Stateful Strategic Regression

Anonymous Authors[1]

Abstract

Automated decision-making tools increasingly assess individuals to determine if they qualify for high-stakes opportunities. A recent line of research investigates how strategic agents may respond to such scoring tools to receive favorable assessments. While prior work has focused on the *short-term* strategic interactions between a decision-making institution (modeled as a principal) and individual decision-subjects (modeled as agents), we investigate interactions spanning *multiple time-steps*. In particular, we consider settings in which the agent's effort investment today can accumulate over time in the form of an internal *state*—impacting both his future rewards and that of the principal. We characterize the Stackelberg equilibrium of the resulting game and provide novel algorithms for computing it. Our analysis reveals several intriguing insights about the role of multiple interactions in shaping the game's outcome: We establish that in our stateful setting, the class of all linear assessment policies remains as powerful as the larger class of all monotonic assessment policies. More importantly, we show that with multiple rounds of interaction at her disposal, the principal is more effective at incentivizing the agent to accumulate effort in her desired direction. Our work addresses several critical gaps in the growing literature on the societal impacts of automated decision-making—by focusing on *longer time horizons* and accounting for the *compounding* nature of decisions individuals receive over time.

1. Introduction

Automated decision-making tools increasingly assess individuals to determine whether they qualify for life-altering opportunities in domains such as lending (12), higher education (15), employment (18), and beyond. These assessment tools have been widely criticized for the blatant disparities they produce through their scores (20, 2). This overwhelming body of evidence has led to a remarkably active area of research into understanding the societal implications of algorithmic/data-driven automation. Much of the existing work on the topic has focused on the *immediate* or *short-term* societal effects of automated decision-making. (For example, a thriving line of work in Machine Learning (ML) addresses the unfairness that arises when ML predictions inform high-stakes decisions (8, 10, 14, 4, 1, 7, 5) by defining it as a form of predictive disparity, e.g., inequality in false-positive rates (10, 2) across social groups.) With the exception of several noteworthy recent articles (which we discuss shortly), prior work has largely ignored the *processes* through which algorithmic decision-making systems can *induce, perpetuate, or amplify* undesirable choices and behaviors.

Our work takes a *long-term perspective* toward modeling the interactions between individual decision subjects and algorithmic assessment tools. We are motivated by two key observations: First, algorithmic assessment tools often provide predictions about the *latent* qualities of interest (e.g., creditworthiness, mastery of course material, or job productivity) by relying on *imperfect* but *observable* proxy attributes that can be directly evaluated about the subject (e.g., past financial transactions, course grades, peer evaluation letters). Moreover, their design ignores the *compounding* nature of advantages/disadvantages individual subjects accumulate over time in pursuit of receiving favorable assessments (e.g., debt, knowledge, job-related skills). To address how individuals *respond* to decisions made about them through modifying their observable characteristics, a growing line of work has recently initiated the study of the *strategic* interactions between decision-makers and decision-subjects (see, e.g., (6, 11, 16, 13, 9)). This existing work has focused mainly on the *short-term* implications of strategic interactions with algorithmic assessment tools—e.g., by modeling it as a *single round* of interaction between a principal (the decision-maker) and agents (the decision-subjects) (13). In addition, existing work that studies interactions over time assume that agents are myopic in responding to the decision-maker's policy (3, 19, 17, 6). We expand the line of inquiry

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

to *multiple rounds* of interactions, accounting for the impact of actions today on the outcomes players can attain tomorrow.

Our multi-round model of principal-agent interactions. We take the model proposed Kleinberg & Raghavan (13) as our starting point. In Kleinberg & Raghavan's formulation, a principal interacts with an agent *once*, where the interaction takes the form of a Stackelberg game. The agent receives a score $y = f(\boldsymbol{\theta}, \mathbf{o})$, in which $\boldsymbol{\theta}$ is the principal's choice of assessment parameters, and $\mathbf{o}$ is the agent's observable characteristics. The score is used to determine the agent's merit with respect to the quality the principal is trying to assess. (As concrete examples, y could correspond to the grade a student receives for a class, or the FICO credit score of a loan applicant.) The principal moves first, publicly announcing her assessment rule $\boldsymbol{\theta}$ used to evaluate the agent. The agent then best responds to this assessment rule by deciding how to invest a *fixed* amount of effort into producing a set of observable features $\mathbf{o}$ that maximize his score y. Kleinberg & Raghavan characterize the assessment rules that can incentivize the agent to invest in specific types of effort (e.g., those that lead to real *improvements* in the quality of interest as opposed to *gaming* the system). We generalize the above setting to $T > 1$ rounds of interactions between the principal and the agent and allow for the possibility of certain effort types rolling over from one step to the next. Our key finding is that longer time horizon provides the principal additional latitude in the range of effort sequences she can incentivize the agent to produce.

