Causal Strategic Learning with Competitive Selection

Kiet Q. H. Vo[1,2], **Muneeb Aadil**[1,2], **Siu Lun Chau**[1], **Krikamol Muandet**[1]

[1]CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
[2]Saarland University, Saarbrücken, Germany
huynh.vo@cispa.de, s8muaadi@stud.uni-saarland.de, siu-lun.chau@cispa.de, muandet@cispa.de

Abstract

We study the problem of agent selection in causal strategic learning under multiple decision makers and address two key challenges that come with it. Firstly, while much of prior work focuses on studying a fixed pool of agents that remains static regardless of their evaluations, we consider the impact of selection procedure by which agents are not only evaluated, but also selected. When each decision maker unilaterally selects agents by maximising their own utility, we show that the optimal selection rule is a trade-off between selecting the best agents and providing incentives to maximise the agents' improvement. Furthermore, this optimal selection rule relies on incorrect predictions of agents' outcomes. Hence, we study the conditions under which a decision maker's optimal selection rule will not lead to deterioration of agents' outcome nor cause unjust reduction in agents' selection chance. To that end, we provide an analytical form of the optimal selection rule and a mechanism to retrieve the causal parameters from observational data, under certain assumptions on agents' behaviour. Secondly, when there are multiple decision makers, the interference between selection rules introduces another source of biases in estimating the underlying causal parameters. To address this problem, we provide a cooperative protocol which all decision makers must collectively adopt to recover the true causal parameters. Lastly, we complement our theoretical results with simulation studies. Our results highlight not only the importance of causal modeling as a strategy to mitigate the effect of gaming, as suggested by previous work, but also the need of a benevolent regulator to enable it.

1 Introduction

Machine Learning (ML) has gained significant popularity in facilitating personalised decision making across diverse domains such as healthcare (Wiens et al. 2019; Chau et al. 2021; Ghassemi and Mohamed 2022), criminal justice (Kleinberg et al. 2018), college admissions (Harris et al. 2022), hiring (Deshpande, Pan, and Foulds 2020), and credit scoring (Björkegren and Grissen 2020). In these critical domains, mutual trust between decision makers and agents who are affected by the decisions is of utmost importance. As a result, the decision makers might need to render algorithmic rules transparent to all stakeholders. However, this transparency can incentivise agents to strategically adjust their variables to receive more favorable decisions, resulting in either genuine improvements or gaming (Bechavod et al. 2021). Although in both scenarios agents receive better decision outcomes, gaming is undesirable for the decision makers as it negatively impacts their utility. Learning under strategic behavior is well-studied in both economics and machine learning (Hardt et al. 2016; Perdomo et al. 2020; Dranove et al. 2003; Dee et al. 2019; Munro 2022). Our work aligns with research efforts to identify causal features that reduce gaming effects and to promote genuine agent improvements (Miller, Milli, and Hardt 2020), an approach often referred to as causal strategic learning (CSL).

Let us consider a college admission example from Harris et al. (2022). The college, acting as the *decision maker (DM)*, aims to evaluate applicants (*agents*) by predicting their prospective college GPAs based on their submitted high school GPAs and SAT scores. For transparency, the college makes this evaluation rule public. In response, applicants can strategically direct their efforts on certain exams (high school or SAT) to optimise their evaluations. Recognising this strategic approach, the college's objective is to formulate and publicise an evaluation rule that maximises the expected college GPA (or *agents' outcome*) for all applicants. Envision a scenario where a student's college GPA is causally determined by their high school GPA only, yet the deployed rule considers both exam results. There is potential for gaming behavior under this rule, if an applicant emphasises their SAT preparation over their high school GPA, since this might boost their evaluation without necessarily improving the actual college academic performance.

The above example underscores the necessity of incorporating causal knowledge into decision making to incentivise agents towards genuine improvement, aligning with what Miller, Milli, and Hardt (2020) have proven. CSL presents numerous challenges. For example, Alon et al. (2020) explore mechanism designs that incentivise agents to respond with the intended outcomes of the DM, assuming knowledge of the true underlying causal structure. Similarly, Munro (2022) also assumes knowledge of casual information and incorporates stochasticity into their released decision rule to discourage gaming. However, without prior causal knowledge, learning the true causal mechanism in practice is challenging due to confounding bias in observational data. To address this, Shavit, Edelman, and Axelrod (2020) show that

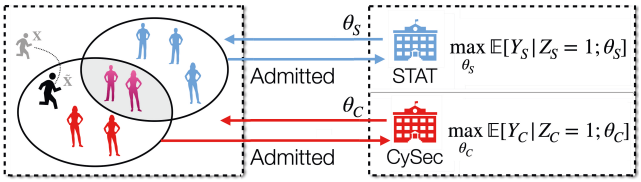


Figure 1: Causal strategic learning with two decision makers. The statistics (STAT) and cyber-security (CySec) departments, each applies selection rules $\boldsymbol{\delta}_{\theta_S}$ and $\boldsymbol{\delta}_{\theta_C}$ to maximise their enrolled students' expected outcomes (future GPA). The parameters θ_S and θ_C are made public, prompting future students to strategically alter their attributes ($\mathbf{X}$ to $\tilde{\mathbf{X}}$) to boost admission chances.

a DM can publicise a sequence of evaluation rules specifically to eliminate confounding bias and achieving causal identifiability. In contrast, Harris et al. (2022) consider scenarios in which the DM can utilise the evaluation rule itself as instrumental variable, and identify the true causal mechanism via instrumental variable regression (Angrist, Imbens, and Rubin 1996; Newey and Powell 2003; Hartford et al. 2017; Singh, Sahani, and Gretton 2019; Muandet et al. 2020). While much of previous CSL research focuses on evaluating (and motivating) agents in light of strategic feedback from a single DM's perspective, our research extends further, considering not just evaluating, but also selecting agents based on their evaluations. This brings in additional challenges, notably the introduction of selection bias, which undermines previous causal identifiability results. Additionally, we venture into situations with multiple DMs competing to select agents. We believe this work is well-motivated for real-world strategic learning scenarios that involve competitive selection, such as in hiring and loan application.

Continuing from our motivating example, consider that we now have multiple college departments (as DMs), e.g. statistics and cyber-security, competing not only to evaluate applicants but also to select them based on their evaluations (see Figure 1). Unlike previous methods, each department (DM) aims to optimise the expected GPA of their enrolled students, rather than focusing on all applicants. This natural objective nonetheless leads to a dilemma between selecting the top-performing candidates and motivating general candidates to improve. Furthermore, a selection rule focusing solely on top candidates can disincentivise self-improvement, potentially lowering future college GPAs (see Corollary 3.2). Additionally, as the optimal selection rule has to rely on incorrect (non-causal) predictions of agents' outcomes, their chances of being selected can be diminished compared to if evaluations were based on accurate (causal) predictions (see Corollary 3.3). We refer to an agent's prospective outcome and selection chances collectively as *agent welfare*. To safeguard such welfare, we adopt a regulator's viewpoint, proposing regulations for the DM to follow, such that their resulting optimal decision rule will lead to neither deterioration of agents' outcomes nor excessive reduction in agents' selection chance. As such regulation requires DM to have access to causal parameters, we provide conditions for a single DM to achieve causal identifiability under selection bias. With multiple DMs, the selection bias is now harder to correct for due to the interference between decision rules. In particular, it is difficult for any individual DM to predict an agent's strategic response when that agent is incentivised by all DMs. Additionally, anticipating their compliance behavior is challenging since this agent can adhere to at most one DM's positive decision. Consequently, we propose a cooperative protocol for the DMs to follow so that their causal parameters can be identified, to subsequently safeguard the welfare of agents.

The rest of the paper is outlined as follows. Section 2 introduces the CSL formulation with selection procedure under multiple DMs. Section 3 then discusses the impact of selection in the context of CSL alongside our main results and extensions to the setting of competitive selection. We validate our approach through various simulation studies in Section 4. Finally, we conclude in Section 5. All proofs are provided in the appendices.

2 CSL with Selection

Notations. We denote random variables and random vectors with upper case letters, and their realisations with lower case and bold lower case letters, respectively. Random matrices are also denoted with upper case letters, and their realisations with bold upper case. We write $\{1, \dots, n\}$ as $[n]$.

