errors from absence of noise to a strategy with noise exceeding the estimated level by 10%.

We first present a detailed analysis for the BoS treatment, which is omitted in the main text. A-sharing and B-sharing are the first and the second strategies in Fig. 8 respectively. These strategies can be interpreted as insisting on sharing the surplus over two consecutive periods, switching between the two pure strategy equilibria. The difference between these two strategies is the choice of the first action - A-sharing starts with equilibrium benefiting the player, and B-sharing starts with equilibrium benefiting the opponent. Tournament results suggest that with no or low noise, A-sharing subjects can dominate the tournament and establish sharing in the population. In the absence of noise, A-sharing can perform as well as always choosing the high payoff action. When the noise is significant, the latter strategy becomes increasingly more stable in the tournaments.

The first of the $\tau = 2$ strategies has the same interpretation of coordinating on several equilibria with surplus shared every three rounds according to the rule “BAB”. Interestingly, this is the fairest sharing rule in terms of absolute difference between payoffs of the two players among rules of memory $\tau \leq 3$. However, we do not see enough

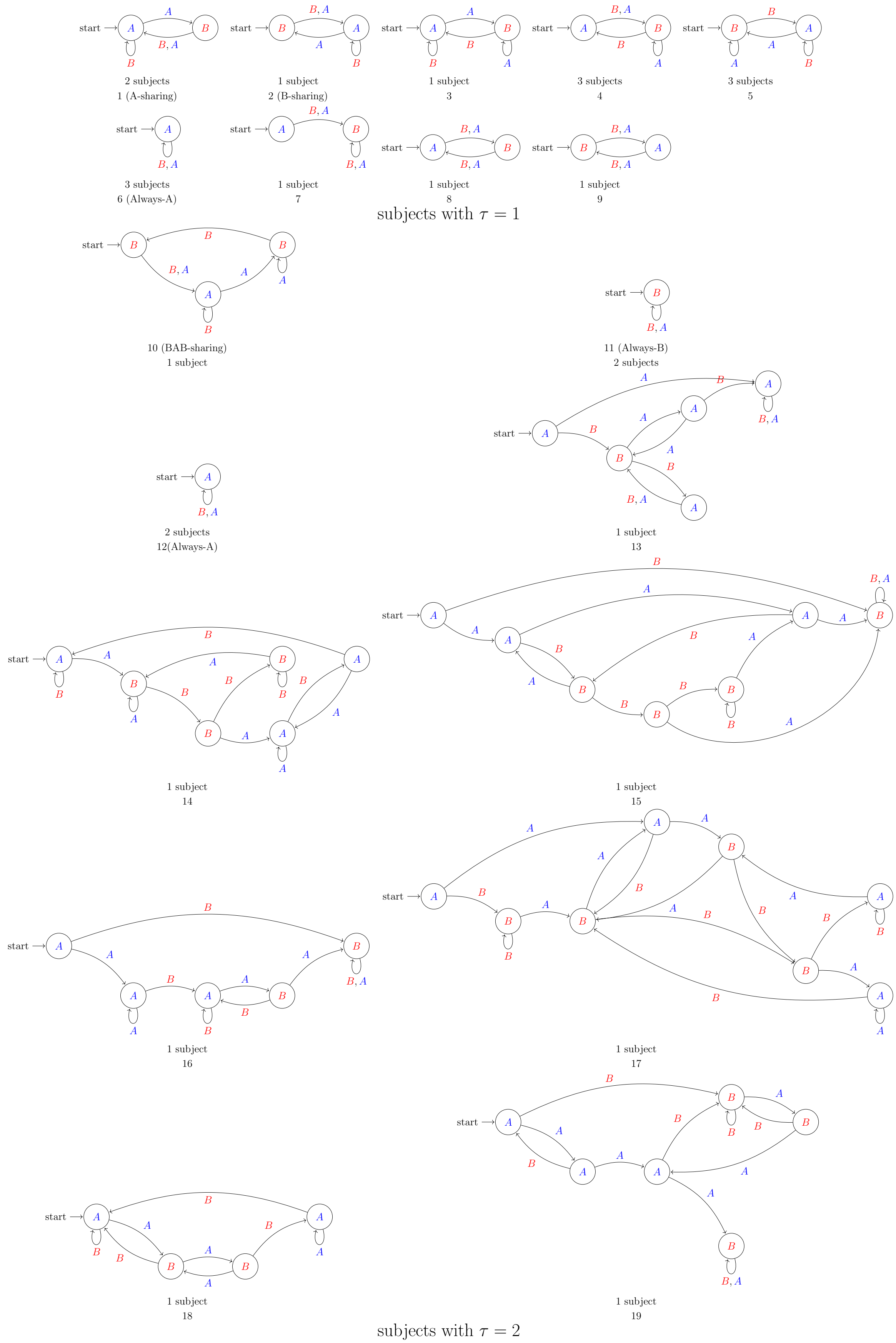


FIGURE 8. Revealed Strategies for Battle of Sexes.

