Long-term Causal Effects of Interventions in Multiagent Economic Mechanisms

Panos Toulis* and David C. Parkes**

*Department of Statistics, Harvard University
** School of Engineering and Applied Science, Harvard University

June 12, 2015

Abstract

The effect of an intervention in an economic mechanism, for example an increase in the reserve price of an auction, is *causal* if the observed effect is better than the counterfactual, i.e., the effect that would be observed under no intervention. As mechanisms are populated by dynamical systems of interacting agents, their response to an intervention fluctuates until the system reaches a new equilibrium. Effects measured in the new equilibrium, the *long-term causal effects*, are more representative of the value of interventions. However, the statistical estimation of long-term causal effects is difficult because it has to rely, for practical reasons, on data observed before the new equilibrium is reached. Furthermore, agent actions do not only depend on the mechanism that the agents are situated in but also on the behavior of others, which complicates the causal evaluation. In this paper, we formalize this problem of estimating long-term causal effects under the *potential outcomes* framework of causal inference [17, 21]. We develop an estimation method that relies on a data augmentation strategy, where agents are assumed to adopt, at each timepoint, a behavior that is latent. This allows us to leverage existing work in behavioral game theory and time-series analysis of compositional data. Our method identifies the long-term causal effects under a set of assumptions that we formulate explicitly. We illustrate our method on a dataset from a real-world behavioral experiment, and discuss open problems to stimulate future research.

1 Introduction

Interventions in an economic mechanism have an effect on its performance. For example, raising the reserve price of an auction has an effect on its revenue, or modifying the rules of a matching procedure for medical residency has an effect on students' incentives. The effect from the intervention is causal if the observed effect is better than the counterfactual, i.e., the effect that would be observed under no intervention. Rigorous empirical evaluation of interventions requires a well-planned experiment and a causal analysis of the observed

outcomes. In this work, we consider problems where the experimental units are agents, the intervention corresponds to a chance in design of an economic mechanism (or game), and the outcomes are agent actions. For instance, in estimating the causal effect of reserve price on auction revenue, the units are advertisers in the auction, the interventions could be two auction formats with different reserve prices, and the outcomes are advertiser bids, aggregated in a way to define the auction revenue.

As argued by R.A. Fisher, the evaluation of intervention effects hinges on an unambiguous interpretation of all possible experiment outcomes, and so "it is always needful to forecast all possible results of the experiment" [11]. However, in multiagent systems, such forecast is complicated by *strategic interference* and *dynamic agent actions*. Strategic interference exists because agents change their actions depending on the behavior of other agents. In an auction, how advertisers bid (and thus the revenue) can be expected to depend on the structure of competition. Interference limits the inferential power of an experiment because the observed outcomes do not provide information about *counterfactuals*, i.e., how agents would act if they were randomized differently into auction environments.[1] The standard method for causal inference assumes away the possibility of interference [22], or assumes that interference arises from a static network of units and employ randomization designs on networks, such as cluster randomization [25], or sequential randomization [24].

Dynamic agent actions present a second challenge to causal inference in multiagent systems, with agents possibly changing their actions dynamically in response to the actions of others. This is important to handle because we are interested to estimate *long-term causal effects*, i.e., to evaluate an intervention after agents have adopted some sort of equilibrium behavior. For instance, raising the reserve price in an auction might increase revenue in the short-run but as agents adapt their bids, or switch to another platform altogether, the long-term effect could be a net decrease [15].

We have three goals in this paper. First, we want to formalize the problem of causal inference in dynamic multiagent systems (Section 2). For that, we work under the *potential outcomes* framework for causal inference [17, 21]. Second, we develop a method for estimation of the defined long-term causal effects (Section 3). Our proposed method relies on a data augmentation strategy, where agents are assumed to adopt, at each timepoint, a behavior that is latent. A game-theoretic model defines the distribution of the actions an agent takes, conditional on the adopted behavior. A temporal model defines the temporal evolution of aggregate agent behaviors. The two models are combined and are fit using short-term data from observed agent actions. The fitted model parameters are then used to predict the long-term agent actions and, thus, estimate the long-term causal effect of interest.

The third goal is to explicitly define a set of assumptions under which our method identifies the long-term causal effect of interest. Strategic interference is alleviated because the assumptions are stated on *aggregate* agent behaviors, whereas individual agent behaviors are allowed to change dynamically in an arbitrary way. Agent behaviors are modeled through a combination of behavioral game theory and time-series models on the latent behavioral space.

[1] The related notion of *equilibrium effects* has received attention in the econometric literature [14, 13]. For example, Athey et. al. [4] developed a method to compare two formats of U.S. timber auctions. Although not causal, their estimates do capture the notion of long-term causal effects.

We apply the methodology to a real-world dataset from a behavioral experiment by Rapoport and Boebel [20]. In this application we combine an instance of the *quantal k-level* (QLk) introduced by Stahl and Wilson [23] as the model from behavioral game theory, with a simple *Vector Autoregressive Model* (VAR) as the temporal model. We use the method to estimate long-term causal effects of changes in the game format on a design objective under a Bayesian mode of inference.

1.1 Related work

The basic approach for estimation of long-term causal effects is to simply assume away strategic interference and dynamic agent actions. In the potential outcomes framework it would be sufficient to make the *stable unit treatment value assumption* (SUTVA) [22], [16, Chapter 3].[2] A more sophisticated approach is to analyze the agent actions as a time series. For example, Brodersen et. al. [6] develop a method to estimate the effects of ad campaigns on website visits. Their method is based on the idea of "synthetic controls", i.e., they create a time-series using different sources of information that would act as the counterfactual to the observed time-series before and after the intervention. However, their problem is macroeconometric and they work with observational data. Thus, there is neither randomized assignment to games, nor strategic interference between agents, nor dynamic agent actions. More crucially, they do not study long-term, equilbrium effects. By construction, in our problem we can leverage behavioral game theory to make, arguably, more informed predictions of counterfactuals to time points after the intervention at which equilibrium has been restored.

An approach that is commonly adopted in econometrics is the *difference-in-differences* (DID) estimator [7, 10, 18]. The DID estimator compares the difference in outcomes before and after the intervention for both the treated and control groups. DID would be unbiased for long-term causal effects if the dependence of agent actions on time and on assignment to a game had an additive structure [1], [3, Section 5.2]. This additivity assumption is stronger than the assumptions we formulate in Section 5. A different approach adopts an assumption that bidders play a game-theoretic equilibrium and assumes assumption of play in observed outcomes and thus does not handle dynamic agent actions and an incomplete observation period [4].

Another approach to causality is through *directed acyclical graphs* (DAGs) [19]. For example, Bottou et. al. [5] study the causal effects on revenue of interventions on the machine learning algorithm that scores online ads in the Bing search engine. Their approach is to create a full DAG among related variables, such as queries, bids, and prices. Through a Causal Markov assumption they can predict counterfactuals for revenue. A key assumption is that the underlying structural equation model remains stable under these manipulations (or interventions), and only edges coming from parents of the manipulated variable need to be removed. As pointed out by Dash [9], this can be implausible when intervening in equilibrium systems.[3]

[2]Under SUTVA an agent's actions over time would remain fixed and would depend only on the agent's assignment to a game (e.g., auction format). Although typical in statistical practice, SUTVA is not sensible for the estimation of long-term causal effects.