Following prior work (Shavit, Edelman, and Axelrod 2020; Harris et al. 2022; Bechavod et al. 2022), we build our setting on the sequential decision making context, following the framework of Stackelberg game. We assume throughout that there exist n decision makers (DMs), with $n \geq 1$, who take turn with agents playing their strategies over T rounds indexed by $t \in [T]$. Let W_{it} be a binary variable representing the decision from DM i for the sole agent who arrives at round t, e.g., whether or not the college i admits this student. At the beginning of each round, each DM publicises their decision rule $\boldsymbol{\delta}_{\boldsymbol{\theta}_{it}}$ parameterised by the parameter vector $\boldsymbol{\theta}_{it} \in \mathbb{R}^m$, i.e.,

$$\boldsymbol{\delta}_{\boldsymbol{\theta}_{it}} : \mathbf{x} \mapsto p\left(W_{it} = 1 \mid X_t = \mathbf{x}\,;\, \boldsymbol{\theta}_{it}\right), \quad i \in [n]$$

where $X_t \in \mathbb{R}^m$ denotes the random vector containing the covariates of the agent in round t and $p(W_{it} = 1 \,|\, X_t = \mathbf{x}\,;\, \boldsymbol{\theta}_{it})$ is a probability that this agent will later receive a positive decision, i.e., being admitted into the college, if they report attributes $X_t = \mathbf{x}$. We assume that $W_{it} \sim \text{Bernoulli}(\boldsymbol{\delta}_{\boldsymbol{\theta}_{it}}(X_t))$. After knowing about $\{\boldsymbol{\delta}_{\boldsymbol{\theta}_{it}}\}_{i=1}^n$, this agent modifies their attributes and then reports the final values $\mathbf{x}$, e.g., SAT score and high school GPA, to all DMs, so as to maximise the chance of receiving favorable decisions. Next, all DMs evaluate this agent using their decision rules and return the selection statuses $\{w_{it}\}_{i=1}^n$. Finally, the agent's compliance to the decisions can be modeled as a random variable $Z_t \sim \text{Categorical}(\{0\} \cup [n])$, whose value dictates which positive decision the agent will comply with[1]. Throughout this work, we focus on the *perfect information* setting where both DMs and agents know

[1]When $Z_t = 0$, the agent either does not comply with any of the positive decisions or does not receive any positive decision.

all information about the decision rules including their parameter vectors (Shavit, Edelman, and Axelrod 2020; Harris et al. 2022). Specifically, for round t_s, the agent knows about $\{\boldsymbol{\delta}_{\boldsymbol{\theta}_{it_s}}, \boldsymbol{\theta}_{it_s}\}_{i=1}^n$ and all DMs know about $\{\{\boldsymbol{\delta}_{\boldsymbol{\theta}_{it}}, \boldsymbol{\theta}_{it}\}_{i=1}^n\}_{t=1}^{t_s-1}$.

Following Harris et al. (2022), we assume that the potential outcome of an agent, $Y_{it} \in \mathbb{R}$, e.g., their future GPA, in any environment i is a linear function of their covariates: $Y_{it} := X_t^\top \boldsymbol{\theta}_i^* + O_{it}$ where $\boldsymbol{\theta}_i^* \in \mathbb{R}^m$ is the true causal parameter vector that maps the covariates X_t to the outcome $Y_{it} \in \mathbb{R}$ and O_{it} is the unobserved noise. In practice, the DMs lack access to the true $\boldsymbol{\theta}_i^*$, so each of them bases their decision on the predicted outcome $\hat{y}_{it} = \mathbf{x}^\top \boldsymbol{\theta}_{it}$ using the agent's covariates $X_t = \mathbf{x}$ where $\boldsymbol{\theta}_{it}$ is a parameter estimate. Finally, we assume that the covariates X_t is a linear function of an agent's baseline and their strategic improvement, namely $X_t := B_t + \mathcal{E}_t \mathbf{a}_t$ where the conversion matrix $\mathcal{E}_t \in \mathbb{R}^{m \times d}$ translates their strategic action $\mathbf{a}_t \in \mathbb{R}^d$ into the improvement upon the baseline $B_t \in \mathbb{R}^m$. The unobserved noise O_{it} is correlated with the agent's baseline B_t and is specific to the environment i, which can be due to the private type of each agent, e.g., a student's socioeconomic background, that can further influence their academic baseline B_t and their cultural fit O_{it} in this environment.

Agents' utilities. Since each agent has access to multiple predicted outcomes $\hat{y}_{it}$ (where $i \in [n]$) alongside their preferred environments, we assume that the agent t aims at maximising the following utility function

$$u(\mathbf{a}_t) := \sum_{i=1}^{n} \gamma_{it} \hat{y}_{it}(\mathbf{a}_t) - \frac{1}{2}\|\mathbf{a}_t\|_2^2 \quad \text{with } \gamma_{it} \geq 0,\ \forall i, t \quad (1)$$

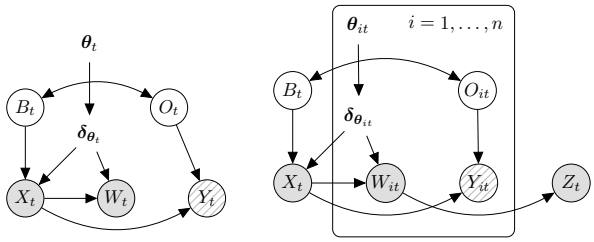
in each round t after being informed of the parameter vectors, where $\{\gamma_{it}\}_{i=1}^n$ represents the preference of this agent. Unlike previous work (Shavit, Edelman, and Axelrod 2020; Harris et al. 2022; Bechavod et al. 2022), the utility function (1) also involves the agent's preference over multiple DMs. For any list of parameter vectors $\{\boldsymbol{\theta}_{1t}, \boldsymbol{\theta}_{2t}, \ldots, \boldsymbol{\theta}_{nt}\}$, it is not difficult to see that the maximiser of (1) is $\mathbf{a}_t = \mathcal{E}_t^\top (\sum_{i=1}^n \gamma_{it} \boldsymbol{\theta}_{it})$; see Appendix A.1 for the full proof.

Decision makers' objectives. We assume that the DMs are utility maximisers each of whom aims to maximise the expected future outcome of the agents that comply with their decisions. Without loss of generality, we specify the objective function for an arbitrary DM i:

$$\max_{\boldsymbol{\theta}_{it}} \mathbb{E}\left[Y_{it}(\{\boldsymbol{\theta}_t^{-i}, \boldsymbol{\theta}_{it}\}) \mid Z_t = i\ ;\ \boldsymbol{\theta}_{it}\right] \quad (2)$$

where we use $\{\boldsymbol{\theta}_t^{-i}, \boldsymbol{\theta}_{it}\}$, $\boldsymbol{\theta}_t^{\text{all}}$, or $\{\boldsymbol{\theta}_{it}\}_{i=1}^n$ to denote a collection of parameters associated with the deployed selection rules. We use the notation $Y_{it}(\{\boldsymbol{\theta}_t^{-i}, \boldsymbol{\theta}_{it}\})$ to highlight that the outcome variable is a function of all parameters $\boldsymbol{\theta}_t^{\text{all}}$ due to agents' strategic behaviour. Furthermore, notice that the expectation also depends on the conditional distribution of the rival DMs' parameters, $p(\boldsymbol{\theta}_t^{-i} | Z_t = i, \boldsymbol{\theta}_{it})$. More detailed discussion will follow in subsequent sections.

In summary, our approach distinguishes itself from previous work in causal strategic learning mainly by its integration of the selection variable W_{it} within a competitive context involving multiple DMs. Figure 2 illustrates the causal graphs associated with our novel setting.



(a) Selection Variable (b) Competitive Selection

Figure 2: Causal graphs for our settings with selection variable W_t (left) and multiple decision makers (right). The patterned nodes Y_t and Y_{it} represent the partial observability nature of these variables.

3 Main Results

Our main results are based on the following two homogeneity assumptions on the strategic responses of agents.

H1. *Homogeneous effort conversion*: for all $t \in [T]$, $\mathcal{E}_t = \boldsymbol{\mathcal{E}}$ for some conversion matrix $\boldsymbol{\mathcal{E}}$.

H2. *Homogeneous preference and compliance*: for each DM i and for all $t \in [T]$, $\gamma_{it} = \gamma_i$ for some $\gamma_i \geq 0$ and $Z_t \perp\!\!\!\perp \{X_t, B_t\} \mid \{W_{it}\}_{i=1}^n$.