(a) 2-player Prisoner's Dilemma

	Pure	Noisy						
Noise level	0	-10%	-5%	-1%	**Actual**	+1%	+5%	+10%
Always-D	70.6%	74.8%	76.1%	81.3%	**81.0%**	80.2%	84.0%	83.3%
Trigger	7.2%	6.6%	5.9%	4.3%	**4.6%**	4.5%	2.9%	4.1%
C-Tit-for-Tat	6.8%	5.2%	4.2%	2.1%	**4.1%**	2.5%	2.3%	2.0%
D-Tit-for-Tat	6.1%	4.9%	6.8%	6.2%	**4.6%**	7.0%	5.4%	5.6%
Other	9.3%	8.5%	7.0%	6.1%	**5.7%**	5.8%	5.4%	5.0%

(b) 3-player Prisoner's Dilemma

	Pure	Noisy						
Noise level	0	-10%	-5%	-1%	**Actual**	+1%	+5%	+10%
Always-D	67.4%	77.1%	84.1%	90.8%	**90.7%**	91.8%	94.3%	94.8%
Trigger	5.3%	2.6%	0.8%	0.2%	**0.2%**	0.0%	0.1%	0.0%
Tit-for-Tat (2 ver.)	2.1%	1.1%	0.5%	0.0%	**0.1%**	0.0%	0.0%	0.0%
Tease-for-Tat	4.7%	5.6%	4.1%	2.9%	**3.0%**	3.0%	2.7%	2.7%
Other (cooperate on P, 4 ver.)	15.2%	9.4%	7.9%	4.3%	**4.4%**	2.9%	1.5%	1.0%

(c) Battle of Sexes

	Pure	Noisy						
Noise level	0	-10%	-5%	-1%	**Actual**	+1%	+5%	+10%
Always-A	24.9%	55.9%	63.0%	67.1%	**69.9%**	67.7%	66.9%	67.6%
A-sharing	23.4%	18.0%	12.7%	10.4%	**9.7%**	7.9%	8.8%	5.9%
B-sharing	8.7%	2.3%	0.8%	4.0%	**2.6%**	4.2%	6.9%	7.6%
Other	13.1%	11.1%	11.4%	8.3%	**7.2%**	10.3%	8.3%	10.2%

TABLE 8. Evolutionary tournament winners: effects of noise (proportion in 1000 tournaments, each runs until a unique winner remains in the population. Evolutionary process (Moran process) without mutations: new automata births proportional to total utility across matches, deaths are uniform)

evidence that would suggest that such sophisticated sharing strategies over more than two periods get established by subjects.

Next, we consider the counterfactual analysis. In both PD treatments we can see that the lower level of errors corresponds to lower probability of unconditional defectors to survive and the higher level of errors increases the probability of unconditional defectors to be the winner. The result behind this is that (conditionally) cooperative strategies are sensitive to the level of errors. Let us note that in 3-PD the tease-for-tat strategy has a stable (but very small) probability of being a winner. This is because tease-for-tat is still a strategy that is very risk-averse and resilient against defectors. According to the same reasoning in 2-PD the D-TFT strategy has a stable probability of surviving (about 5%) regardless of the level of errors.

(a) 2-player Prisoner's Dilemma

Defectors (share of actual)	1	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0
Always-D	80.4%	66.5%	53.6%	37.7%	0.0%
Trigger	5.5%	9.0%	12.8%	21.4%	34.7%
C-Tit-for-Tat	3.6%	5.0%	8.5%	11.6%	18.1%
D-Tit-for-Tat	4.6%	9.3%	11.0%	14.0%	18.1%
Other	5.9%	10.2%	14.1%	15.3%	29.1%

(b) 3-player Prisoner's Dilemma

Defectors (share of actual)	1	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0
Always-D	92.2%	79.5%	71.6%	60.1%	0.0%
Trigger	0.1%	0.7%	0.3%	0.4%	3.1%
Tit-for-Tat (2 ver.)	0.1%	0.2%	0.2%	0.3%	1.4%
Tease-for-Tat	2.5%	6.8%	10.5%	15.3%	40.6%
Other (cooperate on P, 4 ver.)	3.6%	8.0%	10.6%	14.9%	33.2%

