Integrating Causal Inference into Dynamic Incentive Design

Sebastián Bejos[1], Eduardo F. Morales[1], Luis Enrique Sucar[1], and Enrique Munoz de Cote[2]

[1] Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México. {sebastian.bejos,emorales,esucar}@inaoep.mx
[2] Secondmind, Cambridge CB2 1LA, United Kingdom, enrique@people-ai.com

Abstract. The design of incentives that modifies the behavior of self-interested agents to optimize the performance of a Multi-Agent System (MAS) remains a significant challenge. This is mainly due to the fact that after modifying the agents' rewards by incentives, the resulting system outcomes and changes to the agents' joint behavior are generally difficult to predict. Multi-armed bandit approaches in online learning represent interesting solutions to incentive design via exploration–exploitation strategies. However, the design of incentives for MAS that exploit causal feedback to make inferences about the performance of incentives remains uncharted territory, and the incorporation of causal reasoning in this context is an open problem. This paper introduces a way for integrating causal inference to solve dynamic incentive design problems in MAS, using Dynamic Causal Bayesian Optimization (DCBO). We use a generalization of an important representative class within MAS, called Minority Games, to show how dynamic incentive design can be addressed as a causal sequential decision problem incorporating causal reasoning by using DCBO.

Keywords: Multi-agent Systems · Causal Dynamic Incentive Design · Dynamic Causal Bayesian Optimization · Causal Sequential Decision Process · Multi-asset Minority Games

1 Introduction

In Dynamic Incentive Design (DID) problems [12], a central institution modifies the behavior of self-interested agents in a Multi-Agent System (MAS) to optimize the overall performance by introducing an incentive function to modify their individual payoffs. For instance, the central institution may want to drive the system performance to a more desirable behavior that, e.g., maximizes revenue or the social welfare. As an example, consider a ridesharing market MAS like Uber, in which it is common for users in neighborhoods with low user density to waive the platform, given that an habitual strategy among drivers is not to commit to trips in these type of areas but to remain biddable with trips that do not take them to far away from high user density zones. This also prevents the market from growing in these areas due to poor service, resulting in sub-optimal results in this MAS. For this issue, incentives can be designed, such as subsidizing fuel costs or giving a monetary reward for motivate services in this type of zones, to avoid this selfish strategies of the drivers. DID problems can be formulated as a bi-level optimization problem or equivalently as a reverse Stackelberg game [14]; in which an incentive designer agent (the principal agent), sequentially proposes an incentive function that affects the behavior of follower agents, and adapts sequentially this incentive function based on the followers' responses. DID has been a subject of interest to economics and control theory for long time [9]. In recent years there has been an increasing interest in studying DID in machine learning through the lens of online learning [8, 6, 11]. Other recent efforts have adopted the agent-based simulation paradigm and have taken state-of-the-art agent learning methods, such as Multi-Agent Meta Reinforcement Learning [15, 13, 7]. Nevertheless, to the best of our knowledge, there is no work that integrates causal reasoning to address DID problems. The use of causal inference to make predictions about incentive performance and solve for optimal incentives is the main motivation for this research. Causal inference would allow to take full advantage of data samples of observations on past incentive functions to infer the effectiveness of new alternative interventions. We deal with the integration of causal inference into DID problems by incorporating the Dynamic Causal Bayesian Optimization (DCBO) framework to handle the sequential decision process encountered by the principal agent in this scenario.

We define the main aspects of the principal-agents problem [12], the El Farol Bar problem [3, 5], and the DCBO [1] in Section 2 below. After in section 3 we show how to integrate DCBO to solve DID problems by setting the principal-agents problem as a dynamic probabilistic causal model. Considering the general causal structure of the principal-agent problem, in Section 4 we illustrate how the structural equations can be derived from a bi-level optimization formulation of a principal-agent problem, adopting the Multi-Asset Minority Games as an instance for such derivation. Conclusions and future work are given in Section 5.

2 Preliminaries

2.1 The Principal-Agents Problem

We restrict our attention to the principal–agents problem of DID, in which there is only one principal agent and a set $\boldsymbol{N}$ of followers agents, with $|\boldsymbol{N}| = n$, in a MAS, where the underlying model of the environment dynamics can be described as a differential or difference equation in a continuous or discrete time MAS, respectively[1][12]. Let $J_p(v^t, \boldsymbol{u}^t, s^t)$ be the utility function of the principal, with $J_p : \boldsymbol{V} \times \boldsymbol{U}_1 \times \ldots \times \boldsymbol{U}_n \times \boldsymbol{S} \rightarrow \mathbb{R}$, and let $\{J_{a_i}(v^t, \boldsymbol{u}^t, s^t) \mid i \in [n]\}$ be the set of utility functions for the followers agents, with $J_{a_i} : \boldsymbol{V} \times \boldsymbol{U}_1 \times \ldots \times \boldsymbol{U}_n \times \boldsymbol{S} \rightarrow \mathbb{R}$, where $\boldsymbol{V}$ is the action space of the principal, $\boldsymbol{U}_i$ is the action space for the follower

[1] In this work we focus in discrete time multi-agent systems

agent a_i, v^t and $\boldsymbol{u}^t$ are the decisions of the principal and the followers agents, respectively, at time t. $\boldsymbol{S}$ is the state space of the multi-agent system, s^t is the state at time t, and the system state dynamics is given by $s^{t+1} = d(v^t, \boldsymbol{u}^t, s^t)$. Considering a T horizon, there is a game between the principal and the followers agents, where there is an specific order of play. For rounds $t = 1, \ldots, T$, the interaction protocol in the game is as follows:

1. The principal p decides and announces an incentive function $\gamma^t : \boldsymbol{U}_1 \times \ldots \times \boldsymbol{U}_n \to \boldsymbol{V}$ with full, an estimation or no knowledge at all (as applicable) of the utility functions of the followers agents $\{J_{a_i} \mid i \in [n]\}$.
2. Then, with knowledge of this incentive function, each follower agent $a_i \in \boldsymbol{N}$ selects an action $u_i^{t^*}$ that maximizes their utility. So, at the end of this step, the decisions vector $\boldsymbol{u}^{t^*} = (u_1^{t^*}, \ldots, u_i^{t^*}, \ldots, u_n^{t^*})$ is public to the principal agent p, where $u_i^{t^*} \in \arg\max J_{a_i}(\gamma^t(\boldsymbol{u}^t), \boldsymbol{u}^t, s_t)$ for each $i \in [n]$.

The goal of the game is to find $\{(v^{t^*}, \boldsymbol{u}^{t^*})\}_t \in \arg\max \sum_{t \in [T]} J_p(v^t, \boldsymbol{u}^t, s^t)$, i.e., maximize the principal's cumulative utility, by selecting an incentive function γ^t in a set of admissible incentives functions $\boldsymbol{\Gamma} = \{\gamma : \boldsymbol{U}_1 \times \ldots \times \boldsymbol{U}_n \to \boldsymbol{V}\}$, at each time $t \in [T]$. In other words, at each round t the principal selects an incentive function $\gamma^t \in \boldsymbol{\Gamma}$ such that the followers agents chooses an action that leads to the maximization of the principal's utility, which usually is equivalent to the system utility or what is best for the multi-agent system, e.g., maximize the social welfare, i.e., the sum of expected gains of all agents in the long run.

2.2 El Farol Bar Problem and Minority Games

In the *El Farol Bar* problem a set $\boldsymbol{N}$ of n agents have to decide independently on each time t whether to go to the bar, $u_i^t = 1$, or not go, $u_i^t = 0$, i.e., the action space for each agent a_i is $\boldsymbol{U}_i = \{0, 1\}$, with $i \in [n]$. The Farol Bar has a capacity limit, $L < n$, and the bar is enjoyable only if it is not overcrowded, i.e., only if the attendance $A^t = \sum_{i \in [n]} u_i^t$ does not exceed L. In order to make their decisions, agents aim to predict whether the bar will be crowded or not on any given time t based on the past attendances. It is assumed that agents base their predictions on the attendances of a finite number $m \in \mathbb{N}$ of past times, and the information available to agents at time t is encoded in the string $\lambda^t = \big(\zeta[(L - A^{t-1}], \ldots, \zeta[(L - A^{t-m}]\big) \in \{0, 1\}^m$, where $\zeta(\cdot)$ is the Heaviside function, i.e., $\zeta[(L - A^t] = 1$ if the bar was enjoyable $(A^t < L)$ while $\zeta[(L - A^t] = 0$ if the bar was overcrowded $(A^t \geq L)$ at time t.

Given information λ^t, each agent a_i have a set $\boldsymbol{\Lambda}_i = \{\Lambda_{i1}, \ldots, \Lambda_{iK_i}\}$ of functions Λ_{ik} with $k \in [K_i]$ for some $K_i \in \mathbb{N}$ called strategies or predictors, that maps information strings $\{0, 1\}^m$ into the binary actions set $\{0, 1\}$ of go or do not go, i.e., every strategy $\Lambda_{ik} \in \boldsymbol{\Lambda}_i$ is a function such that $\Lambda_{ik} : \{0, 1\}^m \to \{0, 1\}$.

Ranking the strategies of each agent in the El Farol Bar problem can be approached through different learning processes, here we focus on the cumulative utility based ranking [5]. These processes help agents evaluate and adapt their strategies based on their performance as the game progresses. Let $u_{\Lambda_{ik}}^{\lambda^t} \in \{0, 1\}$

denote the prediction of strategy $\Lambda_{ik} \in \boldsymbol{\Lambda}_i$ of agent a_i under the information λ^t at time t, and let $\Omega^t_{\Lambda_{ik}}$ be the cumulative utility of agent a_i using strategy Λ_{ik} up to time t. At every time t, agent a_i selects a strategy with the highest cumulative utility ${\Lambda^t}^*_{ik} \in \arg\max_{\Lambda_{ik}} \Omega^t_{\Lambda_{ik}}$, and acts accordingly, i.e., $u^t_i = u^{\lambda^t}_{\Lambda^*_{ik}}$. In order to decide which strategy to adopt on every time t, agents keep track of their performance via the cumulative utility that is updated according to the rule: $\Omega^{t+1}_{\Lambda^*_{ik}} = \Omega^t_{\Lambda^*_{ik}} + (1 - 2u^{\lambda^t}_{\Lambda^*_{ik}})[A^t - L]$, with the rationale that strategies suggesting not to go ($u^{\lambda^t}_{\Lambda^*_{ik}} = 0$) are rewarded when the attendance is higher than L and punished when it is lower than L (and vice versa when $u^{\lambda^t}_{\Lambda^*_{ik}} = 1$).

For modeling purposes, Minority Games serve as a class of simple models which are able to produce some of macroscopic features being observed in the real financial markets. The basic Minority Game corresponds roughly to the case where $L = \lfloor \frac{n}{2} \rfloor$ of the El Farol Bar problem, but here the agents actions are either to buy, $u^t_i = 1$, or sell, $u^t_i = 1$, an specific stock, to model speculative trading in financial markets [3, 2, 4].

