Algorithmic Decision-Making under Agents with Persistent Improvement

Tian Xie[1], **Xuwei Tan**[1], **Xueru Zhang**[1]

[1]Department of Computer Science and Engineering, the Ohio State University
{xie.1379, tan.1206, zhang.12807}@osu.edu

Abstract

This paper studies algorithmic decision-making under human strategic behavior, where a decision maker uses an algorithm to make decisions about human agents, and the latter with information about the algorithm may exert effort strategically and improve to receive favorable decisions. Unlike prior works that assume agents benefit from their efforts immediately, we consider realistic scenarios where the impacts of these efforts are persistent and agents benefit from efforts by making improvements gradually. We first develop a dynamic model to characterize persistent improvements and based on this construct a Stackelberg game to model the interplay between agents and the decision-maker. We analytically characterize the equilibrium strategies and identify conditions under which agents have incentives to invest efforts to improve their qualifications. With the dynamics, we then study how the decision-maker can design an optimal policy to incentivize the largest improvements inside the agent population. We also extend the model to settings where 1) agents may be dishonest and game the algorithm into making favorable but erroneous decisions; 2) honest efforts are forgettable and not sufficient to guarantee persistent improvements. With the extended models, we further examine conditions under which agents prefer honest efforts over dishonest behavior and the impacts of forgettable efforts.

1 Introduction

In applications such as lending, college admission, hiring, recommendation systems, etc., machine learning (ML) algorithms have been increasingly used to evaluate and make decisions about human agents. Given information about an algorithm, agents subject to ML decisions may behave strategically to receive favorable decisions. How to characterize the strategic interplay between algorithmic decisions and agents, and analyze the impacts they each have on the other, are of great importance but challenging

This paper studies algorithmic decision-making under strategic agent behavior. Specifically, we consider a decision-maker who assesses a group of agents and aims to accept those that are *qualified* for certain tasks based on assessment outcomes. With knowledge of the acceptance rule, agents may behave strategically to increase their chances of getting accepted. For example, agents may invest to genuinely *improve* their qualifications (i.e., honest effort), or they may *manipulate* the observable assessment outcomes to game the algorithm (i.e., dishonest effort). Both types of behaviors have been studied. In particular, Hardt et al. (2016a); Dong et al. (2018); Braverman and Garg (2020); Jagadeesan, Mendler-Dünner, and Hardt (2021); Sundaram et al. (2021); Zhang et al. (2022); Eilat et al. (2022) focus on learning under strategic manipulation, where they proposed various analytical frameworks (e.g., Stackelberg games) to model manipulative behavior, and analyzed models or developed learning algorithms that are robust against manipulation.

Another line of research (Zhang et al. 2020; Harris, Heidari, and Wu 2021; Bechavod et al. 2022; Kleinberg and Raghavan 2020; Chen, Wang, and Liu 2020; Barsotti, Kocer, and Santos 2022; Jin et al. 2022) considers a different setting where agent qualifications (labels) change in accordance with the improvement actions. The goal of the decision-maker is to design a mechanism such that agents are incentivized to behave toward directions that improve the underlying qualifications. Notably, Kleinberg and Raghavan (2020) proposed a mechanism to incentivize individuals to invest in specific improvable features. Their work inherited the classical settings of the Principal-agent model in economics but designed an incentivizing mechanism under a linear classifier. They modeled manipulation and improvement similarly (linear in efforts) and did not consider the persistent and delayed effects of improvement. The mixture of both improvement and manipulation behavior is also studied (Miller, Milli, and Hardt 2020; Chen, Wang, and Liu 2020; Barsotti, Kocer, and Santos 2022; Horowitz and Rosenfeld 2023; Jin et al. 2022). However, these works regarded improvement as a similar action to manipulation where the only difference is it will incur a label change. Another related topic is *performative prediction* (Perdomo et al. 2020; Izzo, Ying, and Zou 2021; Hardt, Jagadeesan, and Mendler-Dünner 2022; Jin et al. 2024b), an abstraction that captures agent actions via model-induced distribution shifts. Details and more related works are presented in Sec. 2.

This paper primarily focuses on honest agents (i.e., agents will invest effort to improve their qualifications), while settings with both improvement and manipulation are also studied. We first propose a novel two-stage Stackelberg game to

model the interactions between decision-maker and agents, i.e., the decision-maker commits to its policy, following which agents best respond. A crucial difference between this study and the prior works is that the existing models all assume that the results of agents' improvement actions are *immediate*, i.e., once agents decide to improve, they experience *sudden* changes in qualifications and receive the return *at once*. However, we observe that in many real-world applications, the impacts of improvement action are indeed *persistent* and *delayed*. For example, humans improve their abilities by acquiring new knowledge, but they make progress gradually and benefit from such behavior throughout their lifetime; loan applicants improve their credit behaviors by repaying all the debt in time, but there is a time lag between such behaviors and the increase in their credit scores. Therefore, it is critical to capture these delayed outcomes in the Stackelberg game formulation.

To this end, we propose a *qualification dynamic* model to characterize how agent qualifications would gradually improve upon exerting honest efforts. Such dynamics further indicate the time it takes for agents to reach the targeted qualifications that are just enough for them to be accepted. The impacts of such time lag on agents are then captured by a *discounted utility model*, i.e., reward an agent receives from the acceptance diminishes as time lag increases. Under this discounted utility model, agents best respond by determining how much effort to exert that maximizes their discounted utilities.

This paper aims to analytically and empirically study the proposed model. With the understanding of the strategic interactions between the decision-maker and agents, we further study how the decision-maker can design an optimal policy to incentivize the largest improvements inside the agent population, and empirically verify the benefits of the optimal policy.

Additionally, we extend the model to more complex settings where (i) agents have an additional option of strategic manipulation and can exert dishonest effort to game the algorithm; (ii) honest efforts exerted by agents are forgettable and may not be sufficient to guarantee persistent improvements, instead the qualifications may deteriorate back to the initial states. We will propose a *model with both manipulation & improvement* and a *forgetting mechanism* to study these settings, respectively. We aim to examine how agents would behave when they have both options of manipulation and improvement, under what conditions they prefer improvement over manipulation, and how the forgetting mechanism affects an agent's behavior and long-term qualifications.

Our contributions can be summarized as follows:

1. We formulate a new Stackelberg game to model the interactions between decision-maker and strategic agents. To the best of our knowledge, this is the first work capturing the delayed and persistent impacts of agents' improvement behavior (Sec. 3).
2. We study the impacts of acceptance policy and the external environment on agents, and identify conditions under which agents have incentives to exert honest efforts. This provides guidance on designing incentive mechanisms to encourage agents to improve (Sec. 4).
3. We characterize the optimal policy for the decision-maker that incentivizes the agents to improve (Sec. 5).
4. We consider the possibility of dishonest behavior and propose a *model with both improvement and manipulation*; we identify conditions when agents prefer one behavior over the other (Sec. 6).
5. We propose a *forgetting mechanism* to examine what happens when honest efforts are not sufficient to guarantee persistent improvement (Sec. 7).
6. We conduct experiments on real-world data to evaluate the analytical model and results (Sec. 8).

2 Related Work

2.1 Strategic Manipulation

Though our work primarily lies in proposing a new model for improvement behaviors, the problem settings are also closely related to strategic classification problems (Hardt et al. 2016a; Ben-Porat and Tennenholtz 2017; Dong et al. 2018; Braverman and Garg 2020; Sundaram et al. 2021; Jagadeesan, Mendler-Dünner, and Hardt 2021; Ahmadi et al. 2021; Eilat et al. 2022; Horowitz and Rosenfeld 2023; Miehling et al. 2019). Hardt et al. (2016a) formulated classification problems with strategic manipulation as a Stackelberg game with deterministic cost functions, where the decision maker optimizes classification accuracy based on individuals' best responses. Afterwards, more sophisticated analytical frameworks were proposed (Dong et al. 2018; Braverman and Garg 2020; Jagadeesan, Mendler-Dünner, and Hardt 2021). Dong et al. (2018) proposed an online algorithm for strategic classification. Braverman and Garg (2020) added randomness to strategic classifiers, while Shao, Blum, and Montasser (2024); Lechner, Urner, and Ben-David (2023) provided a complete analysis of the regret bound for online strategic classification. On the other hand, Sundaram et al. (2021) analyzes the statistical learnability of strategic classification with an SVC classifier. Jagadeesan, Mendler-Dünner, and Hardt (2021) relaxed the *standard microfoundations* assumption where individuals are perfectly rational to *alternative microfoundations* where a proportion of individuals may not be strategic, and proposed a *noisy response model* to tackle the new problem. Zhang et al. (2022) studied the setting where the decision maker and individuals only have knowledge of the feature distributions as random variables. Thus, the strategic manipulation corresponds to a distribution shift and its cost is also a random variable. Eilat et al. (2022) considered the setting where individual responses are dependent and the classifier is learned through *graph neural networks*. Xie and Zhang (2024a) studied the long-term impacts on welfare and fairness under a sequential strategic learning setting.

2.2 Improvement

However, there are other literature considering improvement behavior (Liu et al. 2019; Rosenfeld et al. 2020; Shavit, Edelman, and Axelrod 2020; Alon et al. 2020; Zhang et al. 2020; Chen, Wang, and Liu 2020; Kleinberg and Raghavan 2020; Bechavod et al. 2021; Ahmadi et al. 2022a,b; Raab

and Liu 2021; Xie et al. 2024; Xie and Zhang 2024b; Zhang, Khalili, and Liu 2020). Unlike strategic manipulation, improvement will incur a label change. Liu et al. (2019) studied the conditions where fairness interventions can promote improvement among individuals. Zhang et al. (2020) formulated the label change as a transition matrix where the transition probabilities are deterministic and difficult to estimate. Other works consider both behaviors at the same time. Xie et al. (2024) studied randomness when the agents can both improve and manipulate, while Xie and Zhang (2024b) relaxed the linear assumption on the decision policy and studied the welfare under strategic learning settings. Kleinberg and Raghavan (2020) proposed a mechanism to incentivize individuals to invest in specific features where the individuals have a budget to invest strategically on all features including undesired ones. Their work inherited the classical settings of the Principal-agent model in economics but designed an incentivizing mechanism under a linear machine learning classifier. They modeled manipulation and improvement similarly (linear in efforts) and did not consider the persistent and delayed effects of improvement. By contrast, we first develop a fundamentally different dynamic model to characterize persistent and delayed improvements. Based on the new model, we construct a Stackelberg game to model the interplay between agents and the decision-maker.

Another line of literature focuses on the underlying causal models of strategic classification. Shavit, Edelman, and Axelrod (2020) and Alon et al. (2020) introduced causal inference frameworks into strategic behaviors including manipulation and improvement. Chen, Wang, and Liu (2020) divided the features into immutable features, improvable features and manipulable features and explored linear classifiers which can prevent manipulation and encourage improvement. Jin et al. (2022) also focused on incentivizing improvement and proposed a subsidy mechanism to induce improvement actions and improve social well-being metrics. Barsotti, Kocer, and Santos (2022) conducted several experiments where both improvement and manipulation are possible and both actions incur linear deterministic costs.

2.3 Other Agent Dynamics

In addition to manipulation and improvement, some studies consider different individual behaviors and models. For example, Zhang et al. (2019); Dean et al. (2024a); Hashimoto et al. (2018) focused on participation dynamics where agents at each time decide their participation in decision systems. Perdomo et al. (2020); Hardt, Jagadeesan, and Mendler-Dünner (2022); Jin et al. (2024a) studied *performative prediction* which is an abstract framework for optimization when the deployed model influences agent population. Yin et al. (2024) studied the setting where the population distribution has an unknown dynamic and reinforcement learning is used to improve long-term fairness. We refer interested readers to survey papers (Zhang and Liu 2021; Dean et al. 2024b) for more examples. Also, our work is related to preference shifts and opinion dynamics in recommendation systems, which we refer to Castellano, Fortunato, and Loreto (2009) as a comprehensive survey. Among the rich set of works, Dean and Morgenstern (2022); Gaitonde, Kleinberg, and Tardos (2021) proposed geometric models for opinion polarization and motivate our work.