To build intuition as to why repeated interactions lead to the expansion of incentivizable efforts, consider the following stylized example:

Example 1.1. *Consider the classroom example of Kleinberg & Raghavan where a teacher (modeled as a principal) assigns a student (modeled as an agent) an overall grade y based on his observable features; in this case test and homework score. Assume that the teacher chooses an assessment rule and assigns a score $y = \theta_{TE}TE + \theta_{HW}HW$, where TE is the student's test score HW is his homework score, and $\theta_T, \theta_{HW} \in \mathbb{R}$ are the weight of each score in the student's overall grade. The student can invest effort into any of three activities: copying answers on the test (CT, improves test score), studying (S, improves both test and homework score), and looking up homework answers online (CH, improves homework score). In a one-round setting where the teacher only evaluates the student once, the student may be more inclined to copy answers on the test or look up homework answers online, since these actions immediately improve the score with relatively lower efforts. However, in a multiple-round setting, these two actions do not improve the student's knowledge (which impacts the student's future grades as well), and so these efforts do not carry over to future time steps. When there are multiple rounds of interaction, the student will be incentivized to invest effort into studying, as knowledge accumulation over time takes less effort in the long-run compared to cheating every time. We revisit this example in further detail in Section 2.*

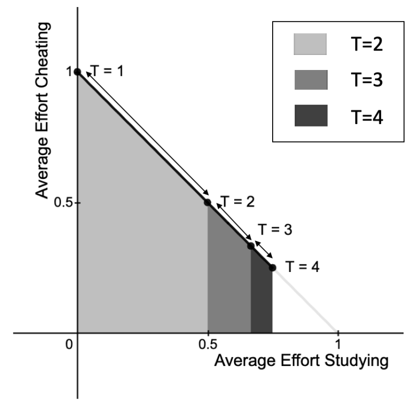


Figure 1: Average effort spent studying vs. average effort spent cheating over time for the example in Appendix A. The line $x + y = 1$ represents the set of all possible Pareto optimal average effort profiles. The shaded region under the line represents the set of average effort profiles which can be incentivized with a certain time horizon. Darker shades represent longer time horizons. In the case where $T = 1$, it is not possible to incentivize the agent to spend any effort studying. Arrows are used to demonstrate the additional set of Pareto optimal average effort profiles that can be incentivized with each time horizon. As the time horizon increases, it becomes possible to incentivize a wider range of effort profiles.

Summary of our findings and techniques. We formalize settings in which the agent's effort investment today can *accumulate* over time in the form of an internal *state*—impacting both his future rewards and that of the principal. We characterize the Stackelberg equilibrium of the resulting game and establish that for the principal, the class of all *linear* assessment policies remains as powerful as the larger class of all *monotonic* assessment policies. In particular, we prove that if there exists an assessment policy that can incentivize the agent to produce a particular sequence of effort profiles, there also exists a linear assessment policy which can incentivize the exact same effort sequence.

Perhaps our most significant finding is that with multiple rounds of assessments at her disposal, the principal is significantly more effective at incentivizing the agent to accumulate effort in her desired direction (as demonstrated in Figure 1 for a simple teacher-student example). In summary, our work addresses two critical gaps in the growing literature on the societal impacts of automated decision-making–by focusing on *longer time horizons* and accounting for the *compounding* nature of decisions individuals receive over time.

2. Problem formulation

In our *stateful* strategic regression setting, a principal interacts with the *same* agent over the course of T time-steps, modeled via a Stackelberg game.[1] The principal moves first, announcing an *assessment policy*, which consists of a *sequence* of assessment rules given by parameters $\{\boldsymbol{\theta}_t\}_{t=1}^T$. Each $\boldsymbol{\theta}_t$ is used for evaluating the agent at round $t = 1, \cdots, T$. The agent then best responds to this assessment rule by investing effort in different activities, which in turn produces a series of observable features $\{\mathbf{o}_t\}_{t=1}^T$ that maximize his overall score. Through each assessment round $t \in \{1, \cdots, T\}$, the agent receives a score $y_t = f(\boldsymbol{\theta}_t, \mathbf{o}_t)$, where $\boldsymbol{\theta}_t$ is the principal's assessment parameters for round t, and $\mathbf{o}_t$ is the agent's observable features at that time. Following Kleinberg & Raghavan, we focus on monotone assessment rules.