[3]Consider, for example, a DAG $X \to Y \leftarrow Z$, and a manipulation that sets the distribution of Y

2 Definition of Long-term Causal Effects

Assume a set of agents $\mathcal{I}$ indexed by i, a set of games $\mathcal{G}$ indexed by j, and a set of actions $\mathcal{A}$ indexed by a. The choice of 'game' corresponds to a design decision about an economic environment. For game j, G_j will denote the characteristics of that game that are relevant, e.g., payoff matrix, rules, and so forth. For example, in an online ad auction the agents can be advertisers, the set of games can be different auction formats (e.g., first-price or second-price auction), and the actions are bids in the auction. The designer wants to run an experiment to select the best game from $\mathcal{G}$ according to some objective R, for example the revenue of an auction format.

In the experiment, each agent is assigned to one game, and the experimenter observes agent actions over time. The objective R depends on agent actions, e.g., revenue depends on agent bids. Ideally, the game designer wants to know, for every game $j \in \mathcal{G}$, how agents *would* bid, and thus what the objective value would be, in the long-term and if *all* agents were assigned to game j.

To formalize, let Z be the $|\mathcal{I}| \times 1$ assignment vector where the ith element Z_i denotes the assignment of agent i to a game; i.e., $Z_i = j$ if agent was assigned to game j. Let $\mathbf{1}$ be the $|\mathcal{I}| \times 1$ vector of ones, then $Z = j\mathbf{1}$ indicates that all agents are assigned to game j. The assignment of agents to games is made uniformly at random such that the marginal probability of each agent being assigned to any game is constant across agents. For simplicity, we assume that the same number of agents is assigned to each game (this is not crucial).

After assignment, the games proceed simultanteously at disecrete time steps indexed by t, from $t = 0$ to $t = t_o - 1$, for some number of data observation steps t_o. The action that agent i *would take* at time t under assignment Z is denoted by $A_{it}(Z)$; the entire vectior of agent actions under assignment Z will be denoted by $A_{i*}(Z) = (A_{i0}(Z), A_{i1}(Z), \ldots, A_{it_o-1}(Z))^\intercal$. Let Δ^p denote the p-dimensional simplex. The *aggregate action* $\alpha_{j,t}(Z) \in \Delta^{|\mathcal{A}|}$ in game j at time t under assignment Z is the frequency of actions $A_{it}(Z)$ of agents assigned to game j; i.e., if $\mathcal{I}_j = \{i \in \mathcal{I} : Z_i = j\}$ is the set of agents assigned to game j, then the ath element of $\alpha_{j,t}(Z)$ is equal to $\sum_{i \in \mathcal{I}_j} \mathbb{I}\{A_{it}(Z) = a\}/|\mathcal{I}_j|$. As before, $\alpha_{j,*}(Z)$ denotes the $|\mathcal{A}| \times t_o$ matrix where $\alpha_{j,k}(Z)$ is the $(k+1)$th column.

The experimenter uses an objective, denoted by R, such that the objective value in game j at time t under assignment Z is $R_{j,t}(Z) = h(\alpha_{j,T}(Z))$, for an appropriate function $h : \Delta^{|\mathcal{A}|} \to \mathbb{R}$. To compare between any two games, say j and j', the experimenter can compare their objective values at some period T that is considered long-run. As the experimenter is interested to adopt only one economic environment after the experiment is done, these objective values need to be calculated at assignments where all agents are assigned to each candidate game. This leads to the folowing definition of *long-term causal effects*.

Definition 2.1. *The causal effect at time T on objective R of game j over game j' is equal to the quantity*

$$\tau(j, j'; T) = R_{j,T}(j\mathbf{1}) - R_{j,T}(j'\mathbf{1}) = h(\alpha_{j,T}(j\mathbf{1})) - h(\alpha_{j,T}(j'\mathbf{1})). \tag{1}$$

independently of X, Z. Then after the manipulation the two edges will need to be removed. However, if in an equilibrium it is required that $Y \approx XZ$, then the two arrows should be reversed after the manipulation. The interpretation of the *Do* operator in equilibrium systems remains an open area without a well-established methodology [8].

The causal effects τ in Definition 2.1 for all pairs j, j' are the *estimands*, i.e., the quantities that the experimenter needs to know to decide which candidate game is the best according to the objective. The estimands capture how agents bid if all are assigned to one single game. However, in any given experiment agents are randomly assigned to games, and not all agents will be assigned to one game. In particular, under the assignment Z in the experiment, we can only observe outcomes (agent actions) $A_{it}(Z)$, for every agent i and from time $t = 0$ to $t = t_o - 1$; all outcomes under different assignments and at different times will be *missing*. The challenge of causal inference is thus to predict these missing outcomes.

3 Method and Assumptions for Estimation of Long-term Causal Effects

Given the definition of long-term causal effects (1), the key challenge is to predict the missing aggregate action $\alpha_{j,T}(j\mathbf{1})$ given only the aggregate actions $\alpha_{j,*}(Z)$, observed under assignment Z in the experiment. This prediction is along two dimensions. First, for every game j we need to extrapolate from the initial assignment Z to assignment $Z = j\mathbf{1}$, i.e., to actions that would be observed had all agents been assigned to play j. Second, given observed data up to some time t_o, we need to extrapolate to $t = T$. The former is the problem of strategic interference, whereas the latter is the problem of dynamic agent actions.

To evaluate such predictions necessarily requires additional assumptions. To illustrate with a simple example, assume that we use the average of aggregate actions $(1/t_o)\sum_{t=0}^{t_o-1}\alpha_{j,t}(Z)$ to estimate the long-term aggregate action $\alpha_{j,T}(j\mathbf{1})$. A sufficient condition for this estimator to be unbiased is to assume $A_{it}(Z) \equiv A_i(Z_i)$, i.e., assume that the action of an agent is the same across time, and depends only on its assignment. We prove this result in Appendix 8.2. This assumption is equivalent to the *stable unit treatment value assumption (SUTVA)* [22], typically used in the causal inference literature. However, SUTVA is not reasonable in our setting because of strategic interference (the assignment determines what competition each agent is facing) and because of dynamics (the agent actions may depend on time as agents adapt to the environment.)

Our approach to resolve these issues is to use a data augmentation strategy where we assume that agents adopt behaviors over time that are latent. A *behavior* is a distribution over actions, and we assume there exists a finite set of behaviors $\mathcal{B} = \{1, 2, \ldots, |\mathcal{B}|\}$ indexed by $b \in \Delta^{|\mathcal{A}|}$. In the auction application, the behavior— formally a distribution over bids —could indicate, for example, how aggressive an agent is as a bidder. In general, we think of behaviors as summarizing patterns of agent actions.