The former condition suggests that all agents exhibit the same strategic response regardless of their individual baselines, i.e., they only differ by their baselines B_t, while the latter condition implies that all agents share the same preference over the n DMs, and any two agents will demonstrate identical compliance behavior based on the given set of selection statuses $\{w_{it}\}_{i=1}^n$. In the context of college admission, the common preference $\{\gamma_i\}_{i=1}^n$ may naturally align with the prestige of the colleges. Intuitively, these two assumptions suggest that while strategic responses may encompass both common and idiosyncratic elements, we solely concentrate on the common part, simplifying our theoretical analyses at the cost of potentially overlooking significant individual variations of agents' strategic behaviour.

Our work thus concerns itself with a *partially* heterogeneous setting. On the contrary, when completely heterogeneous agents are subjected to selection, many variables are rendered dependent; see, e.g., eq. (3), making our theoretical analyses much more cumbersome. However, such homogeneity assumptions do not undermine the impact of selection that we discuss throughout this section since it is likely to persist in a more complex setting. This impact includes the trade-off between choosing capable agents and providing a maximal incentive, e.g., Corollary 3.2, and the selection bias, e.g., Theorem 3.4. To understand the impact of these two assumptions, we provide the sensitivity analyses in Appendix F.2. A relaxation of these assumptions will be considered in future work.

3.1 Impact of Selection Procedure

To illustrate the impact of the selection procedure, we commence with the single DM setting, i.e. $n = 1$. For simplicity, we omit the subscript i and assume that all agents comply with the decisions they receive. Figure 2a shows the associated causal graph. The objective (2), i.e., $\mathbb{E}[Y_t \mid W_t = 1; \boldsymbol{\theta}_t]$, for single DM then becomes

$$\mathbb{E}\left[B_t^\top \boldsymbol{\theta}^* + O_t \mid W_t = 1; \boldsymbol{\theta}_t\right] + \mathbb{E}\left[\left(\mathcal{E}_t \mathbf{a}_t\right)^\top \boldsymbol{\theta}^* \mid W_t = 1; \boldsymbol{\theta}_t\right] = \text{cBP}(\boldsymbol{\theta}_t) + \text{cPI}(\boldsymbol{\theta}_t) \quad (3)$$

where the first and second terms on the right-hand side are referred to as the *conditional base performance* (cBP) and *conditional performance improvement* (cPI), respectively. The former pertains to the agent outcome without strategic behavior, while the latter represents the improvement achieved through strategic behavior. Both cBP and cPI are defined as expected values over the admitted agents, making them functions of the selection parameter $\boldsymbol{\theta}_t$. Additionally, the complexity of cBP and cPI relies on the chosen selection function $\boldsymbol{\delta}_{\boldsymbol{\theta}_t}$. Our objective (3) differs from the marginal expected outcome commonly studied in prior work, where no selection occurs (Shavit, Edelman, and Axelrod 2020; Harris et al. 2022; Bechavod et al. 2022):

$$\mathbb{E}[Y_t;\, \boldsymbol{\theta}_t] = \mathbb{E}\big[B_t^\top \boldsymbol{\theta}^* + O_t\big] + \mathbb{E}\big[\,\left(\mathcal{E}_t \mathbf{a}_t\right)^\top \boldsymbol{\theta}^*;\, \boldsymbol{\theta}_t\big]. \quad (4)$$

We refer to the two terms on the right-hand side of (4) similarly as the *marginal base performance* (mBP) and *marginal performance improvement* (mPI). Observe that maximising (4) amounts to maximising only the mPI, whereas maximising our objective (3) might involve a trade-off between maximising cBP and cPI, as shown below.

Utility maximisation. We further impose the following two assumptions, exclusively for utility maximisation:

S1. **Linear effect**: The selection yields a linear structure of cBP as follows: $\text{cBP}(\boldsymbol{\theta}_t) = \boldsymbol{\alpha}^\top \boldsymbol{\theta}_t + \beta$ for some vector $\boldsymbol{\alpha} \in \mathbb{R}^m$ and constant $\beta \in \mathbb{R}$;

S2. **Bounded parameters**: For all $\boldsymbol{\theta}_t$, $\|\boldsymbol{\theta}_t\|_2 \leq 1$ (Shavit, Edelman, and Axelrod 2020).

On the one hand, Assumption S1 allows us to further simplify the analysis of the DM's behaviour and to further simplify the demonstration of the trade-off between choosing agents and incentivising them, which we discuss later. Even when S1 does not hold, this will only complicate the analysis without changing the implication resulted from Corollary 3.2 and Corollary 3.3. On the other hand, as $\mathcal{Q}(\boldsymbol{\theta}_t)$ is not scale-invariant, we adopt Assumption S2, which was also used by previous work such as Shavit, Edelman, and Axelrod (2020) and Bechavod et al. (2022). As a result, this allows us to restrict $\boldsymbol{\theta}_t$ to some arbitrarily small region and justifies a linear approximation to $\text{cBP}(\boldsymbol{\theta}_t)$. Nevertheless, we acknowledge the limitation of these assumptions and provide a more detailed discussion in Appendix F.2.

We denote the objective (3) by $\mathcal{Q}(\boldsymbol{\theta}_t)$ and expand it as

$$\mathcal{Q}(\boldsymbol{\theta}_t) := \text{cBP}(\boldsymbol{\theta}_t) + \text{cPI}(\boldsymbol{\theta}_t) = \left(\boldsymbol{\alpha}^\top \boldsymbol{\theta}_t + \beta\right) + \gamma \boldsymbol{\theta}_t^\top \boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*$$

where we used the fact that $\mathbf{a}_t^\top = \gamma \boldsymbol{\theta}_t^\top \mathcal{E}_t$ and $\mathcal{E}_t = \boldsymbol{\mathcal{E}}$ as a result of Assumption (H1). Then, we formally state the optimal behaviour of the DM with the next theorem.

Theorem 3.1 (Bounded optimum)**.** *Suppose Assumptions (H1), (H2), (S1), and (S2) hold. Then, the optimal parameter vector for the DM can be expressed as* $\boldsymbol{\theta}^{AO} = (\boldsymbol{\alpha} + \gamma \boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*)/\|\boldsymbol{\alpha} + \gamma \boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*\|_2$.

Since $\mathcal{Q}(\boldsymbol{\theta}_t)$ is linear in $\boldsymbol{\theta}_t$, the DM can obtain an unnormalised version of $\boldsymbol{\theta}^{\text{AO}}$ by regressing $\mathcal{Q}(\boldsymbol{\theta}_t)$ onto $\boldsymbol{\theta}_t$ using ordinary least squares (OLS) regression. As shown in Theorem 3.1, the optimal selection parameter $\boldsymbol{\theta}^{\text{AO}}$ is determined by the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*$ from cBP and cPI, respectively. Intuitively, this implies that an optimal selection rule might be a trade-off between selecting the best agents and incentivising agents to maximise their improvement. Figure 4 (in the supplementary material) illustrates when this trade-off happens and there exists no $\boldsymbol{\theta}_t$ for which both $\text{cBP}(\boldsymbol{\theta}_t)$ and $\text{cPI}(\boldsymbol{\theta}_t)$ are maximised simultaneously. The next corollary formalises this intuition.

Corollary 3.2 (Maximum improvement)**.** *Suppose Assumptions (H1), (H2), (S1), (S2) hold, and* $\gamma > 0$*. If* $\boldsymbol{\alpha} = (k - \gamma) \boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*$ *for some* $k > 0$*, then the maximiser of* $\mathcal{Q}(\boldsymbol{\theta}_t)$ *is also the maximiser of* $\text{cPI}(\boldsymbol{\theta}_t)$*.*

Generally speaking, the vector $\boldsymbol{\alpha}$ represents the causal mechanism translating $\boldsymbol{\theta}_t$ into the base performance of the chosen agents, i.e., $\text{cBP}(\boldsymbol{\theta}_t)$, whereas the vector $\boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*$ denotes the causal mechanism translating $\boldsymbol{\theta}_t$ into the performance improvement of the selected agents, i.e., $\text{cPI}(\boldsymbol{\theta}_t)$. Hence, $\boldsymbol{\theta}_t$ serves not only as a selection parameter but also as the incentive for agents' improvement. From Corollary 3.2, when $k > \gamma$, the two aforementioned causal mechanisms align with each other, i.e., $\cos(\boldsymbol{\alpha}, \boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*) = 1$, then $\boldsymbol{\theta}^{\text{AO}}$ not only selects the best agents (i.e., in terms of cBP) but also is the incentive that maximises their improvement.