Definition 2.1 (Monotone assessment rules)**.** A assessment rule $f(\boldsymbol{\theta}, \cdot) : \mathbb{R}^n \to \mathbb{R}$ is *monotone* if $f(\boldsymbol{\theta}, \mathbf{o}) \geq f(\boldsymbol{\theta}, \mathbf{o}')$ for $o_k \geq o'_k \ \forall k \in \{1, ..., n\}$. Additionally, $\exists k \in \{1, ..., n\}$ such that strictly increasing o_k strictly increases $f(\boldsymbol{\theta}, \boldsymbol{o})$.

For convenience, we assume the principal's assessment rules are linear, that is, $y_t = f(\boldsymbol{\theta}_t, \mathbf{o}_t) = \boldsymbol{\theta}_t^\top \mathbf{o}_t$. Later we show that the linearity assumption is without loss of generality. We also restrict $\boldsymbol{\theta}_t$ to lie in the n-dimensional probability simplex Δ^n. That is, we require each component of $\boldsymbol{\theta}_t$ to be at least 0 and the sum of the n components equal 1.

From effort investments to observable features and internal states. The agent can modify his observable features by investing effort in various activities. While these effort investments are private to the agent and the principal cannot directly observe them, they lead to features that the principal can observe. In response to the principal's assessment policy, The agent plays an *effort policy*, consisting of a *sequence* of effort profiles $\{\mathbf{e}_t\}_{t=1}^T$ where each individual coordinate of $\mathbf{e}_t$ (denoted by $e_{t,j}$) is a function of the principal's assessment policy $\{\boldsymbol{\theta}_t\}_{t=1}^T$. Specifically, the agent chooses his policy $(\mathbf{e}_1, \cdots, \mathbf{e}_T)$, so that it is a best-response to the the principal's assessment policy $(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_T)$.

Next, we specify how effort investment translates into observable features. We assume an agent's observable features in the first round take the form $\mathbf{o}_1 = \mathbf{o}_0 + \boldsymbol{\sigma}_W(\mathbf{e}_1)$, where $\mathbf{o}_0 \in \mathbb{R}^n$ is the initial value of the agent's observable features *before* any modification, $\mathbf{e}_1 \in \mathbb{R}^d$ is the effort the agent expends to modify his features in his first round of interaction with the principal, and $\boldsymbol{\sigma}_W : \mathbb{R}^d \to \mathbb{R}^n$ is the *effort conversion function*, parameterized by W. The effort conversion function is some concave mapping from effort expended to observable features. (For example, if the observable features in the classroom setting are test and homework scores, expending effort studying will affect both an agent's test and homework scores, although it may require more studying to improve test scores from 90% to 100% than from 50% to 60%.)

Over time, effort investment can accumulate. (For example, small businesses accumulate *wealth* over time by following good business practices. Students *learn* as they study and accumulate *knowledge*.) This accumulation takes the form of an internal *state*, which has the form $\mathbf{s}_t = \mathbf{s}_0 + \Omega \sum_{i=1}^{t-1} \mathbf{e}_i$. Here $\Omega \in \mathbb{R}^{d \times d}$ is a *diagonal* matrix in which $\Omega_{j,j}$, $j \in \{1, \ldots, d\}$ determines how much one unit of effort (e.g., in the jth effort coordinate, e_j) rolls over from one time step to the next, and $\mathbf{s}_0$ is the agent's initial "internal state". An agent's observable features are, therefore, a function of both the effort he expends, as well as his internal state. Specifically, $\mathbf{o}_t = \boldsymbol{\sigma}_W(\mathbf{s}_t + \mathbf{e}_t)$ (here $\boldsymbol{\sigma}_W(\mathbf{s}_0)$ is analogous to $\mathbf{o}_0$ in the single-shot setting).

Utility functions for the agent and the principal. Given the above mapping, the agent's goal is to pick his effort profiles so that the observable features they produce maximize the *sum* of his scores over time, that is, the agent's utility $= \sum_{t=1}^T y_t = \sum_{t=1}^T \boldsymbol{\theta}_t^\top \mathbf{o}_t$. Our focus on the sum of scores over time is a conventional choice and is motivated by real-world examples. (A small business owner who applies for multiple loans cares about the cumulative amount of loans he/she receives. A student taking a series of exams cares about his/her average score across all of them.)

The principal's goal is to choose his assessment rules over time so as to maximize cumulative effort investments according to her preferences captured by a matrix Λ. Specifically, the principal's utility $= \left\|\Lambda \sum_{t=1}^T \mathbf{e}_t\right\|_1$. The principal's utility can be thought of as a weighted $L1$ norm of the agent's cumulative effort, where $\Lambda \in \mathbb{R}^{d \times d}$ is a *diagonal* matrix where the element Λ_{jj} denotes how much the principal wants to incentivize the agent invest in effort component e_j.[2]

Constraints on agent effort. As was the case in the single-shot setting of Kleinberg & Raghavan, we assume that the agent's choice of effort $\mathbf{e}_t$ at each time t is subject to a fixed budget B. Without loss of generality, we consider the case

[1] To improve readability, we adopt the convention of referring to the principal as she/her and the agent as he/him throughout the paper.