2.3 Dynamic Causal Bayesian Optimization (DCBO)

A graphical causal model consist of a four tuple $\langle \boldsymbol{W}, \boldsymbol{Z}, P(\boldsymbol{Z}), \boldsymbol{F} \rangle$ and a directed acyclic graph $\mathcal{G}$, where $\boldsymbol{W}$ is the set of observed endogenous variables, $\boldsymbol{Z}$, is a set of exogenous variables expressing a random disturbance distributed according to $P(\boldsymbol{Z})$, and $\boldsymbol{F} = \{f_1, \ldots, f_{|\boldsymbol{W}|}\}$ is a set of functions known as structural equations, such that $W_i = f_i(\boldsymbol{pa}(W_i), Z_i)$, for each $W_i \in \boldsymbol{W}$, with $\boldsymbol{pa}(W_i)$ denoting the parents of W_i [1]. The graph $\mathcal{G}$ encodes the causal relationship between the variables in $\boldsymbol{W}$. Within $\boldsymbol{W}$ we distinguish three different types of variables: non-manipulative variables $\boldsymbol{C}$, treatment variables $\boldsymbol{X}$ that can be set to specific values, i.e., intervene them, and output variable Y that represent the outcome of interest. In order to reason about interventions that are implemented in a sequential manner, i.e., at each time t we decide which intervention to perform in the system. Let $\mathcal{M}^t$ be a dynamic graphical causal model defined as $\mathcal{M}^t = \langle \mathcal{G}^{1:t}, \boldsymbol{W}^{1:t}, \boldsymbol{Z}^{1:t}, P(\boldsymbol{Z}^{1:t}), \boldsymbol{F}^{1:t} \rangle$, where $1:t$ denotes the union of the corresponding graphs, variables or functions up to time t, $\boldsymbol{W}^{1:t} = \boldsymbol{X}^{1:t} \cup \boldsymbol{C}^{1:t} \cup \boldsymbol{Y}^{1:t}$. The goal of DCBO is to find a sequence of interventions, optimizing a target variable, at each time t, in a graphical causal model $\mathcal{M}^t$. Given $\mathcal{M}^t$, at every time step t, we wish to optimize Y_t by intervening on a subset of the manipulative variables $\boldsymbol{X}_t$. The optimal intervention variables $\boldsymbol{X}^*_{s,t}$ and intervention levels $\boldsymbol{x}^*_{s,t}$ are given by:

$$\boldsymbol{X}^*_{s,t}, \boldsymbol{x}^*_{s,t} = \underset{\boldsymbol{X}_{s,t} \in \mathcal{P}(\boldsymbol{X}_t), \boldsymbol{x}_{s,t} \in D(\boldsymbol{X}_{s,t})}{\arg\max} E[Y_t \mid do(\boldsymbol{X}_{s,t} = \boldsymbol{x}_{s,t}), \mathbb{1}_{t>1} \cdot \boldsymbol{I}_{1:t-1}]. \quad (1)$$

where $\mathcal{P}(\boldsymbol{X}_t)$ is the power set of $\boldsymbol{X}^*_{s,t}$, $D(\boldsymbol{X}_{s,t})$ represents the interventional domain of $\boldsymbol{X}_{s,t}$, $\boldsymbol{I}_{1:t-1} = \bigcup_{i=1}^{t-1} do(\boldsymbol{X}^*_{s,i} = \boldsymbol{x}^*_{s,i})$ denotes previous interventions, and $\mathbb{1}_{t>1}$ is the indicator function. The expected value is over the interventional distribution $P(Y_t \mid do(\boldsymbol{X}^*_{s,t} = \boldsymbol{x}^*_{s,t}))$, given $\mathbb{1}_{t>1} \cdot \boldsymbol{I}_{1:t-1}$. DCBO make the assumptions of invariance of causal structure, i.e., $\mathcal{G}^t = \mathcal{G}^{t+1}$ for all $t \in [T]$, and absence of unobserved confounders in $\mathcal{G}_{1:T}$ [1].

3 Causal Dynamic Incentive Design

In this section we show how causal inference can be integrated in the principal-agent problem described in Section 2. This is done by first proposing the general causal structure $\mathcal{G}^t_{pap}$ for a dynamic graphical causal model representing the principal agent problem $\mathcal{M}^t_{pap}$. The dynamics between the principal and the follower agents are modeled in $\mathcal{G}^t_{pap}$ in the following way. The set of observed endogenous variables is given by $\boldsymbol{W^{1:T}} = \{\boldsymbol{V}^{1:T}, \boldsymbol{U}^{1:T}_i, \boldsymbol{J}^{1:T}_{a_i}, \boldsymbol{J}^{1:T}_P\}$, for all $i \in [\boldsymbol{N}]$, where $\boldsymbol{V}^{1:T}$ is the variable representing the principal actions, $\boldsymbol{U}^{1:T}_i$ represents the agent i actions, $\boldsymbol{J}^{1:T}_{a_i}$ represents the agents i utility function, for all $i \in [\boldsymbol{N}]$, and $\boldsymbol{J}^{1:T}_P$ represents the utility function of the principal, all up to time T. The Figure 1 shows the directed graph $\mathcal{G}^{t:t+1}_{pap}$, where the direct causal relation are show for the variables in $\boldsymbol{W^{t:t+1}}$, i.e., for times or rounds t and $t+1$ of the principal-agents protocol (see Section2), and from which all the direct causal relation for $\boldsymbol{W^{1:T}}$ can be inferred. We are using the following color convention for the types of variables that appear in $\mathcal{G}^{t:t+1}_{pap}$ shown in Figure 1 (as well as for the next figures): blue for the variables that are not manipulable, i.e., $\boldsymbol{U}^{1:T}_i \cup \boldsymbol{J}^{1:T}_{a_i} = \boldsymbol{C}^{1:T}$, green for those that are manipulable, i.e., $\boldsymbol{V}^{1:T} = \boldsymbol{X}^{1:T}$ and orange for the target variables, i.e., $\boldsymbol{J}^{1:T}_p = \boldsymbol{Y}^{1:T}$, at each time t.