As games proceed, the behavior that agent i *would adopt* at time t under assignment Z is denoted by $B_{it}(Z) \in \mathcal{B}$; the entire vector of agent behaviors under assignment Z will be denoted by $B_{i*}(Z) = (B_{i0}(Z), B_{i1}(Z), \ldots, B_{it_o-1}(Z))^\intercal$. The *aggregate behavior* $\beta_{j,t}(Z) \in \Delta^{|\mathcal{B}|}$ in game j at time t under assignment Z is the frequency of actions $B_{it}(Z)$ of agents assigned to game j; i.e., if $\mathcal{I}_j$ is the set of agents assigned to game j, then the bth element of $\beta_{j,t}(Z)$ is equal to $\sum_{i\in\mathcal{I}_j}\mathbb{I}\{B_{it}(Z) = b\}/|\mathcal{I}_j|$. Finally, $\beta_{j,*}(Z)$ denotes the $|\mathcal{B}| \times t_o$ matrix where $\beta_{j,k}(Z)$ is the $(k+1)$th column.

Lemma 3.1. *For every game j there exists a left-stochastic, $|\mathcal{A}| \times |\mathcal{B}|$ matrix P_j such that*

$$\mathbb{E}\left(\alpha_{j,t}(Z) | \, \beta_{j,t}(Z)\right) = P_j \beta_{j,t}(Z), \tag{2}$$

for any time t and assignment Z.

Lemma 3.1 establishes a linear relationship between the aggregate action and the latent aggregate behavior. The matrix P_j depends on game j and possibly depends on parameters that need to be estimated. For example, in Section 5 we will use a specific game-theoretic model for which P_j can be analytically computed, and has parameters that correspond to the sophistication of the behaviors that agents can adopt.

Lemma 3.1 transforms the problem of estimating $\alpha_{j,T}(j\mathbf{1})$ to estimating $\beta_{j,T}(j\mathbf{1})$. We decide to work on the latent behavioral space because our strategy for estimation of long-term causal effects will be based on the following high-level algorithm:

Algorithm 1 Estimation of long-term causal effects

1: Model the stochastic process $\beta_{j,t}(Z), t = \{0, 1, \ldots\}$ *independently* of the assignment Z. Estimate the model parameters given observed data $\alpha_{j,*}(Z)$.
2: Concurrently with (1), estimate the conditional distribution of the initial aggregate behavior $\beta_{j,0}(Z)$ under assignment Z given observed data $\alpha_{j,*}(Z)$.
3: Use distribution (2) to estimate the initial aggregate behavior when $Z = j\mathbf{1}$.
4: Use estimates of (1) to estimate the distribution of the long-term aggregate behavior $\beta_{j,T}(j\mathbf{1})$.
5: Use Lemma 3.1 to estimate the distribution of long-term aggregate actions $\alpha_{j,T}(j\mathbf{1})$.

Using Algorithm 1 for both games j and j' can provide estimates for the long-term causal effect $\tau(1)$. For the rest of this section we will derive sufficient assumptions on the aggregate behavior of agents in the experiment, for which Algorithm 1 will identify the long-term causal effects (1).

Assumption 3.1 (Initial behaviors)**.** *Let $\beta_{j,-1}(Z)$ be the aggregate behavior in game j under assignment Z before the assignment is done at $t = 0$. Then, for every game j and assignment Z,*

$$\mathbb{E}\left(\beta_{j,-1}(Z)\right) = \beta^{(0)}.$$

Under Assumption 3.1, at time $t = -1$ (before the assignment), every agent samples a behavior for $t = 0$ independently from a fixed but unknown aggregate behavior $\beta^{(0)} \in \Delta^{|\mathcal{B}|}$. This can be motivated by supposing that the economic environment to which an agent is assigned is not understood by an agent until the agent takes actions within the environment.

Assumption 3.2 (Behavioral ignorability)**.** *For a given assignment Z, let $\mathcal{F}_t$ be the filtration that the process $\beta_{j,t}(Z)$ is adapted to. Under assignment Z the distribution of aggregate behavior in game j at time t, $0 \leq t \leq t_o$, is independent of Z conditional on the characteristics G_j of the game and history of behaviors until $t-1$; i.e.,*

$$Z \perp\!\!\!\perp \beta_{j,t}(Z) \mid \mathcal{F}_{t-1}, G_j.$$

Assumption 3.2 implies that the assignment Z does not add information about the aggregate behavior at time t given the information provided by the aggregate behaviors up to time $t-1$; thus, conditional on the history up to $t-1$, the particular assignment is *ignorable*.[4]

Assumption 3.3 (Temporal model of behaviors)**.** *For a given assignment Z and game j, let $\mathcal{F}_t$ be the filtration that the process $\beta_{j,t}(Z)$ is adapted to. For a known prior π and observation model f, there exist parameters $\theta = (\phi, \psi)$ such that*

$$\beta_{j,0}(Z) \sim \pi(\cdot; \phi)$$
$$\beta_{j,t}(Z) |\ \mathcal{F}_{t-1} \sim f(\cdot|\psi, \mathcal{F}_{t-1}),$$

Assumption 3.3 implies that there exists an underlying model for the temporal evolution of aggregate agent behaviors. The model– the prior π and observation model f –is known but its parameters $\theta = (\phi, \psi)$ are unknown. The values of those parameters may depend on the game j *as well as* the assignment Z.

Theorem 3.1 (Estimation of long-term effects)**.** *Suppose that Assumptions 3.1, 3.2 and 3.3 hold. Then, Algorithm 1 identifies the long-term causal effect* (1) *if parameters $\theta = (\phi, \psi)$ of Assumption 3.3 can be identified as the data observation period $t_o \to \infty$.*

3.1 Discussion

We have explicitly stated assumptions that we use in obtaining our estimation result. Each assumption plays an important role. Without Assumption 3.1, the initial aggregate behavior under assignment Z would not provide an unbiased estimate of the aggregate behavior in period $t = -1$. Without Assumption 3.2 we would not be able to borrow information from the observed behavior time series to the counterfactual one, which is crucial for causal inference. Without Assumption 3.3 we would not be able to make any inference from the observation prior $[t, t_o]$ to some later time T.

4 Application

Rapoport and Boebel [20] conducted a behavioral experiment on a zero-sum, two-agent game. The game was a simultaneous-move game with ten discrete actions, namely $\mathcal{A} = \{a_1, a_2, a_3, a_4, a_5, a'_1, a'_2, a'_3, a'_4, a'_5\}$. The structure of the payoff matrix is given in the Appendix (Table 1). It is parametrized by two values, W and L. The experiment used two different versions of the game corresponding to payments by the row agent when it *won* (W),

[4]This assumption is related to *policy invariance* assumptions adopted for the econometrics of policy effects [13, 14], where, given the *choice* of policy by an agent, the initial process that resulted in this choice does not affect the outcome. For example, given that an individual chooses to participate in a tax benefit program, the way the individual was assigned to the program (e.g., lottery, recommendation, or point of a gun) does not alter the outcome that will observed for that individual. Our assumption is different, because we have a temporal evolution of aggregate behavior and there is no free choice of an agent about the assignment. But it shares the essential aspect of ignorability of assignment under appropriate conditions.