[2] Note that while we only consider diagonal $\Omega \in \mathbb{R}_+^{d \times d}$, our results readily extend to general $\Omega, \in \mathbb{R}_+^{d \times d}$. By focusing on diagonal matrices we have a one-to-one mapping between state and effort components. Non-diagonal Ω corresponds to cases where different effort components can contribute to multiple state components.

where $B = 1$.

Proposition 2.2. *It is possible to incentivize a wider range of effort profiles by modeling the principal-agent interaction over* multiple *time-steps, compared to a model which only considers one-shot interactions.*

See Appendix A for an example which illustrates this phenomena.

3. Equilibrium characterization

The following optimization problem captures the expression for the agent's best-response to an arbitrary sequence of assessment rules.[3] (Recall that d refers to the dimension of effort vectors ($\mathbf{e}_t$'s), and n refers to the number of observable features, i.e., the dimension of $\mathbf{o}_t$'s.)

The set of agent best-responses to a linear assessment policy, $\{\boldsymbol{\theta}_t\}_{t=1}^T$, is given by the following optimization procedure:

$$\{\mathbf{e}_t^*\}_{t=1}^T = \arg\max_{\mathbf{e}_1,\ldots,\mathbf{e}_T} \sum_{t=1}^T \boldsymbol{\theta}_t^\top \boldsymbol{\sigma}_W \left(\mathbf{s}_0 + \Omega \sum_{i=1}^{t-1} \mathbf{e}_i + \mathbf{e}_t \right),$$
$$\text{s.t.} \quad e_{t,j} \geq 0, \quad \sum_{j=1}^d e_{t,j} \leq 1 \; \forall t, j$$

The goal of the principal is to pick an assessment policy $\{\boldsymbol{\theta}\}_{t=1}^T$ in order to maximize the total magnitude of the effort components she cares about, i.e.

$$\{\boldsymbol{\theta}_t^*\}_{t=1}^T = \arg\max_{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_T} \left\| \Lambda \sum_{t=1}^T \mathbf{e}_t(\boldsymbol{\theta}_t, \ldots, \boldsymbol{\theta}_T) \right\|_1,$$
$$\text{s.t.} \quad \boldsymbol{\theta}_t \in \Delta^n \; \forall t$$

Substituting the agent's optimal effort policy into the above expression, we obtain the following formalization of the principal's assessment policy:

Proposition 3.1 (Stackelberg Equilibrium)**.** *Suppose the principal's strategy space consists of all sequences of linear monotonic assessment rules. The Stackelberg equilibrium of the stateful strategic regression game, $\left(\{\boldsymbol{\theta}_t^*\}_{t=1}^T, \{\mathbf{e}_t^*\}_{t=1}^T\right)$, can be specified as the following bilevel multiobjective optimization problem. Moving forward, we omit the constraints on the agent and principal action space for brevity.*

$$\{\boldsymbol{\theta}_t^*\}_{t=1}^T = \arg\max_{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_T} \left\| \Lambda \sum_{t=1}^T \mathbf{e}_t^*(\boldsymbol{\theta}_t, \ldots, \boldsymbol{\theta}_T) \right\|_1,$$
$$\textit{s.t.} \quad \{\mathbf{e}_t^*\}_{t=1}^T = \arg\max_{\mathbf{e}_1,\ldots,\mathbf{e}_T} \sum_{t=1}^T \boldsymbol{\theta}_t^\top \boldsymbol{\sigma}_W \left(\mathbf{s}_0 + \Omega \sum_{i=1}^{t-1} \mathbf{e}_i + \mathbf{e}_t \right)$$

[3]Throughout this section when it improves readability, we denote the dimension of matrices in their subscript (e.g., $X_{a \times b}$ means X is an $a \times b$ matrix).

3.1. Linear assessment policies are optimal

Throughout our formalization of the Stackelberg equilibrium, we have assumed that the principal deploys *linear* assessment rules, when *a priori* it is not obvious why the principal would play assessment rules of this form. We now show that the linear assessment policy assumption is without loss of generality.

We start by defining the concept of *incentivizability* for an effort policy, and characterize it through a notion of a *dominated effort policy*.