.

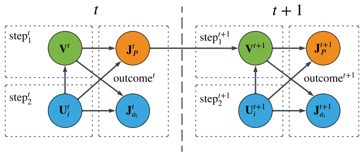

Fig. 1: $\mathcal{G}^{t:t+1}_{pap}$.

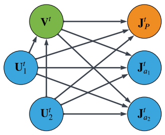

Fig. 2: $\mathcal{G}^t_{pap}$ for $n = 2$.

We are doing an important simplification on the graphical representation in Figure 1 as the real dynamic graphical causal structure most contemplate one variable for each of the agents' action space and one variable for each of the utility functions of each of the agents. Figure 2 shows what the structure $\mathcal{G}^t_{pap}$ would actually look like for the case of $n = 2$ on a given time t. In Figure 1, we are using the variables $\boldsymbol{U}^t_i$ and $\boldsymbol{J}^t_{a_i}$ to represent all the variables for each each of the agents' action space $\boldsymbol{U}^t_1, \ldots \boldsymbol{U}^t_n$ and for each of the utility functions $\boldsymbol{J}^t_{a_1}, \ldots \boldsymbol{J}^t_{a_n}$ of each of the agents, respectively, to keep the drawings understandable and not overly edged.

Figure 3 (a) shows the simplification pattern for variables $\boldsymbol{U}^t_i$ and and its causal relationship with $\boldsymbol{V}^t$. On Figure 3 (b) it is shown what this pattern represents, where the causal structure from variables $\boldsymbol{U}^t_1, \ldots \boldsymbol{U}^t_n$ is given by an empty graph, i.e., no edges or causal relations between $\boldsymbol{U}_1^t, \dots \boldsymbol{U}_n^t$. Figure 3 (c) shows an instance of the other scenario where the causal structure between variables $\boldsymbol{U}_i^t$ is not an empty graph, it is shown a sub-DAG for the action variables $\boldsymbol{U}_i^t$ of four agents in this case. In some application contexts maybe important to perform a causal discovery method to find the underlying causal structure of variables $\boldsymbol{U}_i^t$. We leave this kind of scenario for later work.

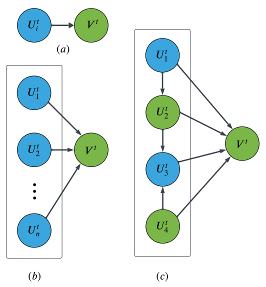

Fig. 3: The pattern $\boldsymbol{U}_i^t \rightarrow \boldsymbol{V}$.

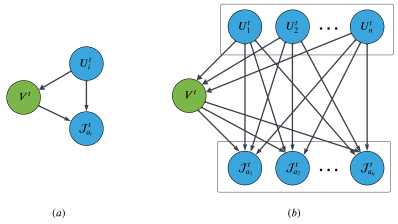

Fig. 4: The directed triangle pattern.

Figure 4 (a) depicts another graphical simplification in the $\mathcal{G}_{pap}^t$ shown in Figure 1, we call it the directed triangle pattern, which represent the direct causal relation of $\boldsymbol{U}_i^t \rightarrow \boldsymbol{V}^t$, $\boldsymbol{V}^t \rightarrow \boldsymbol{J}_{a_i}^t$, and $\boldsymbol{U}_i^t \rightarrow \boldsymbol{J}_{a_i}^t$. In Figure 4 (b) it is shown the unfold directed triangle pattern, i.e., the real structure that it represents. Something similar occurs for the directed triangle pattern representing the direct causal relations of $\boldsymbol{U}_i^t \rightarrow \boldsymbol{V}^t$, $\boldsymbol{V}^t \rightarrow \boldsymbol{J}_p^t$, and $\boldsymbol{U}_i^t \rightarrow \boldsymbol{J}_p^t$. Another aspect that is shown in the graph of Figure 3 (c) is that in some application contexts maybe important to consider interventions on a subset of agents' action space variables. As shown in Figure 3, some U_i^t variables are colored green meaning they are in the set of manipulable variable $\boldsymbol{X}^t$. Manipulating certain U_i^t variables can be helpful for exploring the effects of specific actions taken by a group of agents on the causal model.