3 Problem Formulation

Consider an agent population with m skill sets. Each agent has a *qualification profile* at time t, denoted as a unit m-dimensional vector $q_t \in [0,1]^m$ with $||q_t||_2 = 1$, A decision-maker at each time makes decisions $D_t \in \{0,1\}$ ("0" being reject and "1" accept) about the agents based on their qualification profiles. Let fixed vector $d \in [0,1]^m$ be the ideal qualification profile that the decision maker desires.

Decision-maker's policy. For an agent with qualification profile q_t, the decision-maker assesses whether the agent's profile lines up with the desired qualifications d, and makes decision D_t based on their similarity $x_t := q_t^T d$ using a *fixed* threshold policy $\pi(x_t) = \mathbf{1}(x_t \geq \theta)$, i.e., only agents that are sufficiently fit can get accepted. How to choose threshold θ is discussed in Sec. 5. We assume only agents with initial similarity $x_0 \geq 0$ are interested in positions and only focus on these candidates.

Although the decision policy introduced above focuses on the similarity between q_t and d where qualifications q_t are normalized with the same magnitude for all agents, it can be easily extended to settings where the magnitude/strength of skills also matters and may differ across agents. For example, if the decision-maker prefers students with balanced math and English skills and there are two "balanced" students, the decision-maker is likely to prefer the one with higher scores. Therefore, we propose a **pre-normalization procedure** to take magnitude into account. The idea is to first add an additional dimension to initial qualification profile q_0, which represents the agent's unobservable "irrelevant attribute" (all other skills an agent has that are not important for the decision). Thus, we expand the original q_0 to obtain a $m+1$ dimensional complete qualification profile. Meanwhile, we add a dimension to the ideal qualification profile d with 0 as its value; the new ideal profile becomes $[d; 0]$. Then we can make the following natural assumption:

Assumption 3.1. After adding the dimension of "irrelevant attribute", for all agents, the norms of their complete qualification profiles are the same.

Assumption 3.1 has been supported by literature in machine learning (Liu, Garg, and Borgs 2022) and social science (Holmstrom and Milgrom 1991). The "irrelevant" dimension demonstrates all other skills that belong to an agent but are not important to the decision. Therefore, competency in relevant/measurable attributes implies weakness in irrelevant/immeasurable attributes and the length of the complete qualification profile stays the same for all agents. With Assumption 3.1 and the distribution of q_0 as Q, we formalize the **pre-normalization** procedure in Algorithm 1.

Agent qualification dynamics. We assume agents have information about the ideal profile d (e.g., from application guides, mock interviews). In the beginning, agents with q_0 can choose to improve their profiles by investing an effort $k \in [0,1]$ to acquire the relevant knowledge, but the effort will have delayed and persistent effects over subsequent time

Algorithm 1: Pre-normalization procedure

Require: Joint distribution Q for q_0, n agents with $\{q_0^i\}_{i=1}^{i=n}$ where $q_0^i \in [0,1]^m$, $d \in [0,1]^m$.

Ensure: Normalized $\{q_0^i\}_{i=1}^{i=n}$ (i.e., $q_0^i \in [0,1]^m$ and $\|q_0^i\| = 1$), new $d \in [0,1]^{m+1}$.

1: $d = [d, 0]$.
2: According to Q, find the largest norm $K = \max_{q_0 \sim Q} \|q_0^i\|$ of original profiles.
3: **for** $i \in \{1, \dots n\}$ **do**
4: Calculate norm difference $z^i = \sqrt{K^2 - \|q_0^i\|_2^2}$.
5: $q_0^i = \frac{[q_0^i; z^i]^T}{K} \in [0,1]^{m+1}$.
6: **end for**

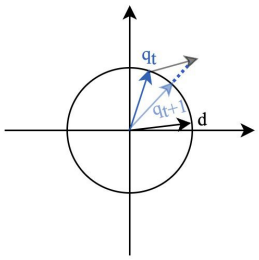


Figure 1: Dynamics of agent's qualification q_t.

stages. The specific value of k depends on the agent's utility and will be introduced at the end of this section. Upon investing k, the agent's qualifications q_t gradually improve over time based on the following:

$$\widetilde{q}_{t+1} = q_t + k \cdot q_t^T d \cdot d\,; \qquad q_{t+1} = \frac{\widetilde{q}_{t+1}}{\|\widetilde{q}_{t+1}\|_2}. \tag{1}$$

(1) suggests that agents at each time improve toward the ideal profile d. How much they can improve depend on their current profile q_t and the effort k. The similarity $q_t^T d$ in the dynamics captures the reinforcing effects: agents that are more qualified could have more resources and are more capable of leveraging the acquired knowledge to improve their skills. Note that the maximum improvement an agent attains at each round is bounded, i.e., the normalized vector q_{t+1} after improvement is always between current qualifications q_t and the ideal profile d. Fig. 1 illustrates the improvement dynamics of qualification q_t in a two-dimensional space.

Dynamics in (1) model the delayed and persistent impacts of improvement action (i.e., effort k). In many real applications, humans acquire knowledge and benefit from repeated practices. They make progress toward the goal gradually, and it takes time to receive the desired outcome from the investment. Indeed, (1) is inspired by the dynamics in Dean and Morgenstern (2022), which was used for modeling preference shifts (details are in Sec. 2.3), where individuals update their opinions/preferences based on their correlations with some influencer (e.g., a political figure) and control the power of the intervention. We believe this is similar to improvement especially when agents improve themselves by imitating some "role models". For example, in a job application scenario, the "influencer" is a current worker who holds information session and introduces her profile d. Then the agents strive to mimic the profile of d by updating q_t. The imitating nature of improvement is well justified in many works (e.g., Raab and Liu (2021); Zhang et al. (2022)), making agent improvement suitable to be modeled in a similar fashion to concept drift/preference shift. Thus, we use (1) to model the evolution of agents' (pre-normalized) qualifications. Based on Prop. 1 of Dean and Morgenstern (2022), we know that q_t converges under dynamics, as formally stated in Lemma 3.2 below.

Lemma 3.2 (Convergence of qualification)**.** *Consider an agent with initial similarity $x_0 := q_0^T d > 0$. If he/she makes an effort k and improves qualification profile q_t based on dynamics in* (1)*, then q_t converges to the desired profile d. The evolution of the similarity $x_t := q_t^T d$ is given by:*

$$x_t^{-2} - 1 = \frac{(x_0)^{-2} - 1}{(k+1)^{2t}} \tag{2}$$

Lemma 3.2 suggests that any agent eventually becomes an ideal candidate with a perfectly aligned profile (i.e., $x_t = q_t^T d = 1$), as long as he/she is interested in the position ($x_0 \geq 0$) and willing to make an effort ($k > 0$). The only difference among agents is the speed of convergence: it takes less time for agents who are more qualified at the beginning (i.e., larger x_0) and/or make more effort (i.e., larger k) to become ideal and get accepted. Note that our work focuses on agent's improvement behavior with *persistent* and *delayed* effects. Although the model is presented in a simplified setting where only a one-step effort is made by the agents at the beginning, it can capture more complicated scenarios where agents repeatedly exert efforts multiple times until they reach the target. Each effort has persistent effects on improving the qualification. Specifically, suppose each agent at time t can exert an effort k_t and the agent's qualification improves based on $\widetilde{q}_{t+1} = q_t + \sum_{\tau=0}^{t} k_\tau \cdot q_t^T d \cdot d$ with $q_{t+1} = \frac{\widetilde{q}_{t+1}}{\|\widetilde{q}_{t+1}\|_2}$ (i.e., every time the agent improves from all accumulated efforts $\sum_{\tau=0}^{t} k_\tau \in [0,1]$ he/she invested so far). For this new dynamics, the overall impacts of these efforts $\{k_t\}_{t\geq 0}$ on improving agent qualification can be *equivalently* characterized by the dynamics (1) with some one-step effort. That is, there exists an effort $k^* \in [0,1]$ such that investing k^* once at the beginning has the same impact on $\lim_{t\to\infty} q_t$ as investing a sequence of efforts $\{k_t\}_{t\geq 0}$ over time. We provide more detailed discussion on this in App. A. We also discuss a special case where the effect of k is decreasing over time and provide further convergence analysis (Thm. A.1) in App. A.

Agent's utility & action. Because it takes time for agents to receive rewards (i.e., get accepted) for their efforts, they may not have incentives to invest if there is a long delay. In practice, people may be more attracted to investments with immediate rewards than delayed rewards, or they may simply not have enough time to wait. For example, students only have limited time to prepare for college applications; credit card applicants may not have incentives to improve

their credit scores and wait to get approval for a specific credit card when there are many instant-approval cards on the market.

To characterize the delayed rewards, we use a discount model and assume the reward each agent receives from the effort k decreases over time. Specifically, let H be the minimum time it takes for an agent to get accepted from the effort $k > 0$. We define **agent's utility** as:

$$U = \frac{1}{(1+r)^H} - k. \tag{3}$$

That is, the utility is the exponentially discounted reward an agent receives from the acceptance minus the effort. $r > 0$ is the discounting factor. Note that the discounted utility model[1] has been widely used in literature such as reinforcement learning (Kaelbling, Littman, and Moore 1996), finance (Meier and Sprenger 2013), and economics (Krahn and Gafni 1993; Samuelson 1937).

Since threshold policy $\pi(x_t) = \mathbf{1}(x_t \geq \theta)$ is used to make decisions, an agent gets accepted whenever the qualification profile is sufficiently aligned with the ideal profile, i.e., $x_t = q_t^T d \geq \theta$. Based on (2), we can derive H as a function of threshold θ, agent's initial similarity x_0, and effort k, i.e.,

$$\begin{aligned} H &= \min_t \{x_t \geq \theta\} = \min_t \left\{ \frac{(x_0)^{-2} - 1}{(k+1)^{2t}} \leq \frac{1}{\theta^2} - 1 \right\} \\ &= \frac{-\ln\left(\sqrt{\frac{(\theta)^{-2}-1}{(x_0)^{-2}-1}}\right)}{\ln(k+1)} \end{aligned} \tag{4}$$

Plug in (3), agent's utility becomes:

$$U := U(k, \theta, r, x_0) = (1+r)^{\frac{\ln\left(\sqrt{\frac{(\theta)^{-2}-1}{(x_0)^{-2}-1}}\right)}{\ln(k+1)}} - k. \tag{5}$$

Therefore, strategic agents will choose to improve their qualifications only if utility $U(k, \theta, r, x_0) > 0$, and they will choose the investment k that maximizes the utility.

Stackelberg game. We model the strategic interplay between the decision-maker and agents as a Stackelberg game, which consists of two stages: (i) the decision-maker first publishes the optimal acceptance threshold θ (details are in Sec. 5); (ii) agents after observing the threshold take actions to maximize their utilities as given in (5).

Manipulation & forgetting. The model formulated above has two implicit assumptions: (i) agents are honest and they improve their qualifications by making actual efforts; (ii) once agents make a one-time effort k to acquire the knowledge, they never forget and can repeatedly leverage this knowledge to improve their profiles based on (1). However, these assumptions may not hold. In practice, agents may fool the decision-maker by directly manipulating x_t to get accepted without improving actual q_t, e.g., people cheat on exams or interviews to get accepted. Moreover, the knowledge agents acquired at the beginning may not be sufficient to ensure repeated improvements. To capture these, we further extend the above model to two settings:

1. *Manipulation:* Besides improving the actual profile q_t by making an effort k, agents may choose to manipulate x_t directly to fool the decision-maker. The detailed model and analysis are in Sec. 6.
2. *Forgetting:* One-time investment k may not guarantee the improvements all the time, i.e., qualifications q_t do not always move toward the direction of ideal profile d, instead it may devolve and possibly go back to starting state q_0. The detailed model and analysis are in Sec. 7.

Objective. In this paper, we study the above interactions between decision-maker and agents. We aim to understand (i) under what conditions agents have incentives to improve their qualifications; (ii) how to design the optimal policy to incentivize the largest improvements inside the agent population; (iii) how the agents would behave when they have both options of manipulation and improvement, and under what conditions agents prefer improvement over manipulation; (iv) how the forgetting mechanism affects agent's behavior and long-term qualifications.