Safeguarding the social welfare. There is therefore a possibility that the deployed selection rule may result in undesirable societal outcomes. For instance, this would involve rejecting agents who, with proper incentivisation, could have been chosen. Another example is when a decision rule selects the best agents but incidentally discourages them from further improvement, which corresponds to the case when $\cos(\boldsymbol{\theta}^{\text{AO}}, \boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}} \boldsymbol{\theta}^*) = -1$. To prevent such situations, a benevolent regulator may opt to enforce a regulation such that a decision rule must result in $\cos(\boldsymbol{\theta}^{\text{AO}}, \boldsymbol{\mathcal{E}} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*) > 0$, thereby guaranteeing that the optimal parameters $\boldsymbol{\theta}^{\text{AO}}$ do not lead to a decline in the selected agents' outcome.

In addition to the inherent trade-off induced by the selection process, Theorem 3.1 also shows that $\boldsymbol{\theta}^{\text{AO}}$ differs from the true causal parameters $\boldsymbol{\theta}^*$ in general. Relying on a selection criterion that uses the optimal parameters $\boldsymbol{\theta}^{\text{AO}}$ results in consistently inaccurate predictions of agents' outcomes. This unjustly reduces an agent's admission chance, compared to when the causal parameters were employed instead. The following corollary then outlines conditions under which the reduction in an arbitrary agent's admission chance can be bounded when DM utilises $\boldsymbol{\theta}^{\text{AO}}$ as the selection parameter, instead of the causal parameter $\boldsymbol{\theta}^*$.

Corollary 3.3 (Bounded reduction)**.** *Suppose Assumptions (H1), (H2), (S1), (S2) hold, and the DM considers only two choices* $\boldsymbol{\theta}_t \in \{\boldsymbol{\theta}^*, \boldsymbol{\theta}^{AO}\}$*. Assume further that: (1)* $\|\boldsymbol{\theta}^*\|_2 \leq 1$*;*

(2) $\boldsymbol{\alpha} = (k-\gamma)\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top\boldsymbol{\theta}^*$ *with* $k, \gamma > 0$*; (3)* $B_t \sim \mathcal{N}(0, \sigma^2 I)$ *where* I *is the identity matrix and* $\sigma \in \mathbb{R}^+$*; (4) The selection function,* $\boldsymbol{\delta}(\mathbf{x}; \boldsymbol{\theta}_t) := \widetilde{\boldsymbol{\delta}}(\hat{y}_t(\boldsymbol{\theta}_t))$*, is increasing in* $\hat{y}_t$ *and is Lipschitz continuous, i.e.* $|\widetilde{\boldsymbol{\delta}}(\hat{y}) - \widetilde{\boldsymbol{\delta}}(\hat{y}')| \leq L|\hat{y} - \hat{y}'|$ *for* $L > 0$*. Then, for any* $M > 0$*:* $p\left(\xi(\boldsymbol{\theta}^*) - \xi(\boldsymbol{\theta}^{AO}) > M\right) \leq \Phi\left(\frac{-M/L - \lambda}{\sigma\|\boldsymbol{\theta}^{AO} - \boldsymbol{\theta}^*\|_2}\right)$ *where* $\xi(\boldsymbol{\theta}_t) := p(W_t = 1 \,|\, B_t; \boldsymbol{\theta}_t)$ *denotes the admission chance of the agent* t*,* Φ *is the CDF of* $\mathcal{N}(0,1)$*, and* $\lambda := \gamma((\boldsymbol{\theta}^{AO})^\top \boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^{AO} - (\boldsymbol{\theta}^*)^\top \boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}^*)$.

Specifically, this corollary rewrites the admission probability of an agent in terms of their baseline B_t and denotes it with $\xi(\boldsymbol{\theta}_t)$. When the counterfactual quantity $\xi(\boldsymbol{\theta}^*) - \xi(\boldsymbol{\theta}^{\text{AO}})$ is positive for an agent, it implies that the admission chance of this agent will be reduced if the DM employs $\boldsymbol{\theta}^{\text{AO}}$ instead of $\boldsymbol{\theta}^*$. As a result, this corollary gives us an upper bound for the probability of this reduction for an unknown agent.

Causal parameters estimation. As shown in Corollary 3.2 and Corollary 3.3, it is necessary for the DM to know the incentivising causal mechanism, $\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top\boldsymbol{\theta}^*$, in order to comply with the regulations, and for the regulator to know $\boldsymbol{\theta}^*$ in order to verify the conditions of Corollary 3.3. Unfortunately, unbiased estimation of $\boldsymbol{\theta}^*$ from observational data alone is impossible without imposing further assumptions on the data-generating process (Peters, Janzing, and Schölkopf 2017). In our case, unobserved common causes of the outcome Y_t and covariates X_t create dependencies between X_t and O_t, rendering it impossible to estimate $\boldsymbol{\theta}^*$ consistently via the OLS regression. To this end, Harris et al. (2022) proposes to view $\boldsymbol{\theta}_t$ as an instrumental variable (IV) and subsequently applies a two-stage least square (2SLS) regression (Cameron and Trivedi 2005) to estimate $\boldsymbol{\theta}^*$. However, existing IV regression approaches are not suitable for our setting because the DM can only observe the outcomes of the selected agents, violating the unconfoundedness assumption of the IV; see Appendix B for the proof.

In what follows, we present an alternative approach to estimate $\boldsymbol{\theta}^*$. This approach can be readily adapted to directly estimate $\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top\boldsymbol{\theta}^*$. To that end, we first consider a ranking-based selection rule that is commonly deployed in practice.

Definition 1 (Ranking selection)**.** *The DM selects an agent* t *based on their relative ranking compared to other agents who are subject to the same selection parameters* $\boldsymbol{\theta}_t$*. Specifically,* $\boldsymbol{\delta}_{\boldsymbol{\theta}_t}(\mathbf{x}) = p\left(X_t^\top \boldsymbol{\theta}_t \leq \mathbf{x}^\top \boldsymbol{\theta}_t\right) = \text{CDF}_{X_t^\top \boldsymbol{\theta}_t}\left(\mathbf{x}^\top \boldsymbol{\theta}_t\right)$.

Based on this selection rule, the higher an agent's evaluation (relative to their peers) the more likely they will be selected. Note that in this work, we do not restrict the DM to the ranking selection rule for utility maximisation. This ranking selection rule is only provided so that the DM can retrieve the true causal parameters, which are useful for designing subsequent selection rules. The next theorem provides an unbiased estimate of the true causal parameter $\boldsymbol{\theta}^*$ in our setting with a selection variable.

Theorem 3.4 (Local exogeneity)**.** *Under Assumptions (H1) & (H2), if there exists a pair of rounds* t *and* t' *such that* $\boldsymbol{\theta}_t = k\boldsymbol{\theta}_{t'}$ *for some* $k > 0$*, then we have:* $\mathbb{E}\left[Y_t \,|\, W_t = 1\,;\, \boldsymbol{\theta}_t\right] - \mathbb{E}\left[Y_{t'} \,|\, W_{t'} = 1\,;\, \boldsymbol{\theta}_{t'}\right] = \left(\mathbb{E}\left[X_t \,|\, W_t = 1\,;\, \boldsymbol{\theta}_t\right] - \mathbb{E}\left[X_{t'} \,|\, W_{t'} = 1\,;\, \boldsymbol{\theta}_{t'}\right]\right)^\top \boldsymbol{\theta}^*$

Intuitively, when all agents exert an equal amount of effort, ranking their covariates X_t is equivalent to ranking their baselines B_t. Therefore, multiplying $\boldsymbol{\theta}_t$ by a positive scalar preserves the ranking. Consequently, we obtain a linear equation that contains no endogenous noise (from Theorem 3.4), allowing for unbiased estimation of the true causal parameters $\boldsymbol{\theta}^*$. Specifically, if we refer to the left-hand side as $\Delta\bar{y}$ and the coefficient on the right-hand side as $\Delta\bar{\mathbf{x}}$, then $\boldsymbol{\theta}^*$ can be estimated by regressing $\Delta\bar{y}$ onto $\Delta\bar{\mathbf{x}}$. We refer to this procedure as Mean-shift Linear Regression (MSLR) and discuss a sample algorithm in Section 4.

3.2 Impact of Competitive Selection

When there are multiple DMs ($n \geq 2$), selecting an agent becomes competitive as their incentives affect the agent's covariates X_t simultaneously. Also, whether an agent complies with any DM is also influenced by other DMs' decisions. Consequently, additional assumptions are required to safeguard the agent's welfare as before.