We focus in the case of no edges between the $U_1^t, \dots, U_n^t$ variables and with none of these variables in the set of manipulable variables $\boldsymbol{X}^t$. Given this developed causal structure $\mathcal{G}_{pap}^t$, it is possible to completely specify the DCBO method (See Section 2) to solve a general principal-agent problem by defining the action space of the principal $\boldsymbol{V}^t$, the action space of the agents $\boldsymbol{U}_i^t$, for all $i \in [n]$, and the structural equations $\boldsymbol{F}^t = \{f_{\boldsymbol{V}^t}^t, f_{\boldsymbol{J}_p^t}^t, f_{\boldsymbol{J}_{a_i}^t}^t\}$, for all $i \in [n]$, which have the following general arrangement due to the causal structure $\mathcal{G}_{pap}^t$: $f_{\boldsymbol{J}_{a_i}^t}^t = f^{\boldsymbol{J}_{a_i}^t}(v^t, \boldsymbol{u}_i^t, \epsilon_{\boldsymbol{J}_a}^t)$, $f_{\boldsymbol{J}_p}^t = f^{\boldsymbol{J}_p^t}(v^t, \boldsymbol{u}_i^t, \epsilon_{\boldsymbol{J}_p}^t)$, and

$$f_{\boldsymbol{V}^t}^t = \begin{cases} f^{\boldsymbol{V}^t}(\boldsymbol{u}_i^t, \epsilon_{\boldsymbol{V}}^t), & \text{if } t = 1, \\ f^{\boldsymbol{V}^t}(\boldsymbol{u}_i^t, \boldsymbol{J}_p^{t-1}, \epsilon_{\boldsymbol{V}}^t), & \text{if } t > 1, \end{cases}$$

where $\epsilon_{\boldsymbol{J}_a}^t, \epsilon_{\boldsymbol{J}_p}^t, \epsilon_{\boldsymbol{V}}^t \in \boldsymbol{Z}^t$ are the exogenous variables expressing a random disturbance distributed according to $P(\boldsymbol{Z}^t)$.

The detailed definition of these structural equations for the causal dynamic graphical model $\mathcal{M}_{pap}^t$ will depend on the application context. In the application settings, it is a common practice to formulate a principal-agent problem as a bi-level optimization problem [12]. Under this type of formulation of a principal-agent problem, the utility of the principal and the agents are presented as the objective functions of the upper level and lower level problem, respectively [14]. In Section 4, we present an interesting instance of the principal agent problem for which we give its bi-level optimization formulation and show how from the objective functions we can obtain directly the structural equations for the $\mathcal{M}_{pap}^t$. It is necessary to mention that at each time t the interaction protocol between the principal and the followers (Seen in Section 2) is respected. That is, the principal first on step 1 decides (or chooses) an incentive function γ from a set of possible incentive function $\boldsymbol{\Gamma} = \{\gamma : \boldsymbol{U}_1 \times \ldots \times \boldsymbol{U}_n \to \boldsymbol{V}\}$. Then, on step 2, with the knowledge of the principal's selected incentive function γ, as this is incorporated in the follower's utility function $\boldsymbol{J}_{a_i}(v^t, \boldsymbol{u}_i^t) = \boldsymbol{J}_{a_i}(\gamma(\boldsymbol{u}_i^t), \boldsymbol{u}_i^t)$, the follower agents decide $\boldsymbol{u}_i^t \in \boldsymbol{U}_i^t$. After the followers agent's decisions, on step 2, the outcome for time t is computed as the variables $\boldsymbol{J}_{a_i}^t$ and $\boldsymbol{J}_p^t$ for every time $t \in [T]$. This interaction protocol for each time t is shown in Figure 1 by partitioning the vertex set of $\mathcal{G}_{pap}^t$ with the sets $step_1^t = \{\boldsymbol{V}^t\}$, $step_2^t = \{\boldsymbol{U}_i^t\}$, and $outcome^t = \{\boldsymbol{J}_{a_i}^t, \boldsymbol{J}_p^t\}$. In the light of all that and the given specification of the general dynamic graphical causal model $\mathcal{M}_{pap}^t$ the agent-principal problem can be solve by finding a sequence of interventions on variables $\boldsymbol{V}^{1:T}$ optimizing the target variables $\boldsymbol{J}_p^{1:T}$, i.e., by computing:

$$v^{t^*} = \gamma^{t^*}(\boldsymbol{u}_i^t) = \operatorname*{arg\,max}_{\gamma^t \in \Gamma} \mathbb{E}\big[\boldsymbol{J}_p^t \mid do(\boldsymbol{V}_t = \gamma^t(\boldsymbol{u}_i^t) = v^t), \mathbb{1}_{t>1} \cdot \boldsymbol{I}_{1:t-1}\big], \qquad (2)$$

with the important remark that intervene the variable $\boldsymbol{V}^t$ is to assign an incentive function $\gamma \in \Gamma$ to variable $\boldsymbol{V}^t$, representing the choose of γ by the principal on step 1, which after step 2 gets a value $v^t \in \boldsymbol{V^t}$ by the causal relation $\boldsymbol{U}_i^t \to \boldsymbol{V}^t$. Moreover, the causal relation $\boldsymbol{J}_p^{t-1} \to \boldsymbol{V}^t$ on $\mathcal{G}_{pap}^t$ stablish the utilization of previous results for $\boldsymbol{J}_p^t$ to inform the exploration for the optimal incentive function selection, using the result of previous interventions $\boldsymbol{I}_{1:t-1}$. We call Causal Dynamic Incentive Design (CDID) to the method described in this section for solving the principal-agent problem.