4 Improvement & Optimal Effort

In this section, we examine the impact of decision threshold θ and the environment (i.e., discounting factor r) on agent behavior. Specifically, we focus on agents with discounted utility ((5)) and identify conditions under which the agents have incentives to improve their qualifications. Note that we do not consider issues of manipulation and forgetting in this section. Based on (5), an agent with $x_0 := q_0^T d$ chooses to improve only if its utility $U(k, \theta, r, x_0) > 0$. To characterize the impact of an agent's one-time investment k on $U(k, \theta, r, x_0)$, we first define a function $C(\theta, r, x_0)$ that summarizes the impacts of all the other factors (i.e., threshold θ, discounting factor r, and initial profile similarity x_0) on agent utility, as defined below.

$$C(\theta, r, x_0) = -\ln\left(\sqrt{\frac{(\theta)^{-2}-1}{(x_0)^{-2}-1}}\right) \cdot \ln(1+r) \tag{6}$$

Based on $C(\theta, r, x_0)$, we can derive conditions under which agents have incentives to improve (Thm. 4.1).

Theorem 4.1 (Improvement & optimal effort)**.** *There exists a threshold $m > 0$ such that for any θ, r, x_0 that satisfies $C(\theta, r, x_0) < m$, the agent has the incentive to improve the qualifications, i.e., agent utility is positive for some efforts $k > 0$. Moreover, there exists a unique optimal effort $k^* \in (0, 1)$ that maximizes the agent utility.*

Thm. 4.1 identifies a condition under which agents have incentives to exert positive effort $k > 0$. This condition depends on factors θ, r, x_0 and can be fully characterized by the function $C := C(\theta, r, x_0)$.

Although the analytical solution of the threshold m is difficult to find, we can numerically solve $m \approx 0.3164$ as shown in App. F.1. In Fig. 2, we illustrate agent utilities U as

[1] Under exponential discounting function, the agent's reward diminishes at a constant rate (Grüne-Yanoff 2015). Our model can also adopt other discounting functions (e.g., hyperbolic discounting) for settings when the agent's reward decreases inconsistently. The qualitative results of this paper still remain the same.

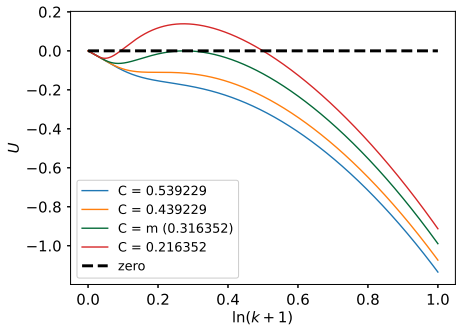


Figure 2: Impact of effort k on agent utility U under different $C := C(\theta, r, x_0)$: $\exists m > 0$ such that agents have incentives to invest and improve their qualifications if $C < m$.

functions of effort under different C. The results show that only when $C < m$ (red curve), an agent can attain positive utility with effort $k > 0$; when $C \geq m$ (green/yellow/blue curve), agents will not invest because the maximum utility is attained at $k = 0$. Moreover, when $C < m$ (red curve), there is a unique optimal effort k that maximizes the utility. These results are consistent with Thm. 4.1.

The condition in Thm. 4.1 further indicates the impacts of policy θ, discounting factor r, and initial state x_0 on agent behavior. Specifically, agents only invest if $C(\theta, x_0, r) < m$ holds. By fixing any two of θ, x_0, r, we can identify the domain of the third factor under which agents invest to improve. These results are summarized in Table 1 and verified in App. B. It shows that for any threshold θ and discounting factor r, agents only improve if their initial qualification profile is sufficiently similar to the ideal profile; the domain of θ also implies the best profile an agent with initial state x_0 can reach after exerting effort: if acceptance threshold θ is larger than the upper bound of θ given in Table 1, then agents will not have incentives to improve.

Table 1: Domain of initial similarity x_0 (or threshold θ) under which agents invest positive efforts.

Domain of x_0 (given θ, r)	$x_0 > \left(1 + (\theta^{-2} - 1) \cdot \exp\left(\frac{2m}{\ln(1+r)}\right)\right)^{-1/2}$
Domain of θ (given x_0, r)	$\theta \leq \left(1 + (x_0^{-2} - 1) \cdot \exp\left(-\frac{2m}{\ln(1+r)}\right)\right)^{-1/2}$

The above results further suggest effective strategies that encourage agents to improve their qualifications, i.e., more agents are incentivized to improve if (i) the decision-maker's acceptance threshold θ is lower; or (ii) the time it takes for agents to succeed after investments is shorter (smaller discounting factor r). Examples of both strategies in real applications are discussed in App. B, which further verify the effectiveness of our proposed model.

5 Decision-maker's policy to incentivize improvement

Sec. 4 studied the impact of threshold θ on agent behavior and provided guidance on incentivizing agents to improve. In practice, although it is more difficult to adjust the discounting factor r, the decision-maker can adjust the threshold policy θ to incentivize the largest possible amount of total improvement, thereby improving the *social welfare*. In this section, we study the optimal policy when the decision-maker is aware of the agent's best response and hopes to incentivize agents to improve.

Suppose the decision-maker has full information about agents and can anticipate their behaviors, i.e., for any decision threshold θ, it knows that agents whose initial similarity $x_0 > x^*(\theta) := \left(1 + (\theta^{-2} - 1) \cdot \exp\left(\frac{2m}{\ln(1+r)}\right)\right)^{-\frac{1}{2}}$ will invest and improve their profiles (by Table 1). Also, we define $x^*(0) = 0$ to let $x^*(\theta)$ be continuous in $[0, 1]$ and denote f as the probability density function of the agent similarity x_0 which is also continuous in $[0, 1]$. Then, we can define $U_d(\theta)$ as the utility of the decision-maker under the threshold as the total amount of agents' improvements:

$$U_d(\theta) = \int_{x^*(\theta)}^{\theta} (\theta - x_0) \cdot f(x_0) dx_0 \tag{7}$$

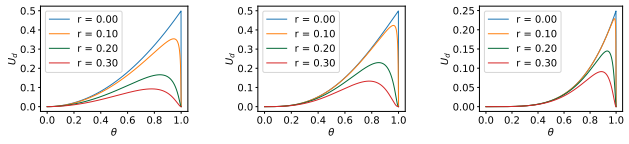
Eq. (7) above demonstrates that the decision-maker aims to maximize the total improvement among the agent population, and its utility is a function of θ. Since $f(x), x^*(\theta)$ are both continuous in $[0, 1]$, utility $U_d(\theta)$ is also continuous. The following Thm. 5.1 further shows the existence of the optimal thresholds $\theta^* \in (0, 1)$.

Theorem 5.1 (Existence of optimal threshold)**.** *For any decision-maker with utility function U_d, there exists at least one $\theta^* \in (0, 1)$ that is optimal under which $U_d(\theta) > 0$. Moreover, θ^* is the unique optimal point of U_d if $\frac{\partial U_d}{\partial \theta}$ has one root within $(0, 1)$.*

To verify Thm. 5.1, we demonstrate the values of U_d under situations where the agent population has different density functions f and different discounting factors r. Specifically, we consider the uniform distribution and Beta distributions with different parameters. Fig. 3 shows $U_d(\theta)$ under different density functions f and discounting factors r. The results illustrate that under these settings, U_d is single-peaked and there is a unique $\theta^* \in (0, 1)$ that is optimal and results in positive utility, which is consistent with Thm. 5.1. The figure also indicates the impact of r on the optimal threshold: as r increases, θ^* increases and the corresponding maximum utility decreases. As formally stated below in Corollary 5.2. We prove Thm. 5.1 and Corollary 5.2 in App. F.2.

Corollary 5.2. *For $U_d(\theta)$ that has a unique maximizer θ^*, optimal θ^* decreases as r increases.*

Importantly, the results of Thm. 5.1 show that the decision-maker can always find an optimal decision threshold θ^* (either numerically or using gradient methods depending on the density function f) to incentivize the largest improvement and promote *social welfare* in practice. While the above results all assume the decision-maker knows r when determining θ, we can relax this and provide a procedure to estimate r; this is included in App. E.



(a) $f = U(0,1)$ (b) $f = Beta(2,2)$ (c) $f = Beta(3,1)$

Figure 3: Optimal thresholds θ^* under different density functions f and discounting factors r.

6 Impact of Manipulative Behavior

Our analysis and results so far rely on an implicit assumption that agents are honest and they improve qualifications q_t by making actual efforts. However, as mentioned in Sec. 3, agents in practice may fool the decision-maker by strategically manipulating $x_t = q_t^T d$ to get accepted without improving q_t. Next, we extend our model in Sec. 3 by considering the possibility of such manipulative behavior.

Model with both manipulation & improvement. We extend the model in Sec. 3 where agents after observing θ have an additional option to manipulate x_0 directly. If they choose to ***improve***, they make a one-time effort $k \in [0,1]$ to acquire relevant knowledge and gradually improve their qualifications q_t over time based on (1). If they choose to ***manipulate***, they only increase x_t at every round to fool the decision-maker without changing the actual profile q_t. Similar to the literature on strategic classification (Hardt et al. 2016a), the manipulation comes at the cost and the risk of being caught.

Specifically, let $c(x', x) \geq 0$ be the *manipulation cost* it takes for an agent to increase its similarity from x to x', and $P \in [0,1]$ be the *detection probability* of manipulation during an agent's entire application process. Agents, once getting caught manipulating x_t, will never be accepted.

Degree of manipulation. If agents choose to manipulate, they will increase x_t at every round to fool the decision-maker, and they manipulate in a way that minimizes the manipulation cost and the risk of being detected. We make the following natural assumptions on c and P:

1. Let $\overline{x}_t$ be the best outcome agents can attain from x_{t-1} at round t by improvement behavior (with largest effort $k = 1$). If $x_t > \overline{x}_t$ for some t, then $P = 1$ because the decision-maker can be certain that x_t is the result of manipulation; otherwise, $P \in [0,1)$ if $x_t \leq \overline{x}_t$.
2. The total manipulation cost it takes for an agent with initial similarity x_0 to be accepted is $c(\theta, x_0)$.

Note that $\overline{x}_t$ above indicates the maximum degree of manipulation of agents: to avoid being detected, an agent should not manipulate x_t more than $\overline{x}_t$. We can compute $\overline{x}_t$ directly from Lemma 3.2 (by setting $k = 1$), i.e., $\overline{x}_t = \left(\frac{x_{t-1}^{-2}-1}{4}+1\right)^{-\frac{1}{2}}$. For agents who manipulate, if the total manipulation cost needed to get accepted is $c(\theta, x_0)$ and detection probability $P = 1$ whenever $x_t > \overline{x}_t$, then agents will always manipulate toward $\overline{x}_t$ to maximize its utility. As a result, agents who manipulate can be regarded as they mimic the improvement behavior with the largest effort $k = 1$.

Let $\widetilde{U}$ be agent's **utility under manipulation**, which is the benefit an agent obtains from acceptance (when not being detected) minus the manipulation cost, i.e.,

$$\widetilde{U} = (1-P)\cdot(1+r)^{\frac{-\ln\left(\sqrt{\frac{(\theta)^{-2}-1}{(x_0)^{-2}-1}}\right)}{\ln 2}} - c(\theta, x_0), \qquad (8)$$

where the benefit is derived based on (5) (with $k = 1$).

Agent's best response. Suppose agents have full information about detection probability P and discounting factor r, after observing the acceptance threshold θ, they best respond by choosing the action (i.e., improvement/manipulation/do nothing) that maximizes their utilities, i.e., if $\widetilde{U} > \max_k U$, they choose to manipulate; otherwise, they improve by exerting optimal effort $k^* = \arg\max_k U$.

Next, we examine under what conditions agents prefer improvement over manipulation.