We assume each DM aims at unilaterally maximising the expected outcome of their own agents. We denote the objective of DM i as $\max_{\boldsymbol{\theta}_{it}} \mathcal{Q}_i\left(\boldsymbol{\theta}_t^{all}\right) = \max_{\boldsymbol{\theta}_{it}} \mathbb{E}\left[Y_{it} \mid Z_t = i; \boldsymbol{\theta}_t^{\text{all}}\right]$ and expand the expectation as

$$\mathbb{E}\left[B_t^\top \boldsymbol{\theta}_i^* + O_{it} \mid Z_t = i; \boldsymbol{\theta}_t^{\text{all}}\right] + \Big(\sum_{j=1}^{n} \gamma_j \boldsymbol{\theta}_{jt}\Big)^\top \boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_i^* \quad (5)$$

where the first and second terms are denoted similarly as $\text{cBP}_i(\boldsymbol{\theta}_t^{\text{all}})$ and $\text{cPI}_i(\boldsymbol{\theta}_t^{\text{all}})$, respectively, and $\mathbf{a}_t^\top = \left(\sum_{j=1}^n \gamma_{jt}\boldsymbol{\theta}_{jt}\right)^\top \boldsymbol{\mathcal{E}}$ is again the agents' optimal strategic action. We highlight that the objective (5) depends not only on $\boldsymbol{\theta}_{it}$ but also on $\boldsymbol{\theta}_t^{-i}$ due to the interaction between DMs via competitive selection. As a result, $\mathcal{Q}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_t^{-i}\})$ can be seen as a family of objective functions parameterised by $\boldsymbol{\theta}_t^{-i}$. When DM i is an expected-utility maximiser, we would maximise the expectation of $\mathcal{Q}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_t^{-i}\})$ to marginalise out the effect of $\boldsymbol{\theta}_t^{-i}$. However, this requires knowledge on the conditional distribution $p(\boldsymbol{\theta}_t^{-i} | Z_t = i, \boldsymbol{\theta}_{it})$ which the DM i does not have. To tackle this challenge, we consider the worst-case scenario in which all rival DMs cooperate to minimise the objective and study how DM i can in response maximise this worst-case objective function. We show in Appendix C that our solution, specifically in this case, is also a maximin strategy of the DM i.

Utility maximisation. The objective (5) is difficult to optimise as it depends not only on the choice of selection rules but also on the behaviour of other DM's objective. To simplify the analysis, we rely on the following assumptions, exclusively for utility maximisation:

M1. **Partially additive interaction between DMs**: For an arbitrary DM i, their cBP_i can be decomposed as $\text{cBP}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_t^{-i}\}) = g_i(\boldsymbol{\theta}_{it}) + h_i(\boldsymbol{\theta}_t^{-i}) + c_i$ for some function g_i, h_i and constant c_i.

M2. **Linear self-effect**: The contribution of DM i to cBP_i admits a linear structure, i.e. $g_i(\boldsymbol{\theta}_{it}) = \boldsymbol{\alpha}_i^\top \boldsymbol{\theta}_{it} + \beta_i$ for some vector $\boldsymbol{\alpha}_i \in \mathbb{R}^m$ and constant $\beta_i \in \mathbb{R}$;

M3. **Bounded parameters**: For all $\boldsymbol{\theta}_{it}$, $\|\boldsymbol{\theta}_{it}\|_2 \leq 1$.

Assumptions (M2) & (M3) are extensions of (S1) and (S2), whereas (M1) is an additional assumption.

Proposition 1 (Dominant strategy)**.** *Suppose Assumptions (H1), (H2), and (M1) hold for DM i. Then,* $\arg\max_{\boldsymbol{\theta}_{it}} \mathcal{Q}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_\bullet^{-i}\}) = \arg\max_{\boldsymbol{\theta}_{it}} \mathcal{Q}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_\diamond^{-i}\})$ *for any pair of distinct values $\boldsymbol{\theta}_\bullet^{-i}$ and $\boldsymbol{\theta}_\diamond^{-i}$ of $\boldsymbol{\theta}_t^{-i}$.*

This proposition shows that the monotonicity of the objective $\mathcal{Q}_i$ remains unaffected regardless of other values $\boldsymbol{\theta}_{jt}$ released by the DMs j where $j \neq i$. Based on Assumption (M1) and this result, any DM i is provided with a condition to maximise all the objective functions within the family $\mathcal{Q}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_t^{-i}\})$ simultaneously.

Theorem 3.5 (Bounded optimum, extended)**.** *Suppose that (H1), (H2), and (M1)-(M3) hold. Then, the optimal parameter vector for any DM i takes the form $\boldsymbol{\theta}_i^{AO} = (\boldsymbol{\alpha}_i + \gamma_i \boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_i^*)/\|\boldsymbol{\alpha}_i + \gamma_i \boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_i^*\|_2$.*

As each objective function $\mathcal{Q}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_\diamond^{-i}\})$, conditioned on some arbitrary $\boldsymbol{\theta}_t^{-i} := \boldsymbol{\theta}_\diamond^{-i}$, is a linear function of $\boldsymbol{\theta}_{it}$, the DM i can obtain an un-normalised version of $\boldsymbol{\theta}_i^{\text{AO}}$ by regressing $\mathcal{Q}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_\diamond^{-i}\})$ onto $\boldsymbol{\theta}_{it}$ using the OLS regression.

Safeguarding the social welfare. Like the previous setting, we provide below an extension of Corollary 3.2 on maximum agents' improvement, with the difference being that $\boldsymbol{\theta}_i^{\text{AO}}$ is also a dominant strategy for cPI$(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_t^{-i}\})$.

Corollary 3.6 (Maximum improvement, extended)**.** *Suppose Assumptions (H1), (H2), and (M1)-(M3) hold for an arbitrary DM i, and that $\gamma_i > 0$. If $\boldsymbol{\alpha}_i = (k_i - \gamma_i)\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_i^*$ for some $k_i > 0$, then $\boldsymbol{\theta}_i^{AO}$ maximises both $\mathcal{Q}_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_t^{-i}\})$ and cPI$_i(\{\boldsymbol{\theta}_{it}, \boldsymbol{\theta}_t^{-i}\})$, regardless of $\boldsymbol{\theta}_t^{-i}$.*

As a result, if the interaction between DMs and agents exhibits additive structures, regulations can be solely imposed on DM i to ensure improved average outcome of the agents who are selected by (and comply with) DM i. Next, we extend Corollary 3.3 regarding agents' admission chance for the environment i.

Corollary 3.7 (Bounded reduction, extended)**.** *Suppose assumptions (H1), (H2), (M1)-(M3) hold for all DMs and each DM considers only two choices $\boldsymbol{\theta}_{jt} \in \{\boldsymbol{\theta}_j^*, \boldsymbol{\theta}_j^{AO}\}$ for $j \in [n]$. Let i be an arbitrary DM and assume further that: (1) $\|\boldsymbol{\theta}_i^*\|_2 \leq 1$; (2) $\boldsymbol{\alpha}_j = (k_j - \gamma_j)\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_j^*$ with $k_j, \gamma_j > 0$ for $j \in [n]$; (3) $\boldsymbol{\theta}_{jt}^\top \boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top (\boldsymbol{\theta}_i^{AO} - \boldsymbol{\theta}_i^*) \geq 0$ for $j \neq i$; (4) $B_t \sim \mathcal{N}(0, \sigma^2 I)$; (5) The selection function $\boldsymbol{\delta}_i(\mathbf{x}; \boldsymbol{\theta}_{it}) := \widetilde{\boldsymbol{\delta}}_i(\hat{y}_{it}(\boldsymbol{\theta}_t^{all}))$ is increasing in $\hat{y}_{it}$ and is Lipschitz continuous, i.e., $|\widetilde{\boldsymbol{\delta}}_i(\hat{y}_{it}) - \widetilde{\boldsymbol{\delta}}_i(\hat{y}'_{it})| \leq L|\hat{y}_{it} - \hat{y}'_{it}|$ for $L > 0$. Then, for any $M > 0$: $p\big(\xi_i(\boldsymbol{\theta}_\bullet^{all}) - \xi_i(\boldsymbol{\theta}_\diamond^{all}) > M\big) \leq \Phi\Big(\frac{-M/L - \lambda}{\sigma\|\boldsymbol{\theta}_i^{AO} - \boldsymbol{\theta}_i^*\|_2}\Big)$ where $\boldsymbol{\theta}_\bullet^{all} = \{\boldsymbol{\theta}_t^{-i}, \boldsymbol{\theta}_i^*\}$, $\boldsymbol{\theta}_\diamond^{all} = \{\boldsymbol{\theta}_t^{-i}, \boldsymbol{\theta}_i^{AO}\}$, and $\xi_j(\boldsymbol{\theta}_t^{all}) := p(W_{jt} = 1 \,|\, B_t; \boldsymbol{\theta}_t^{all})$ denotes the admission chance of the agent t from the DM j, given released decision parameters $\boldsymbol{\theta}_t^{all}$. The function Φ is the CDF of $\mathcal{N}(0,1)$ and $\lambda := \big(\mathbf{a}_t(\boldsymbol{\theta}_\diamond^{all})\big)^\top \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_i^{AO} - \big(\mathbf{a}_t(\boldsymbol{\theta}_\bullet^{all})\big)^\top \boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_i^*$.*