Definition 3.2 (Incentivizability)**.** An effort policy $\{\mathbf{e}_t\}_{t=1}^T$ is *incentivizable* if there exists an assessment policy $\{f(\boldsymbol{\theta}_t, \cdot)\}_{t=1}^T$ for which playing $\{\mathbf{e}_t\}_{t=1}^T$ is *a* best response. (Note: $\{\mathbf{e}_t\}_{t=1}^T$ need not be the *only* best response.)

Definition 3.3 (Dominated Effort Policy)**.** We say the effort policy $\{\mathbf{e}_t\}_{t=1}^T$ is *dominated by* another effort policy if an agent can achieve the same or higher observable feature values by playing another effort policy $\{\mathbf{a}_t\}_{t=1}^T$ that does not spend the full effort budget on at least one time-step.

Note that an effort policy which is dominated by another effort policy will never be played by a rational agent no matter what set of decision rules are deployed by the principal, since a better outcome for the agent will always be achievable.

Theorem 3.4. *For any effort policy $\{\mathbf{e}_t\}_{t=1}^T$ that is not dominated by another effort policy, there exists a linear assessment policy that can incentivize it.*

See Appendix C for the complete proof. We characterize whether an effort *policy* $\{\mathbf{e}_t\}_{t=1}^T$ is dominated or not by a linear program, and show that a *subset* of the dual variables correspond to a linear assessment policy which can incentivize it. Kleinberg & Raghavan present a similar proof for their setting, defining a linear program to characterize whether an effort *profile* $\mathbf{e}_t$ is dominated or not. They then show that if an effort profile is *not* dominated, the dual variables of their linear program correspond to a linear assessment rule which can incentivize it. While the proof idea is similar, their results do not extend to our setting because our linear program must include an additional constraint for every time-step to ensure that the budget constraint is always satisfied. We show that by examining the complementary slackness condition, we can upper-bound the gradient of the agent's cumulative score with respect to a subset of the dual variables $\{\boldsymbol{\lambda}_t\}_{t=1}^T$ (where each upper bound depends on the "extra" term γ_t introduced by the linear budget constraint for that time-step). Finally, we show that when an effort policy is not dominated, all of these bounds hold with equality and, because of this, the subset of dual variables $\{\boldsymbol{\lambda}_t\}_{t=1}^T$ satisfy the definition of a linear assessment policy which can incentivize the effort policy $\{\mathbf{e}_t\}_{t=1}^T$.

References

[1] Agarwal, A., Beygelzimer, A., Dudík, M., Langford, J., and Wallach, H. A reductions approach to fair classification. *arXiv preprint arXiv:1803.02453*, 2018.

[2] Angwin, J., Larson, J., Mattu, S., and Kirchner, L. Machine bias. *Propublica*, 2016.

[3] Bechavod, Y., Ligett, K., Wu, Z. S., and Ziani, J. Gaming helps! learning from strategic interactions in natural dynamics. *arXiv preprint arXiv:2002.07024*, 2020.

[4] Celis, L. E., Huang, L., Keswani, V., and Vishnoi, N. K. Classification with fairness constraints: A meta-algorithm with provable guarantees. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 319–328. ACM, 2019.

[5] Corbett-Davies, S. and Goel, S. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*, 2018.

[6] Dong, J., Roth, A., Schutzman, Z., Waggoner, B., and Wu, Z. S. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 55–70, 2018.

[7] Donini, M., Oneto, L., Ben-David, S., Shawe-Taylor, J., and Pontil, M. Empirical risk minimization under fairness constraints. *arXiv preprint arXiv:1802.08626*, 2018.

[8] Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pp. 214–226. ACM, 2012.

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

[10] Hardt, M., Price, E., Srebro, N., et al. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*, pp. 3315–3323, 2016.

[11] Hu, L., Immorlica, N., and Vaughan, J. W. The disparate effects of strategic manipulation. In *Proceedings of the 2nd ACM Conference on Fairness, Accountability, and Transparency*, 2019.

[12] Jagtiani, J. and Lemieux, C. The roles of alternative data and machine learning in fintech lending: evidence from the lendingclub consumer platform. *Financial Management*, 48(4):1009–1029, 2019.

[13] Kleinberg, J. and Raghavan, M. How do classifiers induce agents to invest effort strategically? *ACM Transactions on Economics and Computation (TEAC)*, 8(4):1–23, 2020.

[14] Kleinberg, J., Mullainathan, S., and Raghavan, M. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016.

[15] Kučak, D., Juričić, V., and ambić, G. Machine learning in education - a survey of current research trends. *Annals of DAAAM & Proceedings*, 29, 2018.

[16] Milli, S., Miller, J., Dragan, A. D., and Hardt, M. The social cost of strategic classification. In *Proceedings of the 2nd ACM Conference on Fairness, Accountability, and Transparency*, 2019.