Theorem 6.1. *Suppose manipulation cost $c(x', x) = (x' - x)_+$ and threshold $\theta \geq \bar{\theta}$ for some $\bar{\theta} \in (0,1)$. For any discounting factor r, there exists $\widehat{P} \in (0,1)$ such that the followings hold:*

1. If $P = 0$, then $\exists \widehat{x} \in (0,1)$ such that agents manipulate only when initial similarity $x_0 \in (\widehat{x}, \theta)$.

2. If $P \in (0, \widehat{P}]$, then $\exists \widehat{x_1}, \widehat{x_2}$ such that agents manipulate only when initial $x_0 \in (\widehat{x_1}, \widehat{x_2})$.

3. If $P > \widehat{P}$, then agents never choose to manipulate.

Thm. 6.1 considers scenarios when the threshold is sufficiently high, and identifies conditions under which manipulation is preferred by agents in these settings. It shows agent behavior highly depends on the risk of manipulation (i.e., detection probability P). The specific values of $\widehat{P}$, $\widehat{x}$, $\widehat{x_1}$, $\widehat{x_2}$ in Thm. 6.1 depend on θ, r. In particular, $\widehat{P}$ increases as r increases. Indeed, we can empirically find $\widehat{P}$, $\widehat{x}$, $\widehat{x_1}$, $\widehat{x_2}$ and verify the theorem. These are illustrated in App. C and Sec. 8.

7 Forgetting Mechanism

The analysis in previous sections relies on the assumption that once agents make a one-time effort k to acquire the knowledge, they never forget and can repeatedly leverage this knowledge to improve their profiles based on (1). This may not hold in practice when the knowledge agents acquired at the beginning are not sufficient to guarantee repeated improvements. In this section, we extend the qualification dynamics ((1)) by incorporating the *forgetting mechanism*, i.e., qualification profile q_t does not always move toward the direction of ideal profile d, instead, it may devolve and possibly go back to the initial q_0. Note that we only consider honest agents who do not manipulate. By modifying (1), we define the new **qualification dynamics with forgetting** as follows.

$$\begin{aligned} \widetilde{q}_{t+1} &= q_t + (k\cdot d + (1-k)\cdot q_0)\cdot q_t^T d \qquad (9)\\ q_{t+1} &= \frac{\widetilde{q}_{t+1}}{\|\widetilde{q}_{t+1}\|_2} \end{aligned}$$

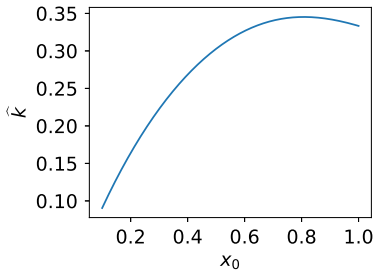


Figure 4: Upper bound $\widehat{k}$ of the optimal effort as a function of x_0.

Let $\widetilde{d} := k \cdot d + (1-k) \cdot q_0$, then new dynamics in (9) implies that at each round, qualification profile q_t is pushed toward the direction of $\widetilde{d}$, i.e., a convex combination of ideal profile d and initial qualifications q_0. Whether q_t improves towards d or deteriorates back to q_0 depends on the investment k: with more effort k, the degree of forgetting is less; there is no forgetting if all the knowledge is acquired ($k = 1$). Under the new dynamics, we can derive the convergence of the qualification profile as follows.

Theorem 7.1 (Convergence of qualification under forgetting)**.** *Consider an agent with initial similarity $x_0 = q_0^T d > 0$ whose qualifications q_t follow dynamics in* (9)*. Suppose the agent makes investment $k > 0$, then q_t converges to profile d^* and the similarity $x_t = q_t^T d$ satisfies:*

$$(x_t^*)^{-2} - 1 < \frac{(x_0^*)^{-2} - 1}{(k_u + 1)^{2t}} \tag{10}$$

where $d^* = \frac{\widetilde{d}}{\|\widetilde{d}\|}$, $x_t^* = q_t^T d^*$, and $k_u = \|\widetilde{d}\| \cdot x_0$.

Thm. 7.1 implies that convergence still holds when qualifications evolve with forgetting. Unlike the scenarios without forgetting where q_t eventually converges to the ideal profile d regardless of k (Lemma 3.2), q_t now converges to d^*, i.e., a profile between initial qualifications q_0 and ideal profile d, which is closer to q_0 with smaller investment k. It shows that if agents do not exert enough effort and the acquired knowledge is not sufficient, then they will not make satisfactory improvements.

Agent's utility and improvement action. Denote agent utility under the forgetting mechanism as $\widehat{U}(k, \theta, r, x_0)$. Unlike settings without forgetting where we can derive the exact time H it takes for agents to be accepted and find utility U ((5)), the analytical form of $\widehat{U}(k, \theta, r, x_0)$ is not easy to derive. Nonetheless, we can still show that there exist scenarios under which agents have incentives to improve, even though the best attainable profile is a profile d^* between initial q_0 and the ideal d.

Theorem 7.2. *For any threshold θ (resp. discounting factor r), there exists a discounting factor r (resp. threshold θ) such that agent's utility $\widehat{U}(\bar{k}, \theta, r, x_0) > 0$ for some $\bar{k} \in (0, \widehat{k})$, i.e., agents have the incentive to make a positive effort. The upper bound of the optimal effort is $\widehat{k}$ given by*

$$\widehat{k} = \min\left(\frac{\widehat{x}_0^2}{2\widehat{x}_0^2 + 2\widehat{x}_0^3}, \frac{x_0 \cdot (x_0^2 + x_0 - \sqrt{x_0^4 - x_0^2 + 1})}{2x_0^2 + 2x_0^3 - 1}\right) \tag{11}$$

where $\widehat{x}_0$ is the root of $2x_0^2 + 2x_0^3 - 1 = 0$ within $(0, 1)$.

Thm. 7.2 implies that there exists (θ, r) such that agents best respond by improving their qualifications, and the optimal effort is upper bounded by $\widehat{k}$. Indeed, we can numerically find the upper bound $\widehat{k}$ as a function x_0 (shown in Fig. 4). Because $\widehat{k} < 0.35$ for all x_0, the actual effort invested by any agent is less than 0.35, and the qualifications q_t converge to a profile d^* that is between q_0 and $0.35 \cdot d + 0.65 \cdot q_0$. This means the improvement an agent can make under the forgetting mechanism may be limited, suggesting that the agents may not improve to be qualified when the tasks are challenging.

8 Experiments

We validate theoretical results by conducting experiments on Exam score (Kimmons 2012) and FICO score (Reserve 2007) dataset[2]. For both datasets, scores serve as the agent's initial similarity x_0, and we assume agents interact with a decision maker based on the Stackelberg game in Sec. 3. We first fit these scores with beta distributions, i.e., $x_0 \sim \text{Beta}(v, w)$, and then use them to derive the followings:

1. The optimal decision threshold θ^* for the decision-maker to incentivize the largest amount of improvement and promote *social welfare*, and the total improvement induced by θ^*.
2. The percentage of agents who choose to manipulate under the decision-maker's optimal policy.

Exam Score Data. It is a synthetic dataset containing 1000 students' exam scores on 3 subjects including math, reading, and writing (Kimmons 2012). We first average over 3 subjects and normalize the averaged score to $[0, 1]$. Then, we fit two beta distributions to the normalized scores of males and females and obtain $x_0 \sim \text{Beta}(4.86, 2.37), \text{Beta}(4.15, 1.79)$ (see Fig. 10 in App. D).

With these distributions, we can compute the optimal decision thresholds and the corresponding total improvement under different discounting factors r. As shown in Fig. 5, for both males and females, the experimental results are similar. When r increases, θ^* always decreases and the total amount of improvement becomes lower. This illustrates how larger discounting factors harm agents' improvement. Additionally, we consider settings with both manipulation and improvement. Fig. 5 also shows the percentages of agents who prefer to manipulate under θ^*. It shows that agents are less likely to manipulate as detection probability P increases.

FICO Score Data. We adopt the data pre-processed by Hardt et al. (2016b), which contains CDF of credit scores of four racial groups (Caucasian, African American, Hispanic, Asian). For each group, we fit a Beta distribution and

[2]https://github.com/osu-srml/Alg-Persistent-Improvement

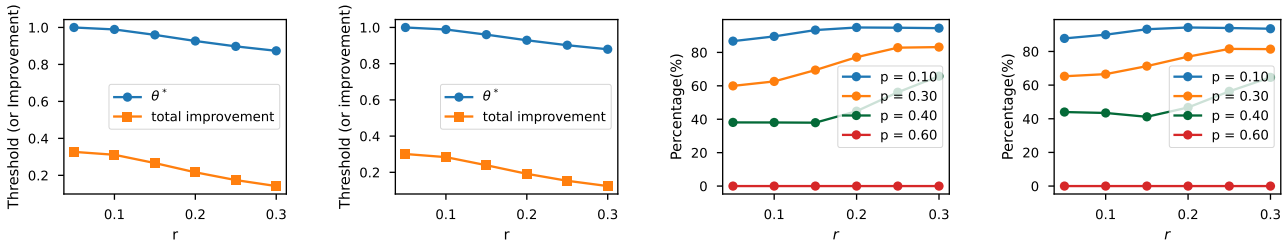


Figure 5: From the left to the right are: optimal thresholds to incentivize improvement for males/females; manipulation probability under the thresholds for males/females for **Exam data**.

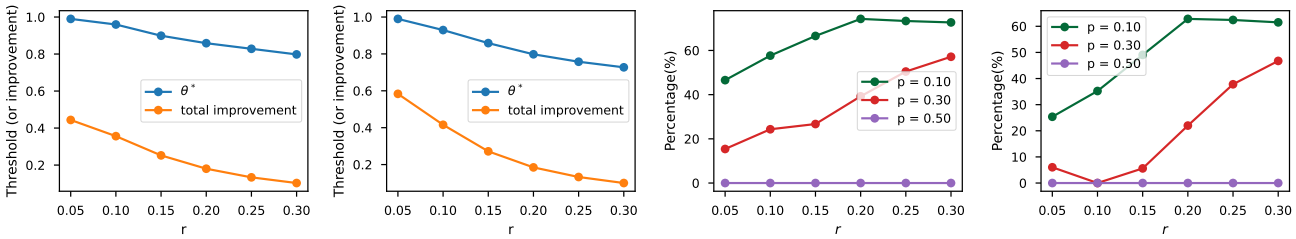


Figure 6: From the left to the right are: optimal thresholds to incentivize improvement for Caucasians and African Americans; manipulation probability under the thresholds for Caucasians and African Americans for **FICO data**.

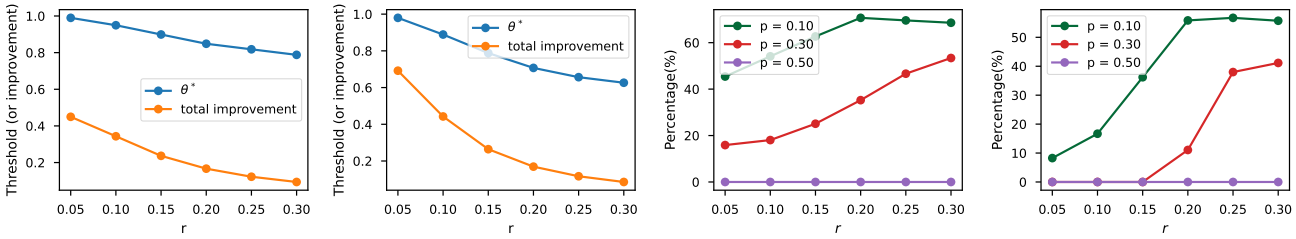


Figure 7: Optimal thresholds to incentivize improvement (left two plots) and manipulation probability under the thresholds (right two plots) for Asians and Hispanic of the **FICO data**.

obtain four distributions: Beta$(1.11, 0.97)$ for Caucasian, Beta$(0.91, 3.84)$ for African American, Beta$(0.99, 1.58)$ for Hispanic, Beta$(1.35, 1.13)$ for Asian (see Fig. 11 in App. D). The results for Caucasians and African Americans are shown on the first column in Fig. 7, while the results for Asian and Hispanic are shown the second column.