As demonstrated in the corollary, this extension necessitates additional conditions compared to Corollary 3.3 to ensure the protection of agents' chances of being selected, thus effectively highlighting the impact of competitive selection. Firstly, as shown in Section 2, the agent's best response is influenced by all DMs in which the agent has an interest, i.e., $\gamma_j > 0$. This dynamic creates competition among DMs to incentivise agents effectively. Secondly, the compliance status z_t of an agent depends on its selection statuses, $\{w_{jt}\}_{j=1}^n$, which in turn are affected by the selection rules $\boldsymbol{\delta}_{\boldsymbol{\theta}_{jt}}$ for $j \in [n]$, resulting in another competition in evaluating and selecting agents.

The third condition in Corollary 3.7 is of particular interest. Recall that under this corollary, any $\boldsymbol{\theta}_j^{\text{AO}}$ is simply a normalised version of $\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_j^*$. Consequently, this condition implies that when agents prefer only DMs whose environments are sufficiently similar to each other (reflected via the $\boldsymbol{\theta}_j^*$ and $\boldsymbol{\theta}_i^*$), then the benevolent regulator can enforce the regulation as outlined in Corollary 3.6 for all DMs to protect the agents from unjust reductions in admission chances. We provide a more detailed discussion on this third condition in Appendix D. We delay the discussion on an agent's admission chance into other environments $j \neq i$ in the appendix.

Causal parameters estimation. Similar to the previous setting, all DMs must have access to the causal mechanism $\boldsymbol{\mathcal{E}}\boldsymbol{\mathcal{E}}^\top \boldsymbol{\theta}_i^*$ to be able to comply with the regulation in Corollary 3.6 and the regulator must know $\boldsymbol{\theta}_i^*$ to let agents know which environments are sufficiently similar (Corollary 3.7). Unfortunately, DMs encounter additional challenges estimating causal parameters $\{\boldsymbol{\theta}_i^*\}_{i=1}^n$ under competitive selection. Specifically, they cannot correct the selection bias alone due to the interference with other DMs, which we demonstrate empirically in the following section. To address this, we propose a cooperative protocol for the regulator. This ensures unbiased estimation of causal parameters for all DMs.

Definition 2 (Cooperative protocol)**.** *Let $[n]$ be the set of all n DMs. If for any two arbitrary rounds t and t', a non-empty subset of DMs, $S \subseteq [n]$, employs the ranking selection rule (Definition 1) and their respective decision parameters satisfy $\boldsymbol{\theta}_{it} = k_i \boldsymbol{\theta}_{it'}$ for some $k_i > 0$ for all $i \in S$, then we say that S follows the cooperative protocol in these two rounds.*

This cooperative protocol extends the condition in Theorem 3.4 and suggests that DMs should synchronise the releases of their positively scaled parameters. When all DMs follow the cooperative protocol for multiple pairs of rounds, they have a way to obtain unbiased estimates of the causal parameters $\boldsymbol{\theta}_i^*$ as shown in the next theorem.

Theorem 3.8 (Local exogeneity, extended)**.** *Under Assumptions (H1) and (H2). If all DMs follow the cooperative protocol (i.e., $\exists S : S = [n]$ in Definition 2) in two rounds t and t', then:* $\mathbb{E}\left[Y_{it} \,|\, Z_t = i\,;\, \boldsymbol{\theta}_t^{all}\right] - \mathbb{E}\left[Y_{it'} \,|\, Z_{t'} = i\,;\, \boldsymbol{\theta}_{t'}^{all}\right] = \Big(\mathbb{E}\left[X_t \,|\, Z_t = i\,;\, \boldsymbol{\theta}_t^{all}\right] - \mathbb{E}\left[X_{t'} \,|\, Z_{t'} = i\,;\, \boldsymbol{\theta}_{t'}^{all}\right]\Big)^\top \boldsymbol{\theta}_i^*$.

Recall that in the previous setting, the ranking of agents can be preserved by scaling the selection parameters with a positive scalar. With the cooperative protocol and Assumption (H2), we can now also preserve the enrollment distribution of agents. We can then deploy the same MSLR procedure from the single DM settings to retrieve the causal parameters. Further details are discussed in Section 4.

Algorithm 1: Mean-shift Linear Regression (MSLR)

Require: a subset of n_s decision makers out of all n decision makers, where $1 \leq n_s \leq n$. These decision makers use ranking selection (Definition 1).
Parameters: number of rounds T, block's length η.

1: $D_i \leftarrow \{\}$ for $i = 1, \ldots, n_s$
2: **for** $t \in \{1, \ldots, T\}$ **do**
3: blockindex $\leftarrow \lfloor t/(\eta+1) \rfloor$
4: **if** blockindex $\%2 = 0$ **then**
5: $\boldsymbol{\theta}_{it} \sim p(\boldsymbol{\theta}_{it})$ for $i = 1, \ldots, n_s$
6: **else**
7: $t' \leftarrow t - \eta$
8: **for** $i \in \{1, \ldots, n_s\}$ **do**
9: $\boldsymbol{\theta}_{it} = k_{it}\boldsymbol{\theta}_{it'}$ with $k_{it} > 0$
10: $\Delta\bar{y}_i \leftarrow (\bar{y}_{it} \mid z_t = i) - (\bar{y}_{it'} \mid z_{t'} = i)$
11: $\Delta\bar{\mathbf{x}}_i \leftarrow (\bar{\mathbf{x}} \mid z_t = i) - (\bar{\mathbf{x}} \mid z_{t'} = i)$
12: $D_i \leftarrow D_i \cup \{\Delta\bar{y}_i, \Delta\bar{\mathbf{x}}_i\}$
13: **end for**
14: **end if**
15: **end for**
16: **for** $i \in \{1, \ldots, n_s\}$ **do**
17: $\hat{\boldsymbol{\theta}}_i^* \leftarrow$ Regress $\Delta\bar{Y}_i$ onto $\Delta\bar{\mathbf{X}}_i$ with OLS and the data set D_i
18: **end for**

4 Experiments

We complement our theoretical results with simulation studies. Starting with the single DM setting, our experiments first compare the optimal decision parameters $\boldsymbol{\theta}^{\text{AO}}$ and the causal parameter $\boldsymbol{\theta}^*$ in terms of utility maximisation, and then we demonstrate that our algorithms estimate $\boldsymbol{\theta}^*$ consistently. We then generalise the experiments to multiple DMs. Further experimental details are included in Appendix F. The code to reproduce our experiments is publicly available.[2]

Experimental setup. Following Harris et al. (2022), we generate a synthetic college admission dataset. In particular, covariates $X_t = (X_t^{\text{SAT}}, X_t^{\text{HS GPA}})^\top$ represent SAT score and high school GPA of the student arriving at round t, while Y_{it} represents the college GPA after enrolling in college i. A confounding factor is simulated to indicate the private type of a student's background: *disadvantaged* and *advantaged*. The distribution of the *disadvantaged* students' baseline B_t has a lower mean than that of their *advantaged* counterparts and the same applies for the distribution of noise O_{it}. After $\mathbf{b}_t$ is simulated and all colleges publicise their parameters $\{\boldsymbol{\theta}_{it}\}_{i=1}^n$, we compute $\mathbf{x}_t = \mathbf{b}_t + \boldsymbol{\mathcal{E}}\mathbf{a}_t$ and $\hat{y}_{it} = \mathbf{x}_t^\top \boldsymbol{\theta}_{it}$ for $i \in [n]$. Unlike Harris et al. (2022), each DM i now assigns an admission status $w_{it} \in \{0, 1\}$ to the student at round t using a variant of the ranking selection rule. Precisely, the student is admitted into college i if their prediction $\hat{y}_{it}$ lies within the top ρ-percentile of all applicants where $\rho \in [0, 1]$ and we set $\rho = 0.5$. Further discussion of this variant of ranking selection is included in Appendix F.1. As ranking selection (Definition 1) requires access to the distribution $p(X_t^\top \boldsymbol{\theta}_{it})$, we estimate it by simulating 1000 students in each round.[3] Afterwards, the compliance $z_t \in [n]$ is computed to indicate the college in which this student enrols, based on the admission statuses $\{w_{it}\}_{i=1}^n$. Finally, for students enrolled in college i at round t, i.e., $z_t = i$, we compute the target college GPA $y_{it} = \mathbf{x}_t^\top \boldsymbol{\theta}_i^* + o_{it}$. The true causal parameters $\boldsymbol{\theta}_i^* = (\theta_i^{*,\text{SAT}}, \theta_i^{*,\text{HS GPA}})^\top$ are distributed as normal distribution around $\boldsymbol{\theta}_i^* = (0, 0.5)^\top$, which was inferred from a real world dataset by Harris et al. (2022).

