Contents

1 Introduction **1**
- 1.1 Our Model and Contribution . . . 2
- 1.2 Illustrative Example . . . 3
- 1.3 Summary of Results and Roadmap . . . 3

2 Model **4**
- 2.1 Repeated Contract Setting: Discrete Version . . . 5
- 2.2 Repeated Contract Setting: Continuous Version . . . 6

3 Linear Contracts **8**
- 3.1 Free-Fall Contracts are Optimal Linear Contracts . . . 8
- 3.2 Implication to Welfare and Agent's Utility . . . 8
- 3.3 General Contracts and Free-Fall . . . 9

4 Unknown Time Horizon **9**

A Further Related Work **14**

B Proof of Theorem 3.1 (Optimal Dynamic Linear Contract) **14**

C Proof of Theorem 3.2 (Win-Win) **19**
- C.1 Proof of Claim C.1 . . . 20

D General Contracts with Single-Dimensional Scaling **22**

E Proof of Thorem 4.2 (Unknown Time Horizon) **25**

F Proof of Theorem 4.3 (Unknown Time Horizon – Converse) **29**

G General Contracts **30**
- G.1 Programmatic LP Search for Sub-Optimal Free Fall Against Non-Linear Contracts . 31

H Simplifying Tool: Reductions from Discrete to Continuous Time **31**
- H.1 Proof of Theorem 2.4 . . . 31
- H.2 Reduction from Discrete to Continuous Time with Unknown Time Horizons 34
- H.3 Mean-Based Algorithms in the Partial-Feedback Setting . . . 35

I Final Remarks **36**

J Conclusion **37**

A Further Related Work

Introduction to the Problem. Consider the following motivating examples: a worker joining a new team, a student starting an internship, or a junior professor joining a committee. These agents initially face uncertainty about the required effort and what constitutes good performance. They must decide when to exert more effort and when to reduce it. Peer assessments introduce additional uncertainty and noise. Furthermore, the environment is dynamic, with the value of certain outcomes changing over time. Each agent encounters an implicit and evolving system of incentives that they must adapt to through repeated interactions. This pattern is prevalent in many real-life contractual relationships and is increasingly relevant to AI agents handling complex, open-ended, and computationally intensive tasks. For details to the aforementioned examples, the interested reader can see [39] and references therein. In settings like credit scoring, the evaluation system creates incentives for the agent while remaining opaque to prevent gaming, forcing the agent to act under uncertainty [36].

Simplifying Contracts. Given this complexity, one line of work focuses on identifying settings where simple contracts suffice. Notably, [46] assume constant absolute risk aversion (CARA) utilities and Brownian motion of the output, examining a single payment at the end of the contractual relationship based on all outcomes. Another approach involves deliberately vague contracts, leaving agents uncertain about performance-based compensation (e.g., [3, 10, 30]). [42] explore how to learn an agent's private type through online principal-agent interaction and contract menus. [9] study principal-agent problems over MDPs, where a budgeted principal offers additional rewards, and the agent selects the MDP policy selfishly, without learning. Thus, a naturally arising question is:

How should an agent choose their actions in a contractual relationship involving uncertainty and recurrent interactions?

Our algorithmic perspective introduces a novel, learning-based approach to address the complexity of repeated contracts, leveraging no-regret and general mean-based agents. Below, we discuss why learning methods are natural choices for agents' responses in the context of existing literature.

Optimizing Against No-Regret Learners. From an econometric perspective, agents often respond to repeated strategic interactions in auctions in ways consistent with no-regret learning [57, 58]. Inspired by these findings, [14] explore algorithmic mechanism design, demonstrating that no-regret learning methods are natural responses for agents. No-regret learning has been extensively studied in repeated games (e.g.,[11, 15, 34, 44, 55, 62, 2, 2, 51, 65, 63, 38, 37]), auctions and economic interactions (e.g.,[24, 18, 50]), and Stackelberg security games (e.g., [8]). For a comprehensive overview, see [61]. By assuming agents employ no-regret learning instead of complex strategic reasoning, we propose a new approach to repeated contracting.