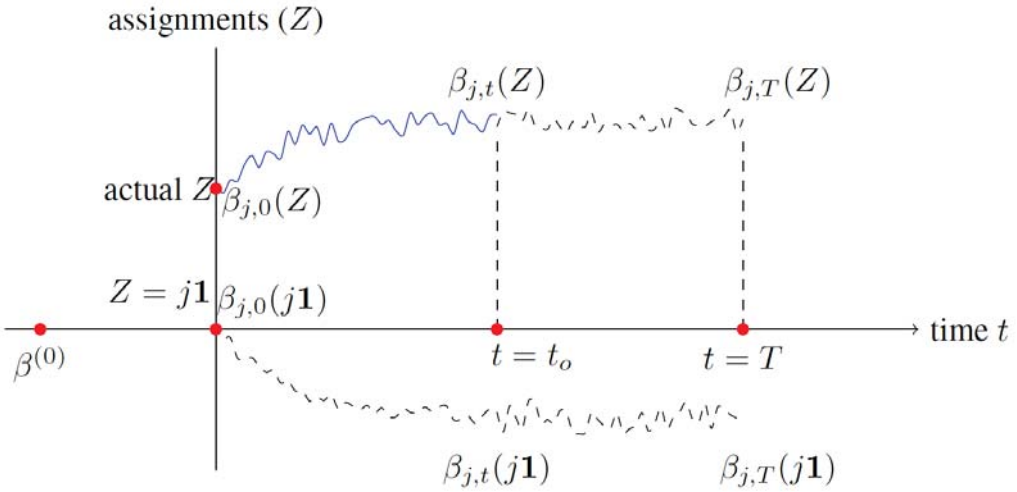

Figure 1: Graphical depiction of Algorithm 1. The vertical line represents the space of assignments Z. For each assignment Z there is a separate dynamic $\beta_{j,t}(Z)$. The goal of inference is to predict the behaviors $\beta_{j,T}(Z)$ at $t = T$ under a hypothetical assignment $Z = j\mathbf{1}$, i.e., when all agents are assigned to game j. Our strategy is to learn the parameters of the time series $\beta_{j,t}(Z)$ under the actual assignment Z using the aggregate action data from $t = 0$ to $t = to$ (blue line). Then, we estimate the distribution of the initial aggregate behavior $\beta_{j,0}(Z)$ under assignment Z. Under randomization, this unbiasedly estimates the initial behaviors $\beta^{(0)}$ adopted by agents before any assignment. Since, $\beta_{j,0}(j\mathbf{1}) = \beta^{(0)}$, we can then use the aforemenioned time series parameters to estimate $\beta_{j,T}(j\mathbf{1})$ starting from $\beta_{j,0}(j\mathbf{1})$. Assumptions 3.1-3.3 are sufficient conditions for this estimation to be unbiased.

or *lost* (L): they used $(W, L) = (\$10, -\$6)$ for game 1 and $(W, L) = (\$15, -\$1)$ for game 2.[5] In our notation, $\mathcal{G} = \{1, 2\}$ is the set of games.

Forty subjects (agents), $\mathcal{I} = \{1, 2, \ldots, 40\}$, were randomized to each game (20 subjects per game), and each agent played both as row agent and as column agent in a match-up with two different agents. Every match-up lasted two sessions (periods) of 60 rounds, where each round consisted of a selection of a strategy from each agent and a (possible) payoff. The aggregate data for the experiment are given in Table 2 of the Appendix, which reports the distribution of actions adopted by agents within each period.

Although their experiment had a different purpose, we can adapt it to apply our causal methodology. For this we define a simple linear objective function. Given this, we ask the following question: "What is the long-term causal effect on the revenue of the game if we switch from $(W, L) = (\$10, -\$6)$ of game 1 to $(W, L) = (\$15, -\$1)$ of game 2?". To evaluate our method, we will consider period four as long-term, and hold out data on that period. Thus, we wish to estimate the revenue of each game in period four *had all agents* been assigned to that game. For simplicity, we define the estimand (1) as

$$\tau = c^{\intercal}(\alpha_{2,T}(\mathbf{1}) - \alpha_{1,T}(\mathbf{0})), \tag{3}$$

where c is a 10×1 vector of uniform $(0, 1)$ numbers; given an element c_a, the agent playing

[5]For example, if the row agent picks action a_1 and the column agent plays a_3' in the first version, then the row agent has to pay \$6 to the column agent.

action a is assumed to pay a constant fee c_a to the game designer. In Section 6 we will compare our method to other related methods on multiple random values of the vector c.

5 Concrete Methodology

Building upon the insights of Section 3, we develop a concrete methodology to estimate long-term causal effects defined in (1). In Section 6 we apply this concrete methodology to the dataset by Rapoport and Boebel [20].

The first component of our method is the game-theoretic model that maps behaviors to distribution of actions. For that, we adopt the *quantal k-response* (QL_k) model that was initially proposed by Stahl and Wilson [23]. The QL_k model has been shown to predict well observed human behavior in several real-world behavioral experiments [26]. In our setting, QL_3 implies that $\alpha_{j,t}(Z)$ follows a multinomial distribution conditional on $\beta_{j,t}(Z)$ with expectation

$$\mathbb{E}\left(\alpha_{jt}(Z)|\,\beta_{jt}(Z)\right) = \Pi_j(\lambda)\cdot\beta_{jt}(Z), \tag{4}$$

where $\lambda = (\lambda_1, \lambda_{1(2)}, \lambda_2)$ are the *precision* parameters of the model, and $\Pi_j(\lambda)$ is a $|\mathcal{A}| \times |\mathcal{B}|$ matrix that depends on λ. The exact derivation for $\Pi_j(\lambda)$ is given in Appendix 8.7. This agrees with Lemma 3.1 for $P_j = \Pi_j(\lambda)$.

Intuitively, when an agent has high precision it is striving for better expected utility; when the precision is zero, the agent picks an action at random. In QL_3 there are three behaviors of increasing sophistication: the first, simplest behavior has precision zero; λ_1 is the precision of the second behavior, $\lambda_{1(2)}$ is the precision of the second behavior as *perceived* by agents adopting the third behavior, and λ_2 is the precision of the third, most sophisticated behavior.

The second component is the temporal behavior model according to Assumption 3.3. For simplicity, we adopt a simple VAR(1) model for $\beta_{j,t}(Z)$. As is typical in the analysis of time-series of compositional data [2, 12], we transform the proportion into a new variable $w_{j,t} \stackrel{\text{def}}{=} \mathrm{logit}(\beta_{jt}(Z))$ and assume that

$$w_{j,t} = \psi_0 w_{j,t-1} + \mu + \psi_1 \epsilon_t, \tag{5}$$

where $\psi_0 \in (0,1), \psi_1 \in \mathbb{R}^+$, and $\mu \in \mathbb{R}^{|\mathcal{B}|-1}$ is a fixed parameter vector and $\epsilon_t \sim \mathcal{N}(0, I)$ is i.i.d. standard normal. The prior on $\beta_{j,0}(Z)$ is a Dirichlet on $\Delta^{|\mathcal{B}|}$ with parameter ϕ.