For each group, we compute the optimal decision threshold and corresponding total improvement under different r. As shown in Fig. 7 (left two plots), for both groups, their corresponding optimal threshold θ^* and the total amount of improvement always decrease as r increases. For settings with both manipulation and improvement (right two plots in Fig. 7), agents are more likely to manipulate under smaller detection probability. When detection probability P is sufficiently large, agents do not have incentives to manipulate.

9 Conclusions & Limitations

This paper studies the strategic interactions between agents and a decision-maker when agent action has delayed and persistent effects. By utilizing a qualification dynamics model and the discounting utility function, we analyze the conditions where agents tend to improve and investigate how the decision-maker can incentivize agents to make the largest improvement. Moreover, we consider the situation where agents can improve or manipulate, and characterize how agents would make improvement or manipulation decisions when their efforts take time to pay back. Finally, we discuss the situation where the tasks are challenging and a forgetting mechanism takes place, thereby expanding the scope of our model.

However, our theoretical results depend on the assumption that both agents and the decision-maker have perfect information about each other so that they always best respond. Extension to cases when each party only has partial or imperfect information is important. Moreover, these theorems are based on the qualification dynamics (1). Although a scenario when it does not hold is studied in Sec. 7, future works should also consider other variants tailored to spe-

cific applications to prevent negative outcomes. For example, they may consider the more complex situation where the decision-maker has a time-varying ideal profile d_t.

Ethical Consideration Statement

We believe our work fills the gap in which agent strategic behaviors are benign and agents' efforts can have long-lasting but diminishing effects. This can be the case under many real-world situations including exam preparation and job application. Thus, our model can improve trustworthy machine learning and decision-making in reality. However, as mentioned in Sec. 9, our work relies on certain assumptions and needs to be used cautiously. Moreover, though we provide a procedure to estimate the discounting factor, performing controlled experiments is not always accessible. Meanwhile, manipulation cost and detection probability are unknown and hard to estimate. Collecting real data and estimating these parameters remain promising research directions in the future.

Acknowledgement

This material is based upon work supported by the U.S. National Science Foundation under award IIS-2202699, by grants from the Ohio State University's Translational Data Analytics Institute and College of Engineering Strategic Research Initiative.

References

Ahmadi, S.; Beyhaghi, H.; Blum, A.; and Naggita, K. 2021. The strategic perceptron. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, 6–25.

Ahmadi, S.; Beyhaghi, H.; Blum, A.; and Naggita, K. 2022a. On classification of strategic agents who can both game and improve. *arXiv preprint arXiv:2203.00124*.

Ahmadi, S.; Beyhaghi, H.; Blum, A.; and Naggita, K. 2022b. Setting Fair Incentives to Maximize Improvement. *arXiv preprint arXiv:2203.00134*.

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

Barsotti, F.; Kocer, R. G.; and Santos, F. P. 2022. Transparency, Detection and Imitation in Strategic Classification. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, 67–73.

Bechavod, Y.; Ligett, K.; Wu, S.; and Ziani, J. 2021. Gaming helps! learning from strategic interactions in natural dynamics. In *International Conference on Artificial Intelligence and Statistics*, 1234–1242.

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

Ben-Porat, O.; and Tennenholtz, M. 2017. Best Response Regression. In *Advances in Neural Information Processing Systems*.

Braverman, M.; and Garg, S. 2020. The Role of Randomness and Noise in Strategic Classification. *CoRR*, abs/2005.08377.

Castellano, C.; Fortunato, S.; and Loreto, V. 2009. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2): 591.

Chen, Y.; Wang, J.; and Liu, Y. 2020. Strategic Recourse in Linear Classification. *CoRR*, abs/2011.00355.

Dean, S.; Curmei, M.; Ratliff, L.; Morgenstern, J.; and Fazel, M. 2024a. Emergent specialization from participation dynamics and multi-learner retraining. In *International Conference on Artificial Intelligence and Statistics*, 343–351. PMLR.

Dean, S.; Dong, E.; Jagadeesan, M.; and Leqi, L. 2024b. Accounting for AI and Users Shaping One Another: The Role of Mathematical Models. *arXiv preprint arXiv:2404.12366*.

Dean, S.; and Morgenstern, J. 2022. Preference Dynamics Under Personalized Recommendations. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, 795–816.

Dong, J.; Roth, A.; Schutzman, Z.; Waggoner, B.; and Wu, Z. S. 2018. Strategic Classification from Revealed Preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, 55–70.

Eilat, I.; Finkelshtein, B.; Baskin, C.; and Rosenfeld, N. 2022. Strategic Classification with Graph Neural Networks.

Gaitonde, J.; Kleinberg, J.; and Tardos, É. 2021. Polarization in geometric opinion dynamics. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, 499–519.

Grüne-Yanoff, T. 2015. Models of temporal discounting 1937–2000: An interdisciplinary exchange between economics and psychology. *Science in context*, 28(4): 675–713.

Hardt, M.; Jagadeesan, M.; and Mendler-Dünner, C. 2022. Performative Power. In *Advances in Neural Information Processing Systems*.

Hardt, M.; Megiddo, N.; Papadimitriou, C.; and Wootters, M. 2016a. Strategic Classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, 111–122.

Hardt, M.; Price, E.; Price, E.; and Srebro, N. 2016b. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems*.

Harris, K.; Heidari, H.; and Wu, S. Z. 2021. Stateful strategic regression. *Advances in Neural Information Processing Systems*, 28728–28741.

Hashimoto, T.; Srivastava, M.; Namkoong, H.; and Liang, P. 2018. Fairness without demographics in repeated loss minimization. In *International Conference on Machine Learning*, 1929–1938. PMLR.

Holmstrom, B.; and Milgrom, P. 1991. Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization*, 24–52.

Horowitz, G.; and Rosenfeld, N. 2023. Causal Strategic Classification: A Tale of Two Shifts. arXiv:2302.06280.

Izzo, Z.; Ying, L.; and Zou, J. 2021. How to Learn when Data Reacts to Your Model: Performative Gradient Descent. In *Proceedings of the 38th International Conference on Machine Learning*, 4641–4650.

Jagadeesan, M.; Mendler-Dünner, C.; and Hardt, M. 2021. Alternative Microfoundations for Strategic Classification. In *Proceedings of the 38th International Conference on Machine Learning*, 4687–4697.

Jin, K.; Xie, T.; Liu, Y.; and Zhang, X. 2024a. Addressing Polarization and Unfairness in Performative Prediction. *arXiv preprint arXiv:2406.16756*.

Jin, K.; Yin, T.; Chen, Z.; Sun, Z.; Zhang, X.; Liu, Y.; and Liu, M. 2024b. Performative federated learning: A solution to model-dependent and heterogeneous distribution shifts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 12938–12946.

Jin, K.; Zhang, X.; Khalili, M. M.; Naghizadeh, P.; and Liu, M. 2022. Incentive Mechanisms for Strategic Classification and Regression Problems. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, 760–790.

Kaelbling, L. P.; Littman, M. L.; and Moore, A. W. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4: 237–285.

Kimmons, R. 2012. Synthetic Exam Scores in a Public School. Technical report, Brigham Young University.

Kleinberg, J.; and Raghavan, M. 2020. How Do Classifiers Induce Agents to Invest Effort Strategically? 1–23.

Krahn, M.; and Gafni, A. 1993. Discounting in the economic evaluation of health care interventions. *Medical care*, 403–418.

Lechner, T.; Urner, R.; and Ben-David, S. 2023. Strategic classification with unknown user manipulations. In *International Conference on Machine Learning*, 18714–18732. PMLR.

Liu, L. T.; Dean, S.; Rolf, E.; Simchowitz, M.; and Hardt, M. 2019. Delayed Impact of Fair Machine Learning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 6196–6200.

Liu, L. T.; Garg, N.; and Borgs, C. 2022. Strategic ranking. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, 2489–2518.

Meier, S.; and Sprenger, C. D. 2013. Discounting financial literacy: Time preferences and participation in financial education programs. *Journal of Economic Behavior & Organization*, 95: 159–174.

Miehling, E.; Dong, R.; Langbort, C.; and Başar, T. 2019. Strategic inference with a single private sample. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2188–2193. IEEE.

Miller, J.; Milli, S.; and Hardt, M. 2020. Strategic Classification is Causal Modeling in Disguise. In *Proceedings of the 37th International Conference on Machine Learning*.

Perdomo, J.; Zrnic, T.; Mendler-Dünner, C.; and Hardt, M. 2020. Performative Prediction. In *Proceedings of the 37th International Conference on Machine Learning*, 7599–7609.

Raab, R.; and Liu, Y. 2021. Unintended selection: Persistent qualification rate disparities and interventions. *Advances in Neural Information Processing Systems*, 26053–26065.

Reserve, U. F. 2007. Report to the congress on credit scoring and its effects on the availability and affordability of credit. In *Board of Governors of the Federal Reserve System*.

Rosenfeld, N.; Hilgard, A.; Ravindranath, S. S.; and Parkes, D. C. 2020. From Predictions to Decisions: Using Lookahead Regularization. In *Advances in Neural Information Processing Systems*, 4115–4126.

Samuelson, P. A. 1937. A note on measurement of utility. *The review of economic studies*, 4(2): 155–161.

Shao, H.; Blum, A.; and Montasser, O. 2024. Strategic classification under unknown personalized manipulation. *Advances in Neural Information Processing Systems*, 36.

Shavit, Y.; Edelman, B. L.; and Axelrod, B. 2020. Causal Strategic Linear Regression. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20.

Sundaram, R.; Vullikanti, A.; Xu, H.; and Yao, F. 2021. PAC-Learning for Strategic Classification. In *Proceedings of the 38th International Conference on Machine Learning*, 9978–9988.

Xie, T.; and Zhang, X. 2024a. Automating Data Annotation under Strategic Human Agents: Risks and Potential Solutions. *arXiv preprint arXiv:2405.08027*.

Xie, T.; and Zhang, X. 2024b. Non-linear Welfare-Aware Strategic Learning. *arXiv preprint arXiv:2405.01810*.

Xie, T.; Zuo, Z.; Khalili, M. M.; and Zhang, X. 2024. Learning under Imitative Strategic Behavior with Unforeseeable Outcomes. *arXiv preprint arXiv:2405.01797*.

Yin, T.; Raab, R.; Liu, M.; and Liu, Y. 2024. Long-term fairness with unknown dynamics. *Advances in Neural Information Processing Systems*, 36.

Zhang, X.; Khalili, M. M.; Jin, K.; Naghizadeh, P.; and Liu, M. 2022. Fairness interventions as (dis) incentives for strategic manipulation. In *International Conference on Machine Learning*, 26239–26264. PMLR.

Zhang, X.; Khalili, M. M.; and Liu, M. 2020. Long-term impacts of fair machine learning. *ergonomics in design*, 28(3): 7–11.

Zhang, X.; Khalili, M. M.; Tekin, C.; and Liu, M. 2019. Group retention when using machine learning in sequential decision making: the interplay between user dynamics and fairness. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 15269–15278.

Zhang, X.; and Liu, M. 2021. Fairness in learning-based sequential decision algorithms: A survey. In *Handbook of Reinforcement Learning and Control*, 525–555.

Zhang, X.; Tu, R.; Liu, Y.; Liu, M.; Kjellstrom, H.; Zhang, K.; and Zhang, C. 2020. How do fair decisions fare in long-term qualification? In *Advances in Neural Information Processing Systems*, 18457–18469.

A Discussion and Generalization of (1)

More details on the dynamics in (1). In the main paper, we assume the influence of the initial effort k is persistent and will enable q_t changes gradually during each round. This is well-supported by the following examples:

1. *Creditworthiness*: To improve creditworthiness, an individual may learn that an ideal profile would be a person with a constant high income and long-lasting good credit history. Therefore, she may exert a significant effort to find a job with a high salary. However, the effort will take several months or even one year for her to finally build up the ideal profile because she needs to work for a while to receive money and build a competitive credit history.
2. *Job application*: An individual who wants to apply for a technology company may learn about the skill set of an ideal candidate from several resources (e.g., the job description, alumni who work at the company, info session) and then exert a significant effort to study the required knowledge. However, it still takes time for her to do exercises and master the skills, resulting in a delay of finally being qualified.