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

[18] Sánchez-Monedero, J., Dencik, L., and Edwards, L. What does it mean to'solve'the problem of discrimination in hiring? social, technical and legal perspectives from the uk on automated hiring systems. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 458–468, 2020.

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

[20] Sweeney, L. Discrimination in online ad delivery. *Queue*, 11(3):10, 2013.

A. Formalizing the classroom example

Example A.1. *We demonstrate this by revisiting the classroom example. Recall that a teacher assigns a student an overall grade $y = \theta_{TE} TE + \theta_{HW} HW$, where TE is the student's test score HW is their homework score, and θ_{TE} & θ_{HW} are the weight of each score in the student's overall grade. The student can invest effort into any of three activities: copying answers on the test (CT, improves test score), studying (S, improves both test and homework score), and looking up homework answers online (CH, improves homework score). Suppose the relationship between observable features and effort $\mathbf{e}$ the agent chooses to spend is defined by the equations*

$$TE = TE_0 + W_{CT} CT + W_{ST} S$$

$$HW = HW_0 + W_{SH} S + W_{CH} CH$$

where TE_0 and HW_0 are the test and homework scores the student would receive if they did not expend any effort. If $W_{CT} = W_{CH} = 3$ and $W_{ST} = W_{SH} = 1$, there is no combination of θ_{TE}, θ_{HW} values the teacher can deploy to incentivize the student to study, because the benefit of cheating is just too great. (See (13) for more detail.)

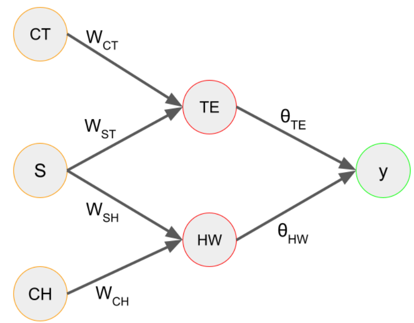
Now consider a multi-step interaction between a teacher and student in which effort invested in studying carries over to future time-steps in the form of knowledge accumulation. The relationships between observable features and effort expended are now defined as

$$TE_t = TE_0 + W_{CT} CT_t + W_{ST} s_t$$

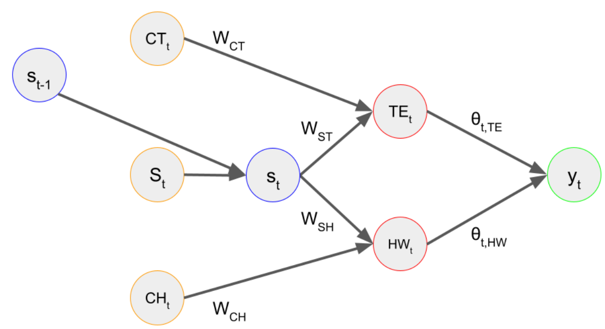
and

$$HW_t = HW_0 + W_{SH} s_t + W_{CH} CH_t$$

where $s_t = \sum_{i=1}^{t} S_i$ is the agent's internal knowledge state. Instead of assigning students a single score y_1, the teacher assigns the student a score y_t at each round by picking $\theta_{t,T}, \theta_{t,HW}$ at every time-step. The student's grade is then the summation of all scores across time. Suppose $T \geq 3$, where T is the number of rounds of interaction. Consider $W_{CT} = W_{CH} = 3$, $W_{ST} = W_{SH} = 1$, and $TE_0 = HW_0 = 0$. Unlike in the single-round setting, it is easy to verify that students can now be incentivized to study by picking $\theta_{t,TE} = \theta_{t,HW} = 0.5 \; \forall t$.



(a) Single-step classroom setting.



(b) Multi-step classroom setting.

Figure 2: Comparison between the single-step and multi-step scenarios in the hypothetical classroom setting. The single-step formulation does not account for changes in the student's internal state over time. In the multi-step formulation, effort put towards studying accumulates in the form of knowledge. Modeling this effort accumulation allows the teacher to incentivize the student to study across a wider range of parameter values. The agent can invest effort in 3 actions: cheating on the test (CT), studying (S), and cheating on the homework (CH). W values denote how much one unit of effort translates to the two observable features, test score (T) and homework score (HW). The student's score (y_t) at each time-step is a weighted average of these two observable features. In the multi-step setting, s_t denotes the student's internal knowledge state at time t.