We can now write down the likelihood for our model. Our likelihood parameters are λ for the QL_3 model and ψ, ϕ for the VAR model including the prior. Consider an initial assignment Z, and the observed data of aggregate actions $\alpha_{j*}(Z)$ at times $t = 0, 1, 2$. The marginal likelihood over the latent aggregate behaviors $B \stackrel{\text{def}}{=} \beta_{j,*}(Z)$ is given by

$$\mathcal{L}(\lambda, \psi, \phi; \alpha_{j*}(Z)) = \int_B \left(\prod_{t=0}^{2} g(\alpha_{jt}(Z)|\beta_{jt}(Z), \lambda)\right) \times h(B|\psi) dB. \tag{6}$$

The term $g(\alpha|\beta, \lambda)$ is the probability of observing aggregate action α given aggregate behavior β and QL_3 parameters λ. This is a simple multinomial $\mathrm{Multinom}(\widehat{\alpha}; N)$, where

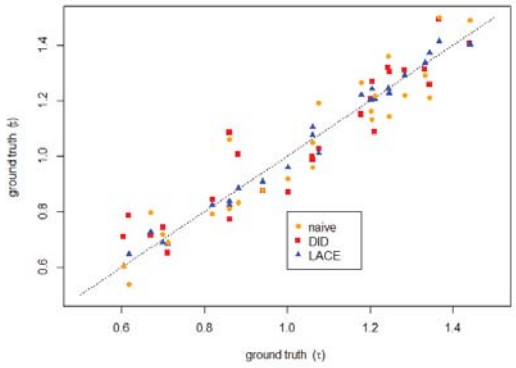

Table 1: Posterior estimates for a subset of parameters

param.	mean (sd)	$[Q_1, Q_3]$
λ_1	.15 (0.05)	[0.1, 0.3]
$\lambda_{(1)2}$	3.1 (0.6)	[2.1, 3.7]
λ_2	1.9 (1.2)	[0.9, 2.8]
$\psi_{0,1}$	0.8 (0.2)	[0.5, 0.9]
$\psi_{0,2}$	0.47 (0.3)	[0.3, .6]

Figure 2: *Left.* Estimates of long-term effects of different methods corresponding to 25 random objective functions. For our estimates (LACE) we sampled 100 times from the posterior predictive distribution of τ and then kept the median. *Right:* Posterior intervals for certain model parameters.

$\widehat{\alpha} = \Pi_j(\lambda)\beta$ and N is the number of agents. The term $h(B|\psi)$ can be decomposed into the product $\prod_{t=1}^{2} f(\beta_{jt}(Z)|\beta_{jt-1}(Z), \psi)\text{Dir}(\beta_{j,0}(Z); \phi)$, where the conditional densities can be computed from model (5), and Dir is the Dirichlet density.

6 Results

We now fit our model to the dataset of Rapoport and Boebel [20] (shown in Table 2 of the Appendix). Details on the priors and MCMC are given in Appendix 8.8. The results are showin in Figure 2.

There are a few interesting observations. First, the model obtains $\lambda_2 > \lambda_1$, i.e., that level-2 agents have better precision than agents at level-1. Interestingly, in this dataset, level-2 agents play as if level-1 agents are very precise (see values for $\lambda_{(1)2}$ in Table of Figure 2). Estimates on vector ψ_0, i.e., the coefficient in the VAR model are significant, indicating a temporal trend in the latent behavioral state. Furthermore, the posterior samples for the latent behaviors (not shown in the figure) indicated that the proportion of not sophisticated agents decreased over time, which is evidence that agents were learning the game.

Finally, estimates of the long-term effect τ is shown in the left part of Figure 2. We sampled 25 random uniform vectors c, which we then used to compute the true values of long-term causal effects from Eq. (13). We observe that our method of estimation of long-term causal effects (LACE) is giving better estimates (mse = 0.045) than the estimates from a naive method (mse = 0.185), which uses outcomes only at $t = 1$, and better than estimates from difference-in-differences (DID, mse = 0.361), which uses outcomes at $t = 1$ and $t = 3$. A more detailed discussion of our method in comparison to standard methods is given in the Appendix 8.5.

7 Discussion

In this paper, we explored the problem of estimating long-term causal effects of interventions in multiagent systems under the Neyman-Rubin causal model of potential outcomes. Two features make this problem conceptually and technically challenging. First, strategic interference among competing agents limits inference from randomized experiments. Second, dynamic actions by agents introduces short-term effects, whereas one is typically interested in long-term effects, i.e., when some sort of equilibrium play has been reached.

Our first contribution is to explicate a set of sufficient assumptions for identification of long-term causal effects. We also provide a method that, given the aforementioned assumptions, identifies long-term causal effects. Our method relies on a latent behavioral space, which allows us to leverage existing tools from behavioral game theory to make more informed statistical predictions of counterfactuals. Working on a real-world dataset from a behavioral experiment [20], we showed how our method can be applied for estimation of a long-term effect of an intervention in the payoff structure of a normal-form game.

However, there are several open issues. One important issue is the strategic interference *between* games. In many interesting situations agents compete with each other across games, or switch to other platforms and so on. Another, more theoretical concern, is whether it is possible to establish necessary assumptions for the identification of long-term causal effects. The formalism of DAG theory could be one approach. We believe that progress in answering such questions will lead to new and fruitful interactions of game theory with experimental design and causal inference.

References

[1] Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, **72**(1), 1–19.

[2] Aitchison, J. (1986). *The statistical analysis of compositional data*. Springer.

[3] Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

[4] Athey, S., Levin, J., and Seira, E. (2011). Comparing open and sealed bid auctions: Evidence from timber auctions. *The Quarterly Journal of Economics*, **126**(1), 207–257.

[5] Bottou, L., Peters, J., Quiñonero-Candela, J., Charles, D. X., Chickering, D. M., Portugualy, E., Ray, D., Simard, P., and Snelson, E. (2013). Couterfactual reasoning and learning systems. *J. Machine Learning Research*, **14**, 3207–3260.

[6] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2014). Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*.

[7] Card, D. and Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. *American Economic Review*, **84**(4), 772–793.

[8] Dash, D. (2005). Restructuring dynamic causal systems in equilibrium. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AIStats 2005)*, pages 81–88.

[9] Dash, D. and Druzdzel, M. (2001). Caveats for causal reasoning with equilibrium models. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 192–203. Springer.

[10] Donald, S. G. and Lang, K. (2007). Inference with difference-in-differences and other panel data. *The review of Economics and Statistics*, **89**(2), 221–233.

[11] Fisher, R. A. (1935). *The design of experiments.* Oliver & Boyd.

[12] Grunwald, G. K., Raftery, A. E., and Guttorp, P. (1993). Time series of continuous proportions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 103–116.

[13] Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation1. *Econometrica*, **73**(3), 669–738.

[14] Heckman, J. J., Lochner, L., and Taber, C. (1998). General equilibrium treatment effects: A study of tuition policy. *American Economic Review*, **88**(2), 3810386.

[15] Holland, J. H. and Miller, J. H. (1991). Artificial adaptive agents in economic theory. *The American Economic Review*, pages 365–370.

[16] Imbens, G. W. and Rubin, D. (2009). *Causal inference in statistics, and in the social and biomedical sciences*. Cambridge University Press New York.

[17] Neyman, J. (1923). On the application of probability theory to agricultural experiments, essay on principles, section 9. *Translated in Statistical Science (1990)*, **5**, 465–480.

[18] Ostrovsky, M. and Schwarz, M. (2011). Reserve prices in internet advertising auctions: A field experiment. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 59–60. ACM.

[19] Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge University Press.

[20] Rapoport, A. and Boebel, R. B. (1992). Mixed strategies in strictly competitive games: A further test of the minimax hypothesis. *Games and Economic Behavior*, **4**(2), 261–283.

[21] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**(5), 688.