(c) Battle of Sexes

Always-A (share of actual)	1	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0
A-sharing	7.7%	11.9%	16.9%	14.5%	20.5%
B-sharing	3.6%	5.7%	5.9%	5.3%	7.8%
Always-A	61.8%	46.6%	35.8%	34.9%	0.0%
Other	12.3%	15.9%	20.0%	20.6%	36.9%

Table 9. Evolutionary tournament winners: effects of defectors (proportion in 1000 tournaments, each runs until a unique winner remains in the population. Evolutionary process (Moran process) without mutations: new automata births proportional to total utility across matches, deaths are uniform)

Table 9 presents the results for the counterfactual analysis on decreasing the number of unconditional defectors in PD and always-A actions in BoS. Decreasing level of defectors in 2-PD results in quite equal split in terms of probability of winners between Triggers and Tit-for-tatters. However, probability that tit-for-tatters win is more responsive for the smaller changes in the share of defectors, while the largest change in the probability of winning for trigger is when the defectors are totally eliminated. Decreasing level of defectors in 3-PD results in tease-for-tatters being most frequent winners. Note that the major difference in this case is that population of tease-for-tatters results in defection being an outcome, even when all unconditional defectors are eliminated. Changing the level of players who always choose A in BoS results in increasing the probability that A-and B-sharing strategies would result in being winners. Hence, these strategies would results in approximation of the fair split of the surplus.

B.2 Observational Data Likelihood ratio tests indicate that for all firms, allowing for deeper memory does not increase predictive power. That is - the restricted model (smaller τ) can not be rejected in favor of unrestricted (larger τ) for any $\tau > 1$ for all firms and for all combinations of inputs. We therefore only report the Vuong test results for $\tau = 1$. The results for other τ are similar. The table below uses power-corrected values, but the difference with other thresholds is minimal, e.g. 1 p.p at most, compared to a fixed threshold of 0.1.

Table 10 presents the results for the results of individual-level likelihood ratio tests for different combination of different history determinants by different time blocks. We concentrate on the strategies of memory one given the results we presented above the table. Moreover, the results of the likelihood ratio test confirm the findings presented in the main text. That is in the time blocks II, III and IV theories which include day-of-the-week as at least as one of the dimensions of the history determinants explains the data better than theories which do not include this determinant. This observation can serve as the evidence that firms learn to cooperate on the day of the week. However, while Byrne and De Roos (2019) claim that they only manage to do so in the last time block, we observe that appearance of tacit collusion happens earlier (blocks II or III).

	All	BP+M	M+D	BP+D	D	M	BP
I.							
All		0%	2%	0%	3%	1%	0%
BP Market	0%		1%	0%	1%	3%	0%
Market Day	0%	0%		0%	0%	0%	0%
BP Day	0%	0%	2%		3%	1%	0%
Day	0%	0%	0%	0%		0%	0%
Market	0%	0%	0%	0%	0%		0%
BP	0%	0%	1%	0%	1%	3%	
II.	All	BP+M	M+D	BP+D	D	M	BP
All		59%	8%	0%	8%	69%	59%
BP Market	0%		2%	0%	2%	29%	0%
Market Day	0%	50%		0%	0%	63%	50%
BP Day	0%	59%	9%		8%	69%	59%
Day	0%	50%	0%	0%		63%	50%
Market	0%	0%	0%	0%	0%		0%
BP	0%	0%	2%	0%	2%	29%	
III.	All	BP+M	M+D	BP+D	D	M	BP
All		75%	6%	5%	10%	76%	81%
BP Market	0%		1%	2%	3%	13%	26%
Market Day	0%	74%		2%	4%	75%	80%
BP Day	0%	74%	3%		3%	75%	79%
Day	0%	74%	0%	0%		74%	78%
Market	0%	0%	0%	1%	1%		21%
IV.	All	BP+M	M+D	BP+D	D	M	BP
BP	0%	0%	0%	0%	0%	5%	
All		79%	1%	12%	0%	79%	79%
BP Market	0%		0%	0%	0%	10%	18%
Market Day	0%	79%		0%	0%	79%	79%
BP Day	0%	79%	0%		0%	79%	79%
Day	0%	79%	0%	0%		79%	79%
Market	0%	0%	0%	0%	0%		14%
BP	0%	0%	0%	0%	0%	2%	

TABLE 10. Share of subjects for whom hypothesis that row model explains data better than the column model can be rejected at 10% level by time blocks