B. Equilibrium derivations

B.1. Agent's best-response effort sequence

A rational agent solves the following optimization to determine his best-response effort policy:

$$\{\mathbf{e}_t^*\}_{t=1}^T = \arg\max_{\mathbf{e}_1,\ldots,\mathbf{e}_T} \sum_{t=1}^{T} (y_t = f_t(\mathbf{e}_1,\ldots,\mathbf{e}_t))$$
$$\text{s.t.} \quad e_{t,j} \geq 0 \; \forall t,j, \quad \sum_{j=1}^{d} e_{t,j} \leq 1 \; \forall t$$

Recall that the agent's score y_t at each time-step is a function of $(\mathbf{e}_1,\ldots,\mathbf{e}_t)$, the sequence of effort expended by the agent so far. Replacing the score y_t and observable features $\mathbf{o}_t$ with their respective equations, we obtain the expression

$$\{\mathbf{e}_t^*\}_{t=1}^T = \arg\max_{\mathbf{e}_1,\ldots,\mathbf{e}_T} \sum_{t=1}^{T} \boldsymbol{\theta}_t^\top \boldsymbol{\sigma}_W (\mathbf{s}_t + \mathbf{e}_t)$$
$$\text{s.t.} \quad e_{t,j} \geq 0 \; \forall t,j, \quad \sum_{j=1}^{d} e_{t,j} \leq 1 \; \forall t$$

where the agent's internal state $\mathbf{s}_t$ at time t is a function of the effort he expends from time 1 to time $t-1$. Replacing $\mathbf{s}_t$ with the expression for agent state, we get

$$\{\mathbf{e}_t^*\}_{t=1}^T = \arg\max_{\mathbf{e}_1,\ldots,\mathbf{e}_T} \sum_{t=1}^{T} \boldsymbol{\theta}_t^\top \boldsymbol{\sigma}_W \left(\mathbf{s}_0 + \Omega \sum_{i=1}^{t-1} \mathbf{e}_i + \mathbf{e}_t\right)$$
$$\text{s.t.} \quad e_{t,j} \geq 0 \; \forall t,j, \quad \sum_{j=1}^{d} e_{t,j} \leq 1 \; \forall t$$

C. Proof of Theorem 3.4

Proof. Let κ be the optimal value of the following linear program:

$$\begin{aligned} V(\{\mathbf{e}_t\}_{t=1}^T) = & \min_{\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_T} \sum_{t=1}^{T} \|\mathbf{a}_t\|_1 \\ \text{s.t.} \quad & W\left(\Omega \sum_{i=1}^{t-1} \mathbf{a}_i + \mathbf{a}_t\right) \geq W\left(\Omega \sum_{i=1}^{t-1} \mathbf{e}_i + \mathbf{e}_t\right), \\ & \mathbf{a}_t \geq \mathbf{0}_d, \; \|\mathbf{a}_t\|_1 \leq 1, \; \forall t \end{aligned} \tag{1}$$

Optimization 1 can be thought of as trying to minimize the total effort $\{\mathbf{a}_t\}_{t=1}^T$ the agent spends across all T time-steps, while achieving the same or greater feature values at every time t compared to $\{\mathbf{e}_t\}_{t=1}^T$. Let $\{\mathbf{a}_t^*\}_{t=1}^T$ denote the set of optimal effort profiles for Optimization 1. If $\{\mathbf{e}_t\}_{t=1}^T \in \{\mathbf{a}_t^*\}_{t=1}^T$, a value of $\kappa = T$ is obtained. A dominated effort policy is formally defined as follows:

Lemma C.1 (Dominated Effort Policy)**.** *An effort policy $\{\mathbf{e}_t\}_{t=1}^T$ is dominated by another effort policy if $\kappa < T$.*

The Lagrangian of Optimization 1 can be written as

$$L = \sum_{t=1}^{T} \|\mathbf{a}_t\|_1 + \sum_{t=1}^{T} \boldsymbol{\lambda}_t^\top W \left(\Omega \sum_{i=1}^{t-1} (\mathbf{e}_i - \mathbf{a}_i) + \mathbf{e}_t - \mathbf{a}_t\right) + \gamma_t (\|\mathbf{a}_t\|_1 - 1) - \boldsymbol{\mu}_t^\top \mathbf{a}_t,$$

where $\boldsymbol{\lambda}_t \geq \mathbf{0}_n, \; \boldsymbol{\mu}_t \geq \mathbf{0}_d, \; \forall t$.