[22] Rubin, D. B. (1980). Comment. *Journal of the American Statistical Association*, **75**(371), 591–593.

[23] Stahl, D. O. and Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, **25**(3), 309–327.

[24] Toulis, P. and Kao, E. (2013). Estimation of causal peer influence effects. In *Proceedings of The 30th International Conference on Machine Learning*, pages 1489–1497.

[25] Ugander, J., Karrer, B., Backstrom, L., and Kleinberg, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337. ACM.

[26] Wright, J. R. and Leyton-Brown, K. (2010). Beyond equilibrium: Predicting human behavior in normal-form games. In *Proc. 24th AAAI Conf. on Artificial Intelligence*.

8 Appendix

8.1 Introduction

Section 8.2 has a simple proof that the initial aggregate behavior $\alpha_{j,0}(Z)$ is an unbiased estimator of the long-term aggregate behavior $\alpha_{j,T}(j\mathbf{1})$. Section 8.3 has proof for Lemma 3.1 and Section 8.4 has proof for Theorem 3.1 of the paper. Section 8.5 has a more lengthy discussion on related methods, in particular, with respect to the Rapoport and Boebel experiment [20]. Section 8.6 presents the dataset from that experiment. Finally, Section 8.7 has the details regading the QLk model, which use to map behaviors to distributions of actions (game-theoretic model).

8.2 Unbiasedness result under SUTVA

To see this result, the expectation over the randomized assignment Z of the ath element of $\alpha_{j,0}(Z)$ is equal to $\mathbb{E}\left((1/|\mathcal{I}_j|)\sum_{i\in\mathcal{I}_j}\mathbb{I}\{A_{i0}(Z)=a\}\right)$ which gives

$$\begin{aligned}
(1/|\mathcal{I}_j|)\sum_{i\in\mathcal{I}_j}\mathbb{E}\left(\mathbb{I}\{A_{i0}(Z)=a\}\right) &= (1/|\mathcal{I}_j|)\sum_{i\in\mathcal{I}}\mathbb{E}\left(\mathbb{I}\{Z_i=j\}\cdot\mathbb{I}\{A_{i0}(Z)=a\}\right)\\
&= (1/|\mathcal{I}_j|)\sum_{i\in\mathcal{I}}\mathbb{E}\left(\mathbb{I}\{Z_i=j\}\cdot\mathbb{I}\{A_i(j)=a\}\right) \quad [\textit{by SUTVA}]\\
&= (1/|\mathcal{I}_j|)\sum_{i\in\mathcal{I}}\mathbb{E}\left(\mathbb{I}\{Z_i=j\}\right)\cdot\mathbb{I}\{A_i(j)=a\} \quad [\textit{only Z is random}]\\
&= (1/|\mathcal{I}_j|)\sum_{i\in\mathcal{I}}\frac{|\mathcal{I}_j|}{|\mathcal{I}|}\cdot\mathbb{I}\{A_i(j)=a\} \quad [\textit{by complete randomization}]\\
&= (1/|\mathcal{I}|)\sum_{i\in\mathcal{I}}\mathbb{I}\{A_i(j)=a\}. \quad [\textit{by SUTVA}] \qquad (7)
\end{aligned}$$

Therefore, $\mathbb{E}\left(\alpha_{j,0}(Z)\right)=\alpha_{j,T}(j\mathbf{1})$.

8.3 Proof of Lemma 3.1

Lemma. *For every game j there exists a left-stochastic, $|\mathcal{A}|\times|\mathcal{B}|$ matrix P_j such that*

$$\mathbb{E}\left(\alpha_{j,t}(Z)|\,\beta_{j,t}(Z)\right)=P_j\beta_{j,t}(Z), \qquad (8)$$

for any time t and assignment Z.

Proof. By definition, a behavior b is a distribution over actions; thus, let p_{ab} denote the probability that an agent adopting behavior b selects action a. Then

$$\begin{aligned}
P(A_{it}(Z)=a) &= \sum_{b\in\mathcal{B}}P(A_{it}(Z)=a|B_{it}(Z)=b)P(B_{it}(Z)=b)\\
&= \sum_{b\in\mathcal{B}}p_{ab}P(B_{it}(Z)=b). \qquad (9)
\end{aligned}$$

Let $\alpha_{j,t}(Z)^a$ be the ath element of $\alpha_{j,t}(Z)$ and $\beta_{j,t}(Z)^b$ be the bth element of $\beta_{j,t}(Z)$. Then, $\alpha_{j,t}(Z)^a=P(A_{it}(Z)=a)$, and $\beta_{j,t}(Z)^b=P(B_{it}(Z)=b)$. The ath element of $\mathbb{E}\left(\alpha_{j,t}(Z)\right)$ is equal to $P(A_{it}(Z)=a)$. Thefore, from Eq. (9) we get $\alpha_{j,t}(Z)^a=\sum_{b\in\mathcal{B}}p_{ab}\beta_{j,t}(Z)^b$, which leads to (2), where P_j is the matrix with p_{ab} as the element in row a and column b. Since $\sum_{b\in\mathcal{B}}p_{ab}=1$ matrix P_j is left-stochastic. □

8.4 Proof of Theorem 3.1

Theorem (Estimation of long-term effects)**.** *Suppose that Assumptions 3.1, 3.2, and 3.3 hold. Then, Algorithm 1 identifies the long-term causal effect if parameters $\theta = (\phi, \psi)$ of Assumption 3.3 can be identified as the data observation period $t_o \to \infty$.*

Proof. Fix a game j (all model parameters, θ, ψ, ϕ are implicitly assumed to have a subscript j). Furthemore, under Assumption 3.1, all expectations are implicitly conditional on the fixed but unkown aggregate behavior $\beta^{(0)}$.

First, given Assumption 3.2, the parameters θ in the behavioral model of Assumption 3.3 are independent of the assignment. Also, given Assumption 3.3 there exists a density for the aggregate behavior at an arbitrary t conditional on the aggregate behavior of at $t = 0$, that depends only on parameter ψ; i.e., $g(\beta_{j,t}(Z)|\beta_{j,0}(Z), \psi)$. To see this, let $\beta_t = \beta_{j,t}(Z)$ for brevity, and let $B_t = (\beta_1, \beta_2, \ldots, \beta_t)$, i.e., it is the history of aggregate behaviors up to t with $\beta_{j,0}(Z)$ removed. Then, we can derive g explicitly as,

$$
\begin{aligned}
g(\beta_t|\beta_0, \psi) &= \int_{B_{t-1}} g\left(\beta_t, B_{t-1}|\beta_0, \psi\right) dB_{t-1} \\
&= \int_{B_{t-1}} g\left(\beta_t|B_{t-1}\beta_0, \psi\right) g(B_{t-1}|\beta_0, \psi) dB_{t-1} \\
&= \int_{B_{t-1}} f(\beta_t|B_{t-1}, \psi) \times g\left(B_{t-1}|\beta_0, \psi\right) dB_{t-1} \quad [\textit{by Assumption 3.3}] \\
&= \int_{B_{t-1}} f(\beta_t|B_{t-1}, \psi) \times g(\beta_{t-1}|B_{t-2}, \beta_0, \psi) g(B_{t-2}|\beta_0, \psi) dB_{t-1} \\
&= \int_{B_{t-1}} f(\beta_t|B_{t-1}, \psi) \times f(\beta_{t-1}|B_{t-2}, \psi) g(B_{t-2}|\beta_0, \psi) dB_{t-1} \\
&= \ldots \\
&= \int_{B_{t-1}} f(\beta_t|B_{t-1}, \psi) \times f(\beta_{t-1}|B_{t-2}, \psi) \ldots \times f(\beta_1|\beta_0, \psi) dB_{t-1}, \qquad (10)
\end{aligned}
$$

where all the terms are given by the observation model of Assumption 3.3.