Model generalization when agents can invest efforts at different time steps. We discuss how the model in the main paper can capture more complicated scenarios where agents repeatedly exert efforts multiple times until they reach the target. Each effort has persistent effects on improving the qualification as shown in Eqn. (12).

$$\widetilde{q}_{t+1} = q_t + \sum_{\tau=0}^{t} k_\tau \cdot q_t^T d \cdot d \tag{12}$$
$$q_{t+1} = \frac{\widetilde{q}_{t+1}}{\|\widetilde{q}_{t+1}\|_2}$$

where $\sum_{\tau=0}^{t} k_\tau \in [0, 1]$. This means the agents are able to invest more effort at arbitrary time steps (e.g., studying more skills in the middle of the preparing process), but the cumulative effort should not exceed 1 (they cannot master 110% of knowledge).

We first prove that there exists an effort $k^* \in [0, 1]$ such that investing k^* once at the beginning has the same impact on $\lim_{t\to\infty} q_t$ as investing a sequence of efforts $\{k_t\}_{t\geq 0}$ over time: define q_t^{min}, q_t^{max} as the "what-if" qualifications if the agents invest $k = 0$ or $k = 1$ at the initial round. Since $\sum_{\tau=0}^{t} k_\tau \in [0, 1]$, we know the q_t must be between q_{min}^t, q_{max}^t. Then because q_t is continuous with respect to k, so we know k^* must exist. Therefore, our model in the main paper can indeed assimilate the more complex setting.

Model generalization when k diminishes with t. In the main paper, k_t is always equal to k, demonstrating the effort has a consistent and persistent effect on the improvement of an individual. According to Lemma 3.2, the similarity x_t approaches 1 at an exponential rate. Thus, the case of $k_t > k$ is not interesting since the convergence is faster and it may not make sense in practice that the effort can be increasingly effective as time goes on. However, in reality, it may be possible that k_t is decreasing. This is a "middle-point" case between the regular improvement in (1) and the forgetting mechanism (9), which may illustrate the "tiredness" when agents stick to improve. However, we can prove that when k_t decreases linearly (i.e., $k_t = \Theta(\frac{k}{t})$), the similarity x_t can only converge to 1 at a speed $\Theta(t^k)$.

Theorem A.1. *When k_t decreases linearly (i.e., $k_t = \Theta(\frac{k}{t+1})$), x_t converges to 1 at a rate $\Theta(t^k)$*

We prove Thm. A.1 in App. F.6. Basically, this result illustrates that the agents will still improve to be qualified if k_t decreases at a linear rate. Specifically, we can rewrite the (2) as:

$$x_t^{-2} - 1 = \frac{(x_0)^{-2} - 1}{(t+1)^{2k}} \tag{13}$$

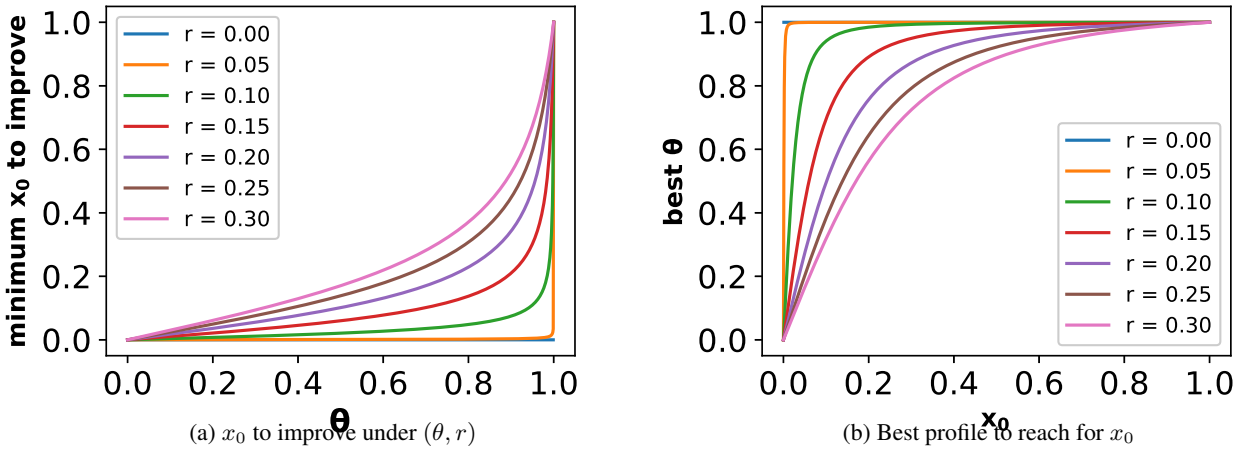
From (13), we can derive similar results of the agents' best responses and work out the thresholds for them to improve.

B Illustration of Table 1

Table 1 illustrate the minimum requirement of x_0 for an individual to improve under different (θ, r), and the best attainable profile for individuals with initial similarity x_0. We illustrate them in Fig. 8.

Discussions of intervention strategies in real applications. Table 1 further suggest effective strategies that encourage individuals to improve their qualifications, i.e., more individuals are incentivized to improve if (i) the decision-maker's acceptance threshold θ is lower; or (ii) the time it takes for individuals to succeed after investments is shorter. Examples of both strategies in real applications are as follows.

1. *Lower acceptance threshold θ in hiring:* Instead of directly recruiting the qualified candidates, companies first lower the standard by offering internship opportunities to encourage applicants to improve, and then offer full-time positions. This two-stage hiring process widens the candidate pool and incentivizes more people to improve.



(a) x_0 to improve under (θ, r)

(b) Best profile to reach for x_0

Figure 8: Illustration of Table 1

2. *Lower discounting factor r in college admission:* Instead of directly rejecting the unqualified high school graduates, universities incentivize them by issuing conditional transfer offers. Once these students meet certain requirements, they get admitted. The conditional acceptances encourage more students to improve by lowering the time it takes for them to receive reward.

Meanwhile, Table 1 also reveals that setting short-term goals will be effective to incentivize individuals to improve. For instance, teachers may set up several quizzes to break down the grade and make students more motivated to improve.

C Illustration of Thm. 6.1

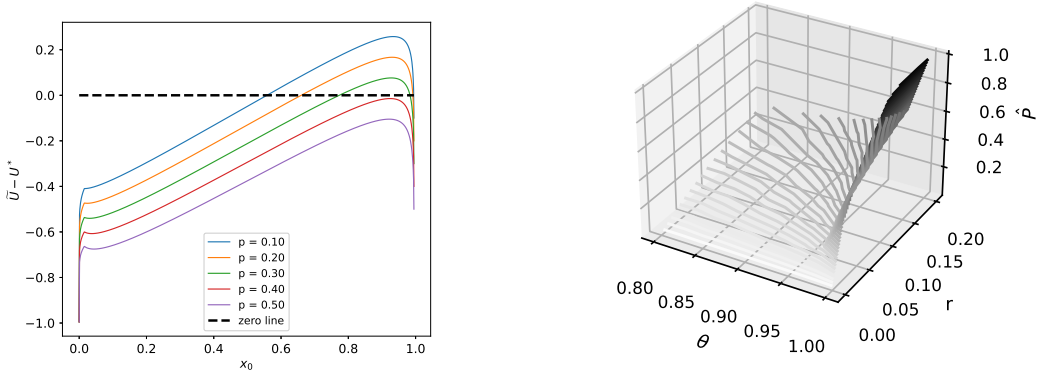


Figure 9: Illustration of Thm. 6.1: the left figure shows $\widetilde{U} - U$ as functions of x_0 under different P when $\theta = 0.995, r = 0.05$; the right plot shows threshold $\widehat{P}$ under different pairs of (θ, r).

Thm. 6.1 identifies conditions under which manipulation (or improvement) is preferred by individuals over the other. As mentioned in Section 6, the specific values of $\widehat{P}$, $\widehat{x}$, $\widehat{x_1}$, $\widehat{x_2}$ in Thm. 6.1 depend on θ, r, and we can empirically find $\widehat{P}$, $\widehat{x}$, $\widehat{x_1}$, $\widehat{x_2}$ and verify the theorem, as illustrated in Figure 9 and Table 2. Specifically, the left plot in Figure 9 shows $\widetilde{U} - U$ as functions of initial similarity x_0 under different detection probability P. Because individuals only prefer to manipulate if $\widetilde{U} - U > 0$, the plot shows the values of $\widehat{P}$, $\widehat{x}$, $\widehat{x_1}$, $\widehat{x_2}$ in Thm. 6.1. The right plot shows threshold $\widehat{P}$ under different pairs of (θ, r), and it shows that $\widehat{P}$ increases as r increases. Table 2 shows ranges $(\widehat{x_1}, \widehat{x_2})$ of initial similarity x_0 under different detection probability P, acceptance threshold θ, and discounting factor r.

Table 2: Ranges $(\widehat{x_1}, \widehat{x_2})$ of initial similarity x_0 under which individuals prefer to manipulate.

θ	r	Detection probability P					
		0	0.1	0.2	0.3	0.4	0.5
0.995	0.1	$(0.364, 0.995)$	$(0.435, 0.994)$	$(0.513, 0.993)$	$(0.596, 0.991)$	$(0.686, 0.984)$	$(0.796, 0.966)$
0.976	0.05	$(0.499, 0.976)$	$(0.613, 0.973)$	$(0.740, 0.958)$	$\emptyset$	$\emptyset$	$\emptyset$
0.953	0.01	$(0.773, 0.953)$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

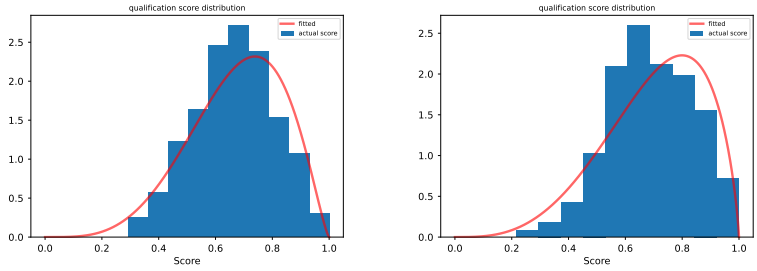


Figure 10: Exam Score: $Beta$ distributions

D Additional Experiments

Exam Score Data Just as Sec. 8 mentions, we acquire the exam score data (Kimmons 2012), preprocess the data and fit beta distributions for both males and females. The fitted distribution and real distribution are shown in Fig. 10.

FICO Score Data Just as Sec. 8 mentions, we fit beta distributions for FICO Score (Hardt et al. 2016b), and obtain four distributions for different racial groups as shown in Fig. 11.

E Estimating the discounting factor r in Sec.5

We can estimate the discounting factor r if given an experimental population. The decision-maker can publish an arbitrary threshold θ and observe the lowest score among all individuals who change their scores, which is $x^*(\theta)$. Then the decision-maker can use any expression in Table 1 to estimate r. Multiple experiments can make the estimation more robust.