In order for stationarity to hold, $\nabla_{\mathbf{a}_t} L(\mathbf{a}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{\gamma}^*) = \mathbf{0}_d \; \forall t$, where $\mathbf{x}^*$ denotes the optimal values for variable $\mathbf{x}$. Applying the stationarity condition to Lagrangian function, we obtain

$$\mathbf{1}_d - W^\top \boldsymbol{\lambda}_t^* - \sum_{i=t+1}^{T} \Omega^\top W^\top \boldsymbol{\lambda}_i^* + \gamma_t^* \cdot \mathbf{1}_d - \boldsymbol{\mu}_t^* = \mathbf{0}_d, \; \forall t \tag{2}$$

Because of dual feasibility, $\boldsymbol{\mu}_t \geq \mathbf{0}_d \; \forall t$. By rearranging Equation 2 and using this fact, we can obtain the following bound on $W^\top \boldsymbol{\lambda}_t^* + \sum_{t=i+1}^{T} \Omega^\top W^\top \boldsymbol{\lambda}_t^*$:

$$W^\top \boldsymbol{\lambda}_t^* + \sum_{i=t+1}^{T} \Omega^\top W^\top \boldsymbol{\lambda}_i^* \leq (1 + \gamma_t^*) \cdot \mathbf{1}_d, \; \forall t \tag{3}$$

Next we look at the complementary slackness condition. For complementary slackness to hold, $\boldsymbol{\mu}_t^{*\top} \mathbf{a}_t^* = 0 \; \forall t$. If $\kappa = T$, then $\{\mathbf{e}_t\}_{t=1}^T \in \{\mathbf{a}_t^*\}_{t=1}^T$ and therefore $\{\mathbf{e}_t\}_{t=1}^T$ is not dominated. If $\{\mathbf{e}_t\}_{t=1}^T$ is not dominated, $\boldsymbol{\mu}_t^{*\top} \mathbf{e}_t = 0 \; \forall t$. This means that if $e_{t,j} > 0$, $\mu_{t,j} = 0, \; \forall t, j$. This, along with Equation 2, implies that

$$\left[W^\top \boldsymbol{\lambda}_t^* + \sum_{i=t+1}^{T} \Omega^\top W^\top \boldsymbol{\lambda}_i^*\right]_j = 1 + \gamma_t^*$$

for all t, j where $e_{t,j} > 0$.

Switching gears, consider the set of *linear* assessment policies $\mathcal{L}$ for which $\{\mathbf{e}_t\}_{t=1}^T$ is incentivizable. The set of linear assessment policies for which $\{\mathbf{e}_t\}_{t=1}^T$ is incentivizable is the set of linear assessment policies for which the derivative of the total score with respect to the agent's effort policy is maximal at the coordinates which $\{\mathbf{e}_t\}_{t=1}^T$ has support on. Denote this set of coordinates as S, and the set of coordinates which $\mathbf{e}_t$ has support on as S_t. Formally,

$$\mathcal{L} = \Bigg\{ \{\boldsymbol{\theta}_t\}_{t=1}^T \Bigg| \left[\nabla_{\mathbf{a}_t} \sum_{i=1}^T \left(y_i = f\left(\{\mathbf{a}_t\}_{t=1}^T, \{\boldsymbol{\theta}_t\}_{t=1}^T \right) \right) \right]_{S_t} = \max_j \left(\nabla_{\mathbf{a}_t} \sum_{i=1}^T y_i \right) \cdot \mathbf{1}_{|S_t|}, \ \forall t \Bigg\}$$

Recall that $\sum_{t=1}^T y_t = \sum_{t=1}^T \boldsymbol{\theta}_t^\top W \left(\mathbf{s}_0 + \Omega \sum_{i=1}^{t-1} \mathbf{a}_i + \mathbf{a}_t \right)$. Therefore, the gradient of $\sum_{t=1}^T y_t$ with respect to $\mathbf{a}_t$ can be written as

$$\nabla_{\mathbf{a}_t} \sum_{t=1}^T y_t = W^\top \boldsymbol{\theta}_t + \sum_{i=t+1}^T \Omega^\top W^\top \boldsymbol{\theta}_i, \ \forall t$$

Note that the form of $\nabla_{\mathbf{a}_t} \sum_{t=1}^T y_t$ is the same as the LHS of Equation 3. We know that if $\{\mathbf{e}_t\}_{t=1}^T \in \{\mathbf{a}_t^*\}_{t=1}^T$ is incentivizable, the inequality in Equation 3 will hold with equality for all coordinates for which $\{\mathbf{e}_t\}_{t=1}^T$ has positive support. Therefore, the derivative is maximal at those coordinates since it is bounded to be *at most* $1 + \gamma_t^*, \forall t$ (due to the KKT conditions for the dominated effort policy linear program). Because of this, $\{\boldsymbol{\lambda}_t^*\}_{t=1}^T$ is in $\mathcal{L}$, which means that $\{\mathbf{e}_t\}_{t=1}^T$ can be incentivized using a linear mechanism.

□