Now, given Assumption 3.1 and the complete randomization in the assignment Z, it follows

$$\mathbb{E}\left(\beta_{j,0}(Z)\right) = \beta^{(0)}, \qquad (11)$$

for any game j and assignment Z (recall that all expectations are conditional on $\beta^{(0)}$ under Assumption 3.1). Intuitively, the initial aggregate behavior at $t = 0$ in game/assignment unbiasedly estimates $\beta^{(0)}$ by randomization, since agents (in aggregate) decide their initial behavior independently of any assignment or game. By Assumption 3.3 it follows,

$$\mathbb{E}_{\pi_\phi}(X) = \beta^{(0)}, \qquad (12)$$

where X is a random variable representing the aggregate behavior at time $t = 0$, and distributed according to the prior π_ϕ with parameter ϕ. Given observed data $\alpha_{j,*}(Z) \equiv \alpha_{j,0:t_o-1}(Z)$, assume estimates $\hat{\psi}, \hat{\phi}$ for the model of Assumption 3.3; in the limit $t_o \to \infty$ the true parameter values can be identified, i.e., $\hat{\psi} \to \psi$ and $\hat{\phi} \to \phi$.

Take $\hat{\beta}_0 = \mathbb{E}_{\pi_{\hat{\phi}}}(X)$, where X is distributed according to $\pi_{\hat{\phi}}$. Then, we claim that the density $g(\cdot|\hat{\beta}_0, \hat{\psi})$ unbiasedly estimates the density of $\beta_{j,T}(j\mathbf{1})$. First, note that $\beta_{j,0}(j\mathbf{1})$ is the aggregate

behavior at $t = 0$ when all agents are assigned to game j. Then, by Assumption 3.1, it follows $\beta_{j,0}(j\mathbf{1}) = \beta^{(0)}$. Thus, by Assumption 3.3, the density of $\beta_{j,T}(j\mathbf{1})$ is $g(\cdot|\beta^{(0)}, \psi)$; the density is defined conditional on $\beta^{(0)}$ because, by Assumption 3.1, agents decide to adopt $\beta^{(0)}$ before the experiment starts and independently of Z. By the continuous mapping theorem, $\hat{\beta}_0 \to \beta^{(0)}$ and $\hat{\psi} \to \psi$, and therefore $g(\cdot|\hat{\beta}_0, \hat{\psi}) \to g(\cdot|\beta_0, \psi)$. Finally, the density of $\alpha_{j,T}(j\mathbf{1})$ can be obtained from $g(\cdot|\hat{\beta}_0, \hat{\psi})$ and a simple linear change of variables according to Lemma 3.1. □

8.5 Discussion of related methods

Consider a simple linear estimand for the Rapoport-Boebel experiment [20]:

$$\tau = c^\intercal(\alpha_{j,\mathbf{1}}(T) - \alpha_{j,\mathbf{1}}(T)). \tag{13}$$

In this section, we discuss how standard methods would estimate (13). Our goal is to illustrate the fundamental assumptions underpinning each method, and compare with our Assumption 3.2. To illustrate we will assume a specific value for $c = (0, 1, 0, 1, 0, 0, 0, 0, 1, 1)^\intercal$, chosen adversarially against the other methods. An objective evaluation is given in the experimental section of the paper. In discussing standard methods, we will mostly be concerned with how point estimates compare to the true value of the estimand $\tau = \$0.054$.

The simplest approach would be to consider only the latest time point, $t = 3$ under the experiment assignment Z, and use the estimate

$$\hat{\tau} = c^\intercal(\alpha_{2,3}(Z) - \alpha_{1,3}(Z)) = -\$0.051. \tag{14}$$

Assuming statistical significance, this estimate would imply that game 2 is worse than game 1 in terms of revenue, contrary to our ground truth that game 2 is actually better. For this estimate to be unbiased for τ, we need $A_{it}(Z) \equiv A_i(Z_i)$, which ignores both strategic interference and the temporal dynamics of agent actions. Such an approach is not uncommon in the literature of treatment effects [16] (Chapter 3), but can only be justified when a strong assumption, such as SUTVA, is reasonable.

A more sophisticated approach is to analyze the agent actions as a time series. For example, Brodersen et. al. [6] developed a method to estimate the effects of ad campaigns on website visits. Their method was based on the idea of "synthetic controls", i.e., they created a time-series using different sources of information that would act as the counterfactual to the observed time-series after the intervention. However, their problem is macroeconometric and they work with observational data. Thus, there is neither experimental randomized assignment to games, nor strategic interference between agents, nor dynamic agent actions. More crucially, they do not study long-term, equilbrium effects. By construction, in our problem we can leverage behavioral game theory to make, arguably, more informed predictions of counterfactuals to time points at which equilibrium after the intervention has been restored.

Another approach, common in econometrics, is the so-called *difference-in-differences* (DID) estimator [7, 10, 18]. In our case, this method is inapplicable because there are no observations before the intervention, but we can still entertain the idea by considering period $t = 1$ as the pre-intervention period. The DID estimator compares the difference in outcomes before and after the intervention for both the "treated" and "control" groups. In our application, this estimator is

$$\hat{\delta} = \underbrace{(c^\intercal(\alpha_{2,3}(Z) - \alpha_{2,1}(Z)))}_{\text{change in revenue for game 2}} - \underbrace{(c^\intercal(\alpha_{1,3}(Z) - \alpha_{1,1}(Z)))}_{\text{change in revenue for game 1}} = -\$0.164. \tag{15}$$

This estimate is also far from the true value, assuming statistical significance. This estimator is unbiased for τ only if there is an additive structure in the actions [1], [3] (Section 5.2), e.g., $\alpha_{j,t}(Z) = \mu_j + \lambda_t + \epsilon_{jt}$, where μ_j is a game-specific parameter, λ_t is a temporal parameter, and ϵ is noise. The DID estimator thus captures a linear trend in the data by assuming a common parameter for both treatment arms (λ_t) that is canceled out in subtraction in Eq. (15). The extent to which additivity assumption is reasonable depends on the application, however, by definition, it implies ignorability of the assignment (i.e., Z does not appear in the model of $\alpha_{j,t}(Z)$), and thus it is a stronger than our Assumption 3.2. [1, 3].

Athey et. al. [4] study the effects of timber auction format (ascending versus sealed bid) on competition for timber tracts. Standard in econometrics for auctions, they estimate bidder valuations from observed data in one auction and impute counterfactual bid distributions in the other auction, under the assumption of equilibrium play in both auctions. This approach makes two critical implicit assumptions that together are stronger than Assumption 3.2. First, the bidder valuation distribution is assumed to be a *primitive* that can be used to impute counterfactuals in other treatment assignments. In other words, the assignment is independent of bidder values, and thus it is strongly ignorable. Second, although imputation is performed for potential outcomes in equilibrium, which captures the notion of long-term effects, inference is performed under the assumption of equilibrium play in the *observed* outcomes, and thus temporal dynamic behavior assumed away.