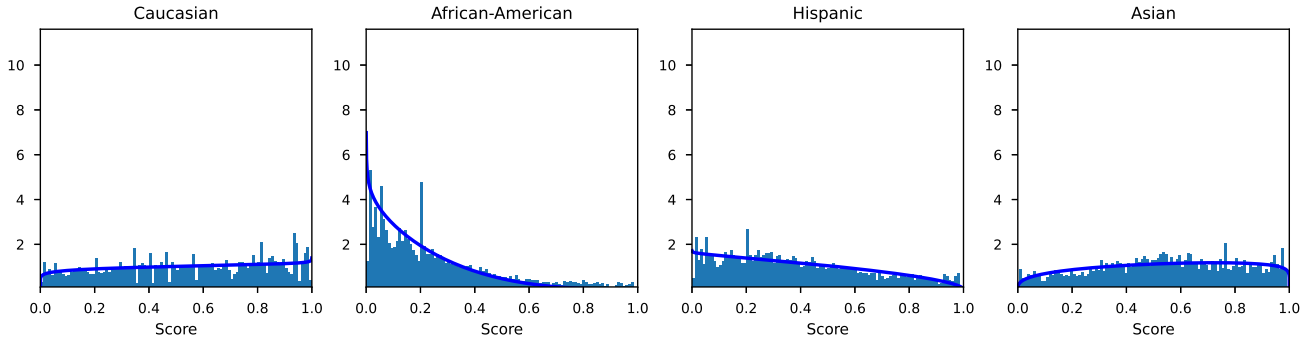


Figure 11: FICO Score: Caucasian (Beta$(1.11, 0.97)$), African American (Beta$(0.91, 3.84)$), Hispanic (Beta$(0.99, 1.58)$), Asian (Beta$(1.35, 1.13)$)

F Proofs

F.1 Proof Details of Thm. 4.1

To derive k^*, we first take the derivative of (5) with respect to k. For simplicity, let $K = k + 1$ and the derivative will not change. Also, let $R = r + 1$ and $G = -\ln\left(\sqrt{\frac{(\theta)^{-2}-1}{(x_0)^{-2}-1}}\right)$. Then show the results as follows:

$$\frac{\partial U}{\partial K} = \ln R \cdot R^{\frac{-G}{\ln K}} \cdot \frac{G}{K \cdot \ln^2 K} - 1 \tag{14}$$

$$\frac{\partial^2 U}{\partial K^2} = \frac{-G \cdot \ln R \cdot R^{\frac{-G}{\ln K}} (\ln^2 K + 2 \ln K - G \cdot \ln R)}{K^2 \cdot \ln^4 K} \tag{15}$$

The denominator of $\frac{\partial^2 U}{\partial K^2}$ is always positive, and the first term $-G \cdot \ln R \cdot R^{\frac{-G}{\ln K}}$ of numerator is always negative.

Also, because $K \in [1, 2]$, $\ln^2 K + 2 \ln K \in (0, \ln^2 2 + 2 \ln 2)$. Thus, we have following situations:

1) If $G \cdot \ln R > \ln^2 2 + 2 \ln 2$, $\frac{\partial^2 U}{\partial K^2}$ is always positive when $K \in [1, 2]$. This means $\frac{\partial U}{\partial K}$ is increasing.

Then, noticing that $\lim_{K \to 1^+} \frac{\partial U}{\partial K} = -1$, we know $\frac{\partial U}{\partial K}$ is always negative when $K \in [1, 2]$. This means U is monotonically decreasing. Also, when $k = 0$, $U = 0$. This ensures U is always non-positive and individuals will never choose to invest any effort.

2) If $G \cdot \ln R \leq \ln^2 2 + 2 \ln 2$, $\frac{\partial^2 U}{\partial K^2}$ is first positive, then negative when $K \in [1, 2]$. Also, if plugging $K = 2$ into (14), we know $\lim_{K \to 2} \frac{\partial U}{\partial K} < 0$. These facts reveal that $\frac{\partial U}{\partial K}$ is firstly increasing from a negative number and then decreasing to a negative number. And there must exist a unique maximum point when $K = K^{'}$, $K^{'}$ should satisfy:

$$\ln^2 K^{'} + 2 \ln K^{'} - G \cdot \ln R = 0 \tag{16}$$

Plug (16) into (14). Denote $\ln K^{'}$ as $t \in [0, \ln 2]$, and denote $\frac{\partial U}{\partial K}$ at $K^{'}$ as L:

$$L = \frac{t + 2}{t \cdot e^{2t+2}} - 1 \tag{17}$$

Then take the derivative of L:

$$\frac{\partial L}{\partial t} = \frac{-2(t+1)^2 \cdot e^{2t+2}}{t^2 \cdot e^{4t+4}} < 0 \tag{18}$$

(18) shows L is decreasing. Also, noticing that $\lim_{t \to 0^+} L(t) = +\infty$ and $\lim_{t \to \ln 2} L(t) < \frac{3}{2e^2} - 1 < 0$, we know there must exist a $t^{'} \in (0, \ln 2)$ as the root of M. We can explicitly solve $t^{'} = 0.1997$.

Thus, we now know that when $t \in [0, t^{'}]$, $L \geq 0$. With the plausible domain of t and (16), we would know: When $G \ln R \in [0, t^{'2} + 2t^{'}]$, $L \geq 0$ and thereby U has an extreme large point with value U^*. At this maximum point, (14) equals 0, and (15) is smaller than 0.

Finally, we derive the condition for $U^* > 0$: Denote $G \ln R$ as C and $\ln K$ as z, U can be simplified to:

$$U = e^{\frac{-C}{z}} - e^z + 1 \tag{19}$$

Because $z \in [0, \ln 2]$, for any t fixed, $\lim_{C \to 0} U = 2 - e^z \geq 0$ and $\lim_{C \to 0} U = 1 - e^z \leq 0$. With the fact that $\frac{\partial U}{\partial C} < 0$, we know U is monotonically decreasing with C, so is U^*. Thus, there must exist a threshold m, when $C < m$, $U^* > 0$. And if $U^* > 0$, individuals will decide to improve. Then Thm. 4.1 is proved and we can numerically solve the threshold $m = 0.316$.

Although we believe exponential discounting is general and fits our setting well, we also note that we can still use derivative analysis when the discounting changes (e.g., hyperbolic discounting). Specifically, if denoting the discounted reward as $d(r, t)$, we would have $U = d(r, H) - k$. Then if taking the derivative we will get $\frac{\partial U}{\partial k} = \frac{\partial d}{\partial H} \cdot \frac{\partial H}{\partial k} - 1$. Noticing that H is known, then discussing the properties of d with different choices of discounting is enough to derive the nature of U.

F.2 Proof Details of Thm. 5.1 and Corollary 5.2

Proof of Thm. 5.1 First prove $U_d(\theta)$ has a maximize $\theta^* \in (0, 1)$:

With the definition of $U_d(\theta)$ in (7), we already know U_d is continuous. We can first observe that $U_d(0) = 0, U_d(1) = 1$. These hold simply because $x^*(0) = 0 and x^*(1) = 1$. Next noticing that for any $\theta \in (0, 1)$, $U_d(\theta) > 0$ holds. This suggests that θ will reach its maximum point according to the Weierstrass extreme value theorem.

Next, noticing that $U_d(\theta) > 0 \in (0, 1)$ we can derive that $\frac{\partial U_d}{\partial \theta}(0) > 0$ and $\frac{\partial U_d}{\partial \theta}(0) < 0$. Then if it only has one root in $(0, 1)$, we would know U_d must first increase and then decrease because there is at most one inflection point. Thus, a unique maximum exists.

Proofs of why Uniform distribution has a unique maximized θ^* If $\frac{\partial U_d}{\partial \theta}$ only has one root. We know it is first larger than 0, then becomes smaller than 0. Next, according to the Leibniz integral rule, we can get:

$$\frac{\partial U_d}{\partial \theta} = \int_{x^*(\theta)}^{\theta} P(x)dx - (\theta - x^*(\theta)) \cdot P(x^*(\theta)) \cdot \frac{\partial x^*(\theta)}{\partial \theta}$$

Use Lagrange's Mean Value Theorem, we can write the above equation as:

$$(\theta - x^*(\theta)) \cdot [P(\theta^{'}) - P(x^*(\theta)) \cdot \frac{\partial x^*(\theta)}{\partial \theta}]$$

where $\theta^{'}$ is between $x^*(\theta), \theta$. Thus, the second term $P(\theta^{'}) - P(x^*(\theta)) \cdot \frac{\partial x^*(\theta)}{\partial \theta}$ must also be first larger than 0 then smaller than 0. Next, noticing that $P(\theta^{'}) = P(x^*(\theta))$ in uniform distribution and $\frac{\partial x^*(\theta)}{\partial \theta}$ is increasing, the equation will be smaller than 0 when $\frac{\partial x^*(\theta)}{\partial \theta} < 1$ and vice versa. Thus, we prove the result for the uniform distribution.

Proof of Corollary 5.2 We now know $\frac{\partial U_d}{\partial \theta} = (\theta - x^*(\theta)) \cdot [P(\theta^{'}) - P(x^*(\theta))] \cdot \frac{\partial x^*(\theta)}{\partial \theta}]$. Then according to the expression of $x^*(\theta)$, it is true that both $x^*(\theta)$ and $\frac{\partial x^*(\theta)}{\partial \theta}$ increase with r. Thus, when the probability distribution remains unchanged, the root of $\frac{\partial U_d}{\partial \theta}$ when r increases becomes smaller.

F.3 Proof Details of Thm. 6.1

Denote $\ln\left(\sqrt{\frac{\theta^{-2}-1}{x_0^{-2}-1}}\right)$ as $G(x_0)$. $G(x_0)$ is always negative and monotonically increasing with $x_0 \in (0, \theta)$.

1. Situation when P = 0: According to Sec. 4 and (8), we can write the maximum improvement utility U^* as $(1+r)^{\frac{G(x_0)}{ln(k^*+1)}} - k^*$, and write manipulation utility $\widetilde{U}$ as $(1+r)^{\frac{G(x_0)}{ln2}} - (\theta - x_0)$.

Then take the derivative of both:

$$\frac{\partial U^*}{\partial x_0} \geq \frac{\partial G}{\partial x_0} \cdot \frac{ln(1+r)}{ln(k^*+1)} \cdot (1+r)^{\frac{G(x_0)}{ln(k^*+1)}} \tag{20}$$

$$\frac{\partial \widetilde{U}}{\partial x_0} = \frac{\partial G}{\partial x_0} \cdot \frac{ln(1+r)}{ln2} \cdot (1+r)^{\frac{G(x_0)}{ln2}} + 1 \tag{21}$$

The "$\geq$" in (20) occurs because k^* is actually a function of x_0, but if we regard k^* at x_0 as a constant, the derivative here serves as a lower bound of $\frac{\partial U^*}{\partial x_0}$.

Firstly, we prove when $x_0 \to \theta$, $U^* < \widetilde{U}$: when $x_0 \to \theta$, we know $k^* \to 0$ since individuals invest an arbitrarily small effort to immediately qualified. However, according to Sec. F.1, k^* should let $\frac{\partial^2 U}{\partial k^2} < 0$. This inequality will give us the bound of k^*: $\ln(k^*+1) > \frac{-G(x_0) \cdot ln(1+r)}{3}$. With this bound, we can plug k^* into (20), and know $\frac{ln(1+r)}{ln(k^*+1)} \to +\infty$, and $(1+r)^{\frac{G(x_0)}{ln(k^*+1)}}$ is larger than a constant because of the bound. Therefore, $\frac{\partial U^*}{\partial x_0} \geq \frac{\partial G}{\partial x_0} \cdot +\infty$. Then according to (21), when $x_0 \to \theta$, $\frac{\partial \widetilde{U}}{\partial x_0} < \frac{\partial G}{\partial x_0} \cdot \frac{ln(1+r)}{ln2} + 1$. Since $\frac{\partial G}{\partial x_0}$ is always positive, when $x_0 \to \theta$, we prove that $\frac{\partial U^*}{\partial x_0} > \frac{\partial \widetilde{U}}{\partial x_0}$. Meanwhile, when $x_0 = \theta$, $U^* = \widetilde{U} = 1$. This means when $x_0 \to \theta$, $U^* < \widetilde{U}$.

Secondly, when $x_0 = 0$: $\widetilde{U} = -\theta$ and $U^* = 0$. So $\widetilde{U} < U^*$ when $x_0 = 0$.

Thus. there must be an intersection between $\widetilde{U}$ and U^*. Then noticing that if we increase θ, $\widetilde{U}$ is always decreasing to converge to function $y = x - 1$, while $U^* \geq 0$ always holds. This suggests when θ is sufficiently close to 1, i.e., there exists a $\bar{\theta}$ and when $\theta > \bar{\theta}$, we can guarantee the first intersection of U^* and $\widetilde{U}$ occurs arbitrarily close to 1, meaning this first intersection is the only intersection.

Let the only *intersection* be $\hat{x}$, we prove situation 1. The shapes of $\widetilde{U}$ and U^* are illustrated in Fig. 12.