4 CDID on the Multi-Asset Minority Game

We explore a generalization of the El Farol Bar problem and the Minority Game, where agents must decide not only whether to go to or not to go to the bar or to buy or to sell a stock but also which bar to attend or which stock to buy or sell from a collection of bars or stocks, where each asset has its own capacity. We call this generalization a Multi-Asset Minority Game. Let M be the number of assets and let each asset has a capacity limit L_j with $L_j < n$ for $j \in [M]$. Let $u_{ij}^t \in \{0, 1\}$ be the decision variable indicating whether agent i decides in to the asset j at time t. Let $\mu_j^t = \big(\zeta[(L_j - A_j^{t-1}], \dots, \zeta[(L_j - A_j^{t-m}]\big) \in \{0,1\}^m$ be the string encoding the information available to agents at time t assuming finite memory of m past times references over the asset $j \in [M]$, with $A_j^t = \sum_{i \in [N]} u_{ij}^t$. The information available to agents at time t is encoded in the following string:

$$\mu^t = \Big(\bigoplus_{j \in [M]} \mu_j^t \Big) \in \{0,1\}^{m|M|},$$

where $\bigoplus_{j \in [M]} \mu_j^t$ denotes de concatenation of the M strings $\mu_j^t \in \{0,1\}^m$. Let $\boldsymbol{\Lambda}_i = \{\Lambda_{i1}, \Lambda_{i2}, \dots, \Lambda_{iK_i}\}$ be the strategies set of agent a_i, $u_{\Lambda_{ik}}^{\mu_j^t} \in \{0,1\}$ denote the prediction of strategy Λ_{ik} with $k \in [K_i]$ for some $K_i \in \mathbb{N}$ of agent i for asset $j \in [M]$, under the information μ_j^t at time t. Let $U^j_{\Lambda_{ik}}(t)$ the the cumulative utility of agent i using strategy Λ_{ik} up to time t for asset $j \in M$. Every time t, agent i selects the strategy Λ_{ik}^* with the highest cumulative utility for each asset $j \in M$, and acts accordingly, i.e., $u_{ij}^t = u_{\Lambda_{ik}^*}^{\mu_j^t}$. Let Λ_{ik}^* be such strategy in $\boldsymbol{\Lambda}_i$, i.e.,

$$\Lambda_{ik}^* = \underset{\Lambda_{ik} \in \boldsymbol{\Lambda}_i}{\arg\max}\, \Omega^j_{\Lambda_{ik}}(t),$$

In order to decide which strategy to adopt for each asset $j \in M$ on every time t, agents keep track of their performance for each asset $j \in M$ via the cumulative utility $\Omega^j_{\Lambda_{ik}}(t)$ that is updated according to the following rule:

$$\Omega^j_{\Lambda_{ik}^*}(t+1) = \Omega^j_{\Lambda_{ik}^*}(t) + (1 - 2u_{\Lambda_{ik}^*}^{\mu_j^t})[A_j^t - L_j],$$

4.1 Bi-level Optimization Formulation of the Multi-Asset Minority Game with Incentives Design

We now extend the Multi-Asset Minority Game (MAMG) to include Dynamic Incentive Design (DID), presenting a formulation of this Multi-Asset Minority Game as a principal-agents problem. For this, we present the Multi-Asset Minority Game with Dynamic Incentives as a bi-level optimization problem below, specifying the principal's (upper level) and agents' (lower level) objective functions, each with their own constraints. The objective function of the principal is given as follows:

$$\text{(Principal)} \quad \max_{\gamma_j^t(\boldsymbol{u}_{ij}) \in \boldsymbol{\Gamma}} \quad \mathbb{E}\left[\sum_{t \in [\boldsymbol{T}]} \sum_{j \in [\boldsymbol{M}]} R_j\big(A_j^t\big) - C_j\big(A_j^t, \gamma_j^t(\boldsymbol{u}_{ij})\big) \right] \tag{3}$$

$$\text{s. a. } \gamma_j^t(\boldsymbol{u}_{ij}) = \begin{cases} \alpha\Big(\frac{L_j - A_j^t}{L_j}\Big) \leq \omega_j^t & \text{if } A_j^t \leq L_j, \text{ where } \omega_j^t \text{ is the budget at } t. \\ \alpha\Big(\frac{A_j^t - L_j}{L_j}\Big) & \text{if } A_j^t > L_j, \text{ for some } \alpha > 0 \in \mathbb{R}. \end{cases}$$

where $R_j\left(A_j^t\right)$ is the revenue function assuming a revenue per unit of attendance (e.g., trading volume) for asset j, and $C_j\left(A_j^t, \gamma_j^t(\boldsymbol{u}_{ij})\right)$ is the cost function, presuming a cost per unit of attendance, and a cost per unit of incentive. Note that we define the incentives in a way that when $A_j^t \leq L_j$ the principal offers a subsidy which is given based on how far the attendance is from the limit, on the other hand, when $A_j^t > L_j$ the principal imposes taxes on the agents according to the excess of attendance. The agents' objective function is shown below:

$$\text{(Agents)} \quad (\boldsymbol{u}_{ij}^*) \in \underset{\boldsymbol{u}_{ij} \in \boldsymbol{U}_{ij}}{\arg\max} \left\{ \mathbb{E}\left[\sum_{t \in [T]} \sum_{j \in [M]} \Big((1 - 2u_{ij}^t)[A_j^t - L_j] \Big) + \gamma_j^t(\boldsymbol{u}_{ij}) \right] \right\}_{i \in [n]}$$

$$\text{s. a.} \sum_{j \in [\boldsymbol{M}]} u_{ij}^t \leq |M|, \quad \forall i \in [n],\ \forall j \in [M],\ \forall t \in [T]. \tag{4}$$

It is important to note that the above are not two independent optimization problems, one for the primary agent's objective function and constraints and the other for the followers' objective functions and constraints. Rather, it is a single bi-level optimization problem, where an optimal solution for the lower level problem, i.e., the one corresponding to the follower agents, is only a feasible solution for the upper level optimization problem, the one corresponding to the principal agent.