Finally, another popular approach to causality is through *directed acyclical graphs* (DAGs) between the variables of interest [19]. For example, Bottou et. al. [5] study the causal effects of the machine learning algorithm that scores online ads in the Bing search engine on the search engine revenue. Their approach is to create a full DAG of the system including variables such as queries, bids, and prices, and make a Causal Markov assumption for the DAG. This allows to predict counterfactuals for the revenue under manipulations of the scoring algorithm, using only observed data generated from the assumed DAG. However, a key assumption of the DAG approach is that the underlying structural equation model is stable under the treatment assignment, and only edges coming from parents of the manipulated variable need to be removed; again, assignment is considered strongly ignorable. As pointed out by Dash [9], this might be implausible in equilibrium systems. Consider, for example, a system where $X \rightarrow Y \leftarrow Z$, and a manipulation that sets the distribution of Y independently of X, Z. Then after manipulation the two edges will need to be removed. However, if in an equilibrium it is required that $Y \approx XZ$, then the two arrows should be reversed after the manipulation. Proper causal inference in equilibrium systems remains an open area without a well-established methodology [8]. However, we do believe our assumptions in Section 3 can have immediate counterparts in the DAG framework.

8.6 Data of behavioral experiment of Rapoport and Boebel [20]

	a'_1	a'_2	a'_3	a'_4	a'_5
a_1	W	L	L	L	L
a_2	L	L	W	W	W
a_3	L	W	L	L	W
a_4	L	W	L	W	L
a_5	L	W	W	L	L

Table 2: Normal-form game in the experiment of Rapoport and Boebel [20].

		row agent				column agent			
Game	Period	a_1	a_2	a_3	a_4	a'_1	a'_2	a'_3	a'_4
1	1	0.308	0.307	0.113	0.120	0.350	0.218	0.202	0.092
1	2	0.293	0.272	0.162	0.100	0.333	0.177	0.190	01.40
1	3	0.273	0.350	0.103	0.123	0.353	0.133	0.258	0.102
1	4	0.295	0.292	0.113	0.135	0.372	0.192	0.222	0.063
2	1	0.258	0.367	0.105	0.143	0.332	0.115	0.245	0.140
2	2	0.290	0.347	0.118	0.110	0.355	0.198	0.208	0.108
2	3	0.355	0.313	0.082	0.100	0.355	0.215	0.187	0.110
2	4	0.323	0.270	0.093	0.105	0.343	0.243	0.168	0.107

Table 3: Frequency of actions for the row agent and the column agent in the experiment by Rapoport and Boebel [20] broken down by game and session. Gray color indicates that we assume the data as hold-out. The frequencies for actions a_5, a'_5 can be inferred.

8.7 QLk model

In QL_k, agents possess increasing levels of sophistication. Following earlier work [26] we adopt $k = 3$, and thus consider a behavioral space with three different behavior types $\mathcal{B} = \{b_0, b_1, b_2\}$.

Recall that a behavior $b \in \mathcal{B}$ represents the distribution of actions that an agent will play after adopting that behavior. In QL_k such distributions depend on an assumption of *quantal response*, which is defined as follows. Let $u \in \mathbb{R}^{|\mathcal{A}|}$ denote a vector such that u_a is the expected utility of an agent taking action $a \in \mathcal{A}$, and let F_j denote the payoff matrix in game j as in Table 2. If an agent is facing another agent with behavior b, then $u = F_j b$. The quantal best-response with parameter λ determines the distribution of actions that the agent will take facing expected utilities u, and is defined as

$$\text{QBR}(u; \lambda) = \text{logistic}(\lambda u), \tag{16}$$

where, for a vector x with elements x_i, $\text{logistic}(x)$ is a vector with elements $\exp(x_i)/\sum_i \exp(x_i)$. The parameter $\lambda \geq 0$ is called the *precision* of the quantal best-response. If λ is very large then

the response is closer to the classical Nash best-response, whereas if $\lambda = 0$ the agent ignores the utilities and randomizes among actions.

Let $\lambda = (\lambda_1, \lambda_{(1)2}, \lambda_2)$ be the precision parameters. Given parameters λ, QL_3 calculates the distribution of actions that agent play as follows:

- Agents who adopt b_0, termed *level-0* agents, have precision $\lambda_0 = 0$, and thus will randomly pick one action from the action space $\mathcal{A}$. Thus, $b_0 = \mathrm{QBR}(u; 0) = (1/|\mathcal{A}|)\mathbf{1}$, regardless of the argument u.

- An agent who adopts b_1, termed *level-1* agent, has precision λ_1 and assumes that is playing against a level-0 type agent. Thus, the agent is facing a vector of utilities $u_1 = F_j b_0$, and so $b_1 = \mathrm{QBR}(u_1; \lambda_1)$.

- An agent who adopts b_2, termed *level-2* agent, has precision λ_2 and assumes is playing against a level-1 agent with precision $\lambda_{(1)2}$. Thus, it estimates that it is facing an agent with behavior $b_{(1)2} = \mathrm{QBR}(u_1; \lambda_{(1)2})$, where $u_1 = F_j b_0$ as above. The expected utility vector of the level-2 agent is $u_2 = F_j b_{(1)2}$, and thus its behavior is $b_2 = \mathrm{QBR}(u_2; \lambda_2)$.

Note that the behaviors b_0, b_1, b_2 depend only on the parameters λ and not on the distribution of behaviors $\beta_{j,t}(Z)$, because in our game an agent plays against only one other agent. Now, let $\Pi_j(\lambda) = [b_0 \;\; b_1 \;\; b_2]$ be the $|\mathcal{A}| \times 3$ matrix with the QL_3 behaviors parametrized by λ. Thus, QL_3 implies that the expected aggregate action is,

$$\mathbb{E}\left(\alpha_{jt}(Z) \middle| \beta_{jt}(Z)\right) = \Pi_j(\lambda) \cdot \beta_{jt}(Z), \tag{17}$$

which agrees with Lemma 3.1 for $P_j = \Pi_j(\lambda)$.

8.8 Bayesian model

We assume diffuse priors for the parameters ϕ, ψ, λ; in particular, we consider $\pi(\lambda_i) \propto \mathrm{Expo}(1/10)$, i.e., an exponential random variable with rate around $1/10$ for the parameters of the quantal best-response, a diffuse beta for ψ_0 and a flat variance prior for ψ_1 i.e., $\pi(\psi_1) \propto (1/\psi_1^2)$. For ϕ we choose a uniform Dirichlet distribution. As exact conditionals are hard to obtain under this model, we employ a Metropolis-Hastings scheme with a proposal distribution that simply disturbs slightly the current model parameters. This is efficient because our parameter space is constrained; for example, we know that $\psi_0 \in (0, 1)$ and that the precision parameters are effectively bounded because the logistic functions of quantal best-response become flat above, or below, a certain threshold. Our ultimate goal is to predict agent actions at $T = 4$ through the posterior predictive distribution of our model. We run our chain for $1e5$ iterations and assume the first half of the samples as burn-in period.