	$\hat{\boldsymbol{\theta}}^{\text{AO}}$	$\hat{\boldsymbol{\theta}}_{\text{OLS}}$	$\hat{\boldsymbol{\theta}}^*$
$\mathcal{Q}(\boldsymbol{\theta}_t)$	$\mathbf{2.530}$ ± 0.006	2.511 ± 0.006	2.511 ± 0.006

Table 1: [Higher is better] Utility $\mathcal{Q}(\boldsymbol{\theta}_t)$ ($\pm$ standard error) of a DM for various values of the parameter $\boldsymbol{\theta}_t$.

Additional details for MSLR. Because there are infinitely many ways to carry out the releases of $\boldsymbol{\theta}_t$ as required by Theorem 3.4 and there are infinitely many ways for multiple DMs to synchronise their releases of $\boldsymbol{\theta}_{it}$ as required by Definition 2, we provide only an instantiation of the MSLR procedure via Algorithm 1 that we use in our experiments. We use the word *coalition* to refer to the subset of n_s DMs who perform this algorithm together. Line 9 refers to the cooperative protocol (Definition 2) and line 10 to 12 refer to the extended theorem on local exogenity (Theorem 3.8). The branching in line 4 (and in line 6) checks whether the current round t is of the type $t_\bullet$ or $t_\diamond$ which we discuss next. Recall that according to Definition 2, DMs i and j are cooperative if they deploy linearly dependent parameter vectors $\{\boldsymbol{\theta}_{it_\bullet}, \boldsymbol{\theta}_{it_\diamond}\}$ in the same pair of two arbitrary rounds $t_\bullet$ and $t_\diamond$. To easily simulate the cooperative and non-cooperative aspects of DMs in our experiments, we control the interval for deploying dependent vectors with the integer constants $\eta_i \in \mathbb{N}^+$ where $t_\bullet + \eta_i = t_\diamond$. Each batch of such linearly dependent vectors gives us a linear equation as shown in Theorem 3.8 and we want to have multiple distinct batches with sufficient span in $\mathbb{R}^m$ so that $\boldsymbol{\theta}_i^*$ is solvable. Because η_i creates a gap between $t_\bullet$ and $t_\diamond$ of the same batch, distinct batches are generated in an interleaved manner using the following formula:

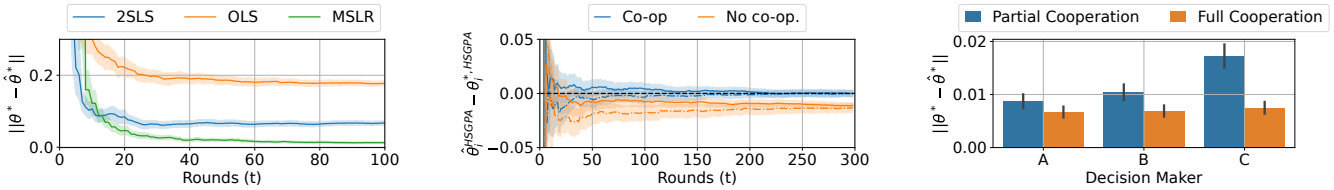
$$t_\bullet = k + \left(\left\lfloor \frac{k-1}{\eta_i} \right\rfloor \times \eta_i \right), \quad t_\diamond = t_\bullet + \eta_i,$$

where $k \in \{1, 2, \ldots\}$ denotes the k-th batch to which $\{\boldsymbol{\theta}_{it_\bullet}, \boldsymbol{\theta}_{it_\diamond}\}$ belong. Finally, we say that a set of DMs deploys the parameter vectors *synchronously* if they deploy the linearly dependent vector at the same frequency (i.e., $\forall i, \eta_i = \eta$ for some constant η), otherwise, we say their deployments are *asynchronous*.

Impact of selection procedure ($n = 1$). We first demonstrate our estimated $\boldsymbol{\theta}^{\text{AO}}$ in fact results in higher utility than other plausible selection parameters such as $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\theta}}^{\text{OLS}}$, echoing the theoretical analysis from Theorem 3.1. We regress $\mathcal{Q}(\boldsymbol{\theta}_t)$ onto $\boldsymbol{\theta}_t$ to estimate $\boldsymbol{\theta}^{\text{AO}}$ (see Section 3.1), and utilise our MSLR algorithm to estimate $\boldsymbol{\theta}^*$, whereas

[2] https://github.com/muandet-lab/csl-with-selection

[3] Having multiple students per round is equivalent to having each student arrive at different rounds subject to the same $\boldsymbol{\theta}^{\text{all}}$.



(a) Estimation errors of causal parameter estimate $\hat{\boldsymbol{\theta}}^*$ as a function of rounds.

(b) Bias in the estimated causal effect of $X_t^{\text{HS GPA}}$ on Y_{it} for 2 DMs, each depicted by a different line style.

(c) Estimation error of each DM when two of them cooperate ($A+B$) and when all of them cooperate ($A+B+C$).

Figure 3: Estimation of the causal parameters $\boldsymbol{\theta}^*$ under various scenarios. Error bars show 95% confidence interval.

	$\hat{\boldsymbol{\theta}}_1^{\text{AO}}$	$\hat{\boldsymbol{\theta}}_1^{\text{OLS}}$	$\hat{\boldsymbol{\theta}}_1^*$
$\hat{\boldsymbol{\theta}}_2^{\text{AO}}$	**2.529** $\pm$0.028	2.507 $\pm$0.029	2.506 $\pm$0.029
$\hat{\boldsymbol{\theta}}_2^{\text{OLS}}$	**2.561** $\pm$0.028	2.546 $\pm$0.029	2.545 $\pm$0.029
$\hat{\boldsymbol{\theta}}_2^*$	**2.560** $\pm$0.029	2.546 $\pm$0.029	2.544 $\pm$0.029

Table 2: [Higher is better] Utilities $\mathcal{Q}_1(\boldsymbol{\theta}_{1t}, \boldsymbol{\theta}_{2t})$ ($\pm$ standard error) of the first DM for various values of $\{\boldsymbol{\theta}_{1t}, \boldsymbol{\theta}_{2t}\}$.

$\hat{\boldsymbol{\theta}}^{\text{OLS}}$ is obtained from performing ordinary regression with $Y_t|W_t = 1$ and $X_t|W_t = 1$. Conforming to Assumption (S2), we use $\|\hat{\boldsymbol{\theta}}^*\|_2$ as the threshold and scale $\hat{\boldsymbol{\theta}}^{\text{AO}}$, such that $\|\hat{\boldsymbol{\theta}}^{\text{AO}}\|_2 = \|\hat{\boldsymbol{\theta}}^*\|_2$ to ensure a fair comparison. On the other hand, if $\hat{\boldsymbol{\theta}}^{\text{OLS}}$ has a larger magnitude than the threshold, we scale it down accordingly (see Appendix F.4 for the detailed explanation). Table 1 reports their utility values $\mathcal{Q}(\boldsymbol{\theta}_t)$. We can see that $\hat{\boldsymbol{\theta}}^{\text{AO}}$ induces the highest utility compared to other plausible options of $\boldsymbol{\theta}_t$. To demonstrate the impact of selection on estimating $\boldsymbol{\theta}^*$, which is needed for the DM to comply with the regulation (Corollary 3.2), we compare the MSLR algorithm (cf. Theorem 3.4) with that of Harris et al. (2022), i.e., 2SLS. Figure 3a shows estimation errors as the number of rounds increases. Unlike the OLS and 2SLS estimates, our estimate of $\boldsymbol{\theta}^*$ is asymptotically unbiased.