4.2 CDID for the Multi-asset Minority Game with Incentives Design.

From the analysis given in Section 3, we can use CDID to solve the Multi-asset Minority Game with incentives design by first defining the action spaces of the principal and follower agent, i.e., variables $\boldsymbol{V}^t$, and $\boldsymbol{U}_i^t$, for all $i \in [\boldsymbol{N}]$, which for this principal-agent MAS we can state as:

$$\boldsymbol{V}^t = \left\{ \boldsymbol{v}^t = (v_1^t, \ldots, v_M^t) \in \mathbb{R}^M \mid v_j^t = \gamma_j^t(\boldsymbol{u}_{ij}^t),\ \text{for } j \in [M], \gamma_j^t \in \mathit{\Gamma} \right\}$$

$$\mathcal{U}_i^t = \left\{ \boldsymbol{u}_i^t = (u_{i1}^t, \ldots, u_{iM}^t) \in \{0,1\}^M \mid u_{ij}^t = u_{\mathit{\Lambda}_{ik}^*}^{\mu^t} \in \{0,1\},,\ \text{for } j \in [M] \right\}$$

Subsequently, given the general structure for the dynamic causal model representing a principal-agent problem $\mathcal{M}_{pap}^t$ as given in Section 3, it simply remains to state the structural equations $\boldsymbol{F}^t = \{f_{\boldsymbol{V}^t}^t, f_{\boldsymbol{J}_P^t}^t, f_{\boldsymbol{J}_{a_i}^t}^t\}$, for all $i \in [n]$. Using the above bi-level optimization formulation of the principal-agent problem for the Multi-asset Minority Game with incentives design, we can set up the structural equations $\boldsymbol{F}^t$ for its dynamic causal model $\mathcal{M}_{pap}^t$ as follows:

$$f^t_{\boldsymbol{J}_p} = f^{\boldsymbol{J}^t_p}(v^t, \boldsymbol{u}^t_i, \epsilon^t_{\boldsymbol{J}_p}) = \mathbb{E}\left[\sum_{j\in[M]} R_j\big(A^t_j\big) - C_j\big(A^t_j, \gamma^t_j(\boldsymbol{u}_{ij})\big) + \epsilon^t_{\boldsymbol{J}_p}\right] \tag{5}$$

$$f^t_{\boldsymbol{J}^t_{a_i}} = f^{\boldsymbol{J}^t_{a_i}}(v^t, \boldsymbol{u}^t_i, \epsilon^t_{\boldsymbol{J}_a}) = \mathbb{E}\left[\sum_{i\in[M]}\Big((1-2u^t_{ij})[A^t_j - L_j]\Big) + \gamma^t_j(\boldsymbol{u}_{ij}) + \epsilon^t_{\boldsymbol{J}_a}\right] \tag{6}$$

$$f^t_{\boldsymbol{V}^t} = \begin{cases} f^{\boldsymbol{V}^t}(\boldsymbol{u}^t_i, \epsilon^t_{\boldsymbol{V}}), & \text{if } t = 1, \\ f^{\boldsymbol{V}^t}(\boldsymbol{u}^t_i, \boldsymbol{J}^{t-1}_p, \epsilon^t_{\boldsymbol{V}}), & \text{if } t > 1, \end{cases} = \begin{cases} \big((0,\dots,0) + \boldsymbol{\epsilon}^t_{\boldsymbol{J}_p}\big) \in \mathbb{R}^M \text{ if } t = 1, \\ \boldsymbol{v}^{t^*} = (v^{t^*}_j) \in \mathbb{R}^M, \text{ if } t > 1, \text{ where} \end{cases} \tag{7}$$

$$v^{t^*}_j = \gamma^{t^*}_j(\boldsymbol{u}^t_{ij}) = \underset{\gamma^t_j \in \boldsymbol{\Gamma}}{\arg\max}\, \mathbb{E}\big[\boldsymbol{J}^t_p \mid do(\boldsymbol{V}^t = \gamma^t_j(\boldsymbol{u}^t_{ij}) = v^t_j), \mathbb{1}_{t>1} \cdot \boldsymbol{I}_{1:t-1}\big],$$

which is exactly Equation 2 from Section 3, but for each asset $j \in [M]$, where $\epsilon^t_{\boldsymbol{J}_a}, \epsilon^t_{\boldsymbol{J}_p}, \epsilon^t_{\boldsymbol{V}} \in \boldsymbol{Z}^t$ for which we assume $P(\epsilon^t_{\boldsymbol{J}_a}) = P(\epsilon^t_{\boldsymbol{J}_p}) = P(\epsilon^t_{\boldsymbol{V}}) \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, with $\mu = 0$ and $\sigma = 1$. Observe that Equations 5 and 6, which correspond to the structural equations of the utility variables $\boldsymbol{J}^t_p$ and $\boldsymbol{J}^t_{a_i}$, for all $i \in [n]$, are practically the objective functions of the bi-level optimization problem for the principal agent (Equation 3) and the follower agents (Equation 4), respectively. Finally, an important design aspect for using the CDID framework and in particular to fully define the set $\boldsymbol{V}^t$ in the MAMG with DID is to define the set of admissible incentive functions $\boldsymbol{\Gamma}$. Many variants of the $\boldsymbol{\Gamma}$ set can be proposed, yet we exhibit two instances of incentive functions families that independently each of them could be established as $\boldsymbol{\Gamma}$, but also the union of both families of incentive functions. For example, to distribute clients optimally across assets, the principal can implement a dynamic pricing scheme by stablish a cost

$$DPS_j(A^t_j) = bp_0 + \beta\frac{A^t_j}{L_j},$$

for access or buying an asset $j \in [M]$ as a function of attendance A^t_j, where bp_0 is the base price and β is a scaling factor. Likewise, agents could be penalized for choosing overcrowded assets by defining a congestion penalty incentive function

$$CPIF(A^t_j) = \beta\big[max\big(0, (A^t_j - L_j)\big)\big],$$

that reduces individual utility based on attendance. Where β is a scaling factor. Each agent choosing an asset j receives a penalty if the asset is over capacity, incentivizing them to opt for less crowded assets.