2. Situation when P > 0: From (8): when $x_0 \to \theta$, $\widetilde{U} \to 1 - P$. However, at this time $U^* \to 1 > 1 - P$. This demonstrates $\hat{x_2}$ must exist.

When $P \to 0$, according to situation 1 and the continuity of $\widetilde{U}$ with respect to P, $\hat{x_1}$ must exist. However, when $P \to 1$, $\widetilde{U}$ is always negative, making $\hat{x_1}$ does not exist.

Thus, there must exist a threshold $\hat{P}$, when $P \leq \hat{P}$, $\hat{x_1}, \hat{x_2}$ exist. Otherwise, $U^* > \widetilde{U}$ is always true.

F.4 Proof Details of Thm. 7.1

Thm. 7.1 can be proved by the inequality $\|\widetilde{d}\| q_t^T d > \|\widetilde{d}\| q_0^T d > \|\widetilde{d}\| q_0^T d q_t^T d^*$. This means we can just let $k_u = \|\widetilde{d}\| q_0^T d = \|\widetilde{d}\| x_0$ and directly apply Lemma 3.2.

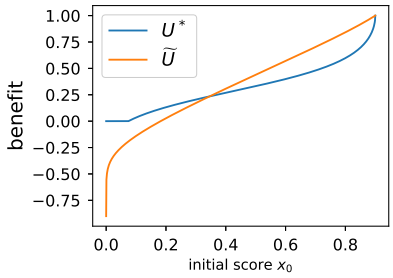
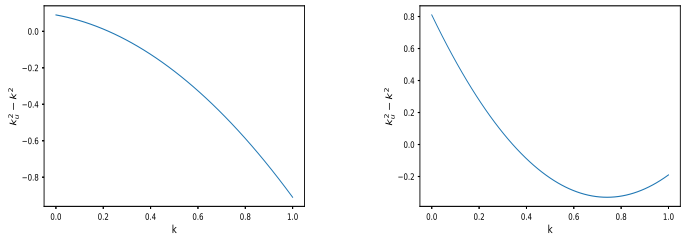


Figure 12: $\widetilde{U}, U^*$



(a) shape when $2x_0^2 + 2x_0^3 - 1 < 0$ (b) shape when $2x_0^2 + 2x_0^3 - 1 > 0$

Figure 13: Shapes of $k_u^2 - k^2$

F.5 Proof Details of Thm. 7.2

First let us prove following two lemmas:

Lemma F.1. *For any initial qualification score x_0, There exists a $\widehat{k} \in (0,1)$, when $k \in [0, \widehat{k})$, $k_u > k$. Let $\widehat{x_0}$ be the only root of $2x_0^2 + 2x_0^3 - 1 = 0$ within $(0,1)$, then $\widehat{k}$ is given by:*

$$\widehat{k} = \min\left(\frac{\widehat{x_0}^2}{2\widehat{x_0}^2 + 2\widehat{x_0}^3}, \frac{x_0 \cdot (x_0^2 + x_0 - \sqrt{x_0^4 - x_0^2 + 1})}{2x_0^2 + 2x_0^3 - 1}\right) \tag{22}$$

Proof. According to Thm. 7.1, $k_u^2 = \|\widetilde{d}\|^2 \cdot x_0^2$ and $\|\widetilde{d}\|^2 = k^2 + (1-k)^2 - 2k(1-k)x_0$. We can get following expression:

$$k_u^2 - k^2 = (2x_0^2 + 2x_0^3 - 1)k^2 - (2x_0^2 + 2x_0^3)k + x_0^2 \tag{23}$$

Firstly, when $2x_0^2 + 2x_0^3 - 1 = 0$, $\widehat{x_0} = 0.565$. Thus, when $k < \frac{\widehat{x_0}^2}{2\widehat{x_0}^2 + 2\widehat{x_0}^3} = 0.319$, $k_u^2 > k^2$.

Except the above situation, We can regard (23) as a quadratic function of k and solve the two roots:

$$\frac{x_0 \cdot (x_0^2 + x_0 \pm \sqrt{x_0^4 - x_0^2 + 1})}{2x_0^2 + 2x_0^3 - 1} \tag{24}$$

We then prove a claim that when $x_0 \in (0,1)$, $\frac{x_0 \cdot (x_0^2 + x_0 + \sqrt{x_0^4 - x_0^2 + 1})}{2x_0^2 + 2x_0^3 - 1}$ is either larger than 1 or smaller than 0:

1) When $2x_0^2 + 2x_0^3 - 1 < 0$, the denominator of (24) is negative, while the numerator is always positive. Thus, (24) is negative.

2) When $2x_0^2 + 2x_0^3 - 1 > 0$:

$$\frac{x_0 \cdot (x_0^2 + x_0 + \sqrt{x_0^4 - x_0^2 + 1})}{2x_0^2 + 2x_0^3 - 1} > \frac{x_0 \cdot (x_0^2 + x_0 + x_0^2)}{2x_0^3 + x_0^2} = 1 \tag{25}$$

(25) means (24) is larger than 1. Thus, the claim is proved.

Thus, $k_u^2 - k^2$ only has one root within $(0,1)$. Also from (23) we know when $k = 0$, $k_u > k$ and when $k = 1$, $k_u \leq k$. With these facts we immediately know: When $k \leq \frac{x_0 \cdot (x_0^2 + x_0 - \sqrt{x_0^4 - x_0^2 + 1})}{2x_0^2 + 2x_0^3 - 1}$, $k_u^2 - k^2 \geq 0$. Otherwise, $k_u^2 - k^2 < 0$. In fact, besides

the exception $2x_0^2 + 2x_0^3 - 1 = 0$, there are only two possibilities of the shape of $k_u^2 - k^2$ as shown in Fig. 13. Because k and k_u are both non-negative, the relationship of the square must be the same for their values.

Then if we define $\hat{k}$ as:

$$\hat{k} = \min\left(\frac{\widehat{x_0}^2}{2\widehat{x_0}^2 + 2\widehat{x_0}^3}, \frac{x_0 \cdot (x_0^2 + x_0 - \sqrt{x_0^4 - x_0^2 + 1})}{2x_0^2 + 2x_0^3 - 1}\right) \tag{26}$$

Then $k_u > k$ when $k \in [0, \hat{k})$. Proved.

Lemma F.2. *For any individual with initial qualification score x_0 and the admission threshold θ, there must exist a r to let there exists a $\bar{k} \in [0, \widehat{k})$, $U(\bar{k}, \theta, r, x_0) > 0$*

Proof. If we let $z = \ln(k+1)$ be z and recall that $C(\theta, x_0, r) = -\ln\left(\sqrt{\frac{(\theta)^{-2}-1}{(x_0)^{-2}-1}}\right) \cdot \ln(1+r)$, we would have $U = e^{\frac{-C}{z}} - e^z + 1$.

For any z there exists C_z, when $C < C_z$, $U > 0$.

So we can just let k be an arbitrary point $\in [0, \widehat{k})$ and we can get the corresponding C_z, then we can only let r satisfy:

$$\ln(1+r) < \frac{C_z}{-\ln\left(\sqrt{\frac{(\theta)^{-2}-1}{(x_0)^{-2}-1}}\right)} \tag{27}$$

Then we find the plausible r. Proved.

Proof of Thm. 7.2. According to Lemma F.1, when $\bar{k} \in [0, \widehat{k})$, $k_u > k$, so the convergence speed of the individual to d^* under forgetting mechanism will be faster than the convergence speed of the individual to d without forgetting mechanism, so that the reward under forgetting mechanism is discounting less. Meanwhile, according to Lemma F.2, there exists a r where $U(\bar{k}, \theta, r, x_0) > 0$. Combine them together, $\widehat{U}(\bar{k}, \theta, r, x_0) > U(\bar{k}, \theta, r, x_0) > 0$ and Thm. 7.2 is proved.

F.6 Proof of Thm. A.1

Assume $k_t = \frac{k}{t+1}$ when $t \geq 0$. From (1) and similar to (Dean and Morgenstern 2022), we know $(q_{t+1}^T \cdot d)^{-2} - 1 = \frac{(q_t^T \cdot d)^{-2} - 1}{k_t + 1}^2$. This will lead to $(q_t^T \cdot d)^{-2} - 1 = \prod_{i=0}^{t-1}(\frac{k}{i+1} + 1)^{-2}((q_0^T \cdot d)^{-2} - 1)$.

Then consider $\prod_{i=0}^{t-1}(\frac{k}{i+1} + 1)^{-1} = \prod_{i=0}^{t-1}(\frac{i+1}{k+i+1}) = \frac{1}{k+1} \cdot \frac{2}{k+2}$.... When $k = 1$, The expression is $\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} ... \frac{t-1}{t} = \frac{1}{t}$, demonstrating the convergence rate is linear. Note that this expression is decreasing as k decreases, so the convergence rate in our model is always slower than linear. Next, consider the general expression $\prod_{i=0}^{t-1}(\frac{i+1}{k+i+1}) = \frac{1}{k+1} \cdot \frac{2}{k+2}$... and $k < 1$. Let $a = \frac{1}{k}$ which is larger than 1, and $j = i + 1$ which is larger than 0. We slightly abuse the definition of a to let it be an integer. Then the expression becomes $\prod_{i=0}^{t-1}(\frac{ja}{1+ja}) = \frac{a}{a+1} \cdot \frac{2a}{2a+1} \cdots \frac{ta}{ta+1}$.

Then for any a we can bound this expression. Basically, we already know $\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} ... \frac{t-1}{t} = \frac{1}{t}$. Noticing that when $a > 1$, it is just equal to erase some terms of this expression. We can utilize this fact to get the lower bound and upper bound:

1. Lower bound: consider the following $a-1$ sets of expressions and each set consists of t terms: $\{\frac{1}{2} \cdot \frac{a+1}{a+2} \cdot \frac{2a+1}{2a+2} \cdots \cdot \frac{(t-1)a+1}{(t-1)a+2}\}$, $\{\frac{2}{3} \cdot \frac{a+2}{a+3} \cdot \frac{2a+2}{2a+3} \cdots \cdot \frac{(t-1)a+2}{(t-1)a+3}\}$..., $\{\frac{a-1}{a} \cdot \frac{2a-1}{2a} \cdot \frac{3a-1}{3a} \cdots \cdot \frac{ta-1}{ta}\}$. Then each of the $a-1$ expressions are smaller than $\prod_{j=1}^{t}(\frac{ja}{1+ja}) = \frac{a}{a+1} \cdot \frac{2a}{2a+1} \cdots \frac{ta}{ta+1}$. Denote $\prod_{j=1}^{t}(\frac{ja}{1+ja}) = \frac{a}{a+1} \cdot \frac{2a}{2a+1} \cdots \frac{ta}{ta+1}$ as I, we will have $I^a \geq \frac{1}{ta+1}$, so the convergence rate is smaller than $\sqrt[a]{ta} = \Theta(t^k)$
2. Upper bound: consider the following $a-1$ sets of expressions and each set consists of t terms: $\{\frac{a+1}{a+2} \cdot \frac{2a+1}{2a+2} \cdots \cdot \frac{ta+1}{ta+2}\}$, $\{\frac{a+2}{a+3} \cdot \frac{2a+2}{2a+3} \cdots \cdot \frac{ta+2}{ta+3}\}$..., $\{\frac{2a-1}{2a} \cdot \frac{3a-1}{3a} \cdots \cdot \frac{(t+1)a-1}{(t+1)a}\}$. Then each of the $a-1$ expressions are larger than $\prod_{j=1}^{t}(\frac{ja}{1+ja}) = \frac{a}{a+1} \cdot \frac{2a}{2a+1} \cdots \frac{ta}{ta+1}$. Denote $\prod_{j=1}^{t}(\frac{ja}{1+ja}) = \frac{a}{a+1} \cdot \frac{2a}{2a+1} \cdots \frac{ta}{ta+1}$ as I, we will have $I^a \leq \frac{1}{(t+1)}$, so the convergence rate is larger than $\sqrt[a]{ta} = \Theta(t^k)$

Thus, take the limit and apply the Sandwich Theorem, the convergence rate is $\Theta(t^k)$.