Impact of competitive selection ($n \geq 2$)**.** Next, we show that $\hat{\boldsymbol{\theta}}_1^{\text{AO}}$ induces the optimal utility $\mathcal{Q}_1(\boldsymbol{\theta}_{1t}, \boldsymbol{\theta}_{2t})$ for the first DM as a dominating strategy. Analogous to the previous experiment, Table 2 shows that our estimate $\hat{\boldsymbol{\theta}}_1^{\text{AO}}$ induces the highest utility $\mathcal{Q}_1$ for the first DM, regardless of $\boldsymbol{\theta}_{2t}$ deployed by the second DM. We normalise the parameter similarly as before and use $\|\hat{\boldsymbol{\theta}}_1^*\|_2$ and $\|\hat{\boldsymbol{\theta}}_2^*\|_2$ as thresholds.

We now demonstrate the impact of competitive selection on the estimation of causal parameter $\boldsymbol{\theta}_i^*$ for $i \in [n]$, which are needed for DMs to comply with our regulations. By Theorem 3.8, they must follow the cooperative protocol (Definition 2) by deploying linearly dependent parameter vectors $\boldsymbol{\theta}_{it} = k_i \boldsymbol{\theta}_{it'}$ in the same pair of rounds t and t'. To this end, we test whether DMs can estimate $\boldsymbol{\theta}^*$ when they deploy linearly dependent vectors (a) synchronously, as required by the protocol (i.e., cooperation), and (b) asynchronously (i.e., no cooperation). Figure 3b shows that cooperation enables all DMs to obtain unbiased estimates of $\theta^{*,\text{HS GPA}}$, the ground-truth causal effect of the high-school GPA covariate. We provide the results for the other covariate in Appendix F.3. Lastly, we demonstrate that following the cooperative protocol is of mutual benefit to all DMs for obtaining accurate estimates of $\boldsymbol{\theta}_i^*$. To this end, we generate the data with $n = 3$, for two scenarios: a group of two DMs (A and B) deploys linearly dependent parameter vectors synchronously, while the remaining DM (C) deploys its respective linearly dependent vector (i) asynchronously (i.e., leading to partial cooperation between DMs), and (ii) synchronously (i.e., full cooperation). We use the converged estimates (i.e., after $T = 100$ rounds) of causal parameters under both scenarios to demonstrate that, in terms of accuracy, not only does the DM C gain substantially by joining the coalition, but it also benefits the current members of the coalition; see Figure 3c.

5 Conclusion

To conclude, we study the problem of causal strategic learning under competitive selection by multiple decision makers. We show that in this setting, optimal selection rules require a trade-off between choosing the best agents and motivating their improvement. In addition, these rules may unjustly reduce the admission chances of agents due to reliance on non-causal predictions. To address these issues, we propose conditions for a benevolent regulator to impose on decision makers, allowing them to recover true causal parameters from observational data and ensure optimal incentives for agents' improvement without excessively reducing their admission chances, thus safeguarding agents' welfare.

Our results rest on assumptions like homogeneous strategic behavior and linearity in agent models. Although these assumptions undoubtedly limit the applicability of our methods, they do not undermine the implication of our work. Intuitively, this inherent trade-off emerges because a DM has only one degree of freedom in designing the selection rule that may result in two distinct effects. Consequently, selecting the best candidates (private reward) and incentivising their improvements (social return) can indeed differ; and when they do, the benevolent regulator, e.g., governments, is needed to align the two. Our finding reinforces causal identification as an essential instrument to achieve this. Future studies could delve into non-linear agent models, fully heterogeneous setting, or scenarios in which certain decision makers cooperate strategically.

Acknowledgments

We thank Jake Fawkes and Nathan Kallus for a fruitful discussion and detailed feedback. We also thank David Kaltenpoth, Jilles Vreeken, and Xiao Zhang for insightful questions and feedback on the preliminary version of this work which was presented at the CISPA ML Day.

References

Alon, T.; Dobson, M.; Procaccia, A.; Talgam-Cohen, I.; and Tucker-Foltz, J. 2020. Multiagent evaluation mechanisms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 1774–1781.

Angrist, J. D.; Imbens, G. W.; and Rubin, D. B. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434): 444–455.

Bechavod, Y.; Ligett, K.; Wu, S.; and Ziani, J. 2021. Gaming Helps! Learning from Strategic Interactions in Natural Dynamics. In Banerjee, A.; and Fukumizu, K., eds., *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, 1234–1242. PMLR.

Bechavod, Y.; Podimata, C.; Wu, S.; and Ziani, J. 2022. Information discrepancy in strategic learning. In *International Conference on Machine Learning*, 1691–1715. PMLR.

Björkegren, D.; and Grissen, D. 2020. Behavior revealed in mobile phone usage predicts credit repayment. *The World Bank Economic Review*, 34(3): 618–634.

Cameron, A. C.; and Trivedi, P. K. 2005. *Microeconometrics: methods and applications*. Cambridge university press.

Chau, S. L.; Ton, J.-F.; González, J.; Teh, Y.; and Sejdinovic, D. 2021. Bayesimp: Uncertainty quantification for causal data fusion. *Advances in Neural Information Processing Systems*, 34: 3466–3477.

Dee, T. S.; Dobbie, W.; Jacob, B. A.; and Rockoff, J. 2019. The causes and consequences of test score manipulation: Evidence from the New York regents examinations. *American Economic Journal: Applied Economics*, 11(3): 382–423.

Deshpande, K. V.; Pan, S.; and Foulds, J. R. 2020. Mitigating Demographic Bias in AI-Based Resume Filtering. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 268–275. Association for Computing Machinery.

Dranove, D.; Kessler, D.; McClellan, M.; and Satterthwaite, M. 2003. Is more information better? The effects of "report cards" on health care providers. *Journal of political Economy*, 111(3): 555–588.

Ghassemi, M.; and Mohamed, S. 2022. Machine learning and health need better values. *npj Digital Medicine*, 5(1): 51.

Hardt, M.; Megiddo, N.; Papadimitriou, C.; and Wootters, M. 2016. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, 111–122.

Harris, K.; Ngo, D. D. T.; Stapleton, L.; Heidari, H.; and Wu, S. 2022. Strategic instrumental variable regression: Recovering causal relationships from strategic responses. In *International Conference on Machine Learning*, 8502–8522. PMLR.

Hartford, J.; Lewis, G.; Leyton-Brown, K.; and Taddy, M. 2017. Deep IV: A Flexible Approach for Counterfactual Prediction. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, 1414–1423. PMLR.

Kleinberg, J.; Lakkaraju, H.; Leskovec, J.; Ludwig, J.; and Mullainathan, S. 2018. Human decisions and machine predictions. *The quarterly journal of economics*, 133(1): 237–293.

Miller, J.; Milli, S.; and Hardt, M. 2020. Strategic classification is causal modeling in disguise. In *International Conference on Machine Learning*, 6917–6926. PMLR.

Muandet, K.; Mehrjou, A.; Lee, S. K.; and Raj, A. 2020. Dual instrumental variable regression. *Advances in Neural Information Processing Systems*, 33: 2710–2721.

Munro, E. 2022. Learning to personalize treatments when agents are strategic. *arXiv preprint arXiv:2011.06528*.

Newey, W. K.; and Powell, J. L. 2003. Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, 71(5): 1565–1578.

Perdomo, J.; Zrnic, T.; Mendler-Dünner, C.; and Hardt, M. 2020. Performative prediction. In *International Conference on Machine Learning*, 7599–7609. PMLR.

Peters, J.; Janzing, D.; and Schölkopf, B. 2017. *Elements of causal inference: foundations and learning algorithms*. The MIT Press.

Shavit, Y.; Edelman, B.; and Axelrod, B. 2020. Causal strategic linear regression. In *International Conference on Machine Learning*, 8676–8686. PMLR.

Singh, R.; Sahani, M.; and Gretton, A. 2019. Kernel instrumental variable regression. *Advances in Neural Information Processing Systems*, 32.

Wiens, J.; Saria, S.; Sendak, M.; Ghassemi, M.; Liu, V. X.; Doshi-Velez, F.; Jung, K.; Heller, K.; Kale, D.; Saeed, M.; Ossorio, P. N.; Thadaney-Israni, S.; and Goldenberg, A. 2019. Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine*, 25(9): 1337–1340.