5 Conclusions

The generation of experimental data in the dynamic incentive design on a MAS is expensive. In this context, such experimental data translates into testing different incentives and measuring their effects on the multi-agent system in question. The field of causal inference features a rich set of tools to evaluate the performance of untested incentives and solve for optimal incentives, thereby allowing to make the most of limited data samples, i.e., experience with past incentive functions. That is, using observations on the MAS and data from past incentive functions effect on the system, causal inference can infer the effectiveness of new alternative interventions by evaluating post-intervention distributions and rate different incentive functions cheaper. The main contribution in this paper is the presentation of CDID method in which causal inference can be incorporated to deal with incentive design in MAS. For this purpose, we leverage of the DCBO method and characterize a generic dynamic causal probabilistic model $\mathcal{M}_{pap}$ representing a principal-agent problem in general, and show how the structural equations of this dynamic causal probabilistic model can be derived from a bi-level optimization formulation of a principal-agents problem in MAS. An interesting result is that in the case of homogeneous follower agents, only three structural equations of the dynamic causal model are needed for the application of DCBO. For the more general case, i.e., contemplating heterogeneous follower agents MAS with g different groups of agents, it would require the formulation of $2g + 1$ structural equations of the dynamic causal model.

5.1 Future Work.

We are working in the validation this proposal using data from Agent Based Modeling simulations [10] of Multi-asset Minorty Games and other MASs, like traffic assignment systems, and compare with solutions based on Bayesian Optimization for DID [11], hoping to reach faster convergence in the search for optimal incentives functions. In this research, we focus on a single principal with several follower agents version of the DID, but another goodness about our proposal is that it can be straightforward extended to contemplate more than one principal agent. Moreover, it can also be extended to consider more than two levels of hierarchy among agents, e.g., that there were principal agents at a third level who incentivize second level principals. Additionally, the proposed framework for DID may be appropriate for considering information asymmetries, such as adverse selection. Which can be characterized as the follower agent's utilities being dependent on some parameter $\theta \in \Theta$ representing the agent's type, so the utility functions J_{a_i} for $i \in [n]$ can be expressed as $J_{a_i}(v, \boldsymbol{u}; \theta)$, and θ is unknown a priori to the principal. We intend to investigate these variants that can be easily accommodated within our framework by incorporating new variables into the probabilistic causal model in future research.

References

1. Aglietti, V., Dhir, N., González, J., Damoulas, T.: Dynamic causal bayesian optimization. Advances in Neural Information Processing Systems **34**, 10549–10560 (2021)
2. Challet, D., Marsili, M., Zhang, Y.C.: Minority games: interacting agents in financial markets. OUP Oxford (2004)
3. Challet, D., Zhang, Y.C.: Emergence of cooperation and organization in an evolutionary game. Physica A: Statistical Mechanics and its Applications **246**(3-4), 407–418 (1997)
4. Coolen, A.C.: The mathematical theory of minority games: statistical mechanics of interacting agents. OUP Oxford (2005)
5. De Martino, A., Marsili, M.: Statistical mechanics of socio-economic systems with heterogeneous agents. Journal of Physics. A, Mathematical and General **39**(43), R465–R540 (2006)
6. Fiez, T., Sekar, S., Zheng, L., Ratliff, L.J.: Combinatorial bandits for incentivizing agents with dynamic preferences. arXiv preprint arXiv:1807.02297 (2018)
7. Guresti, B., Vanlioglu, A., Ure, N.K.: Iq-flow: Mechanism design for inducing cooperative behavior to self-interested agents in sequential social dilemmas (2023)
8. Ho, C.J., Slivkins, A., Vaughan, J.W.: Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. In: Proceedings of the fifteenth ACM conference on Economics and computation. pp. 359–376 (2014)
9. Laffont, J.J., Martimort, D.: The Theory of Incentives: The Principal-Agent Model. Princeton University Press, Princeton, NJ, USA (2001)
10. Manzo, G.: Agent-based models and causal inference. John Wiley & Sons (2022)
11. Mguni, D., Jennings, J., Sison, E., Valcarcel Macua, S., Ceppi, S., Munoz de Cote, E.: Coordinating the crowd: Inducing desirable equilibria in non-cooperative systems. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. p. 386–394. AAMAS '19, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2019)
12. Ratliff, L.J., Dong, R., Sekar, S., Fiez, T.: A perspective on incentive design: Challenges and opportunities. Annual Review of Control, Robotics, and Autonomous Systems **2**(1), 305–338 (2019)
13. Yang, J., Wang, E., Trivedi, R., Zhao, T., Zha, H.: Adaptive incentive design with multi-agent meta-gradient reinforcement learning. In: Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems. p. 1436–1445. AAMAS '22, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2022)
14. Zemkohoo, A., Dempe, S.: Bilevel optimization advances and next challenges (2020)
15. Zheng, S., Trott, A., Srinivasa, S., Parkes, D.C., Socher, R.: The ai economist: Taxation policy design via two-level deep multiagent reinforcement learning. Science advances **8**(18) (2022)