Optimizing Against Mean-Based Learners. Finding an optimal dynamic strategy against a mean-based learner in general games remains an open problem. [26] show an equivalence between this problem and an n-dimensional control problem, where n is the number of actions available to the agent. Non-trivial optimization against a mean-based learner has been achieved only in repeated auction settings, where [14] demonstrate that the designer can extract full welfare as revenue. [25, 17] extend this to prior-free auction settings and multiple agents. However, even for a single agent, the optimal auction strategy, involving alternating between second-price auctions and charging large payments, is impractical and not intended to guide practice [17]. [18] study mechanisms for no-regret agents, incorporating principal learning to avoid common prior assumptions in economic design problems.

B Proof of Theorem 3.1 (Optimal Dynamic Linear Contract)

Proof overview. We will present a series of "rewriting" rules, which will allow us to replace a given dynamic contract π with a simpler, more constrained, dynamic contract π' with utility at least as large as π. At the conclusion of our sequence of rewriting steps, we will see that our contract takes the form of a free-fall contract, thus implying that there is an optimal free-fall contract.

We begin not with a rewriting rule, but instead a general observation about the structure of dynamic linear contracts — namely, that it is impossible for an agent to "skip over" an action. That is, if an agent is playing action i at some point, and action j at some later point, there must exist segments of non-zero duration where the agent plays each of the intermediate actions between i and j. Formally, we can write this as follows.

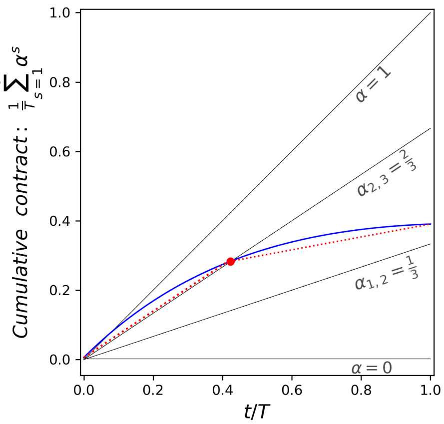


Figure 3: An illustration of Lemma B.2. The plot shows the cumulative contract over time for the contract game shown in Figure 1, repeated to T steps, where both axes are normalized by T. The lemma shows how arbitrary dynamic strategies, as the one shown in the blue curve, can be re-written as piecewise stationary strategies, depicted in the dotted red curve, inducing similar behavior by the agent and the same utilities.

Lemma B.1. *If $\pi = \{(\alpha^k, \tau^k, a^k)\}$ is a dynamic linear contract, then for any k, $|a^k - a^{k+1}| \leq 1$ (i.e., in any two consecutive segments, the learner's action can change by at most one).*

Proof. Note that a^k and a^{k+1} must both be best responses to the average historical contract $\overline{\alpha}^k$, i.e., $a^k, a^{k+1} \in \mathsf{BR}(\overline{\alpha}^k)$. Since $\mathsf{BR}(\alpha)$ is always of the form $\{i\}$ or $\{i, i+1\}$, the conclusion follows. $\square$

Note that the proof of Lemma B.1 relies on the "linear topology" of the best-response regions in Figure 2 (i.e., any non-zero boundary between best-response regions connects two consecutive actions of the agent). This property is *not* true for general contracts or general games; however, we will later see that Lemma B.1 also holds for the class of $\mathbf{p}$-scaled contracts introduced in Appendix D.

We now proceed to introduce our rewriting rules. The first rewriting rule we present is very general (and in fact applies to any game): we will show that without loss of generality, no two consecutive segments of a dynamic contract induce the same action for the learner. Intuitively, for any time interval in which a mean-based agent plays a single action, we can replace the contracts in this interval with their average and obtain overall a revenue-equivalent dynamic contract. Formally, we can phrase this as follows.

Lemma B.2. *Let π be any linear dynamic contract. Then there exists a linear dynamic contract π' such that $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$ and no two consecutive segments of π' share the same agent action ($a^k \neq a^{k+1}$).*

Proof. Let $\pi = \{(\alpha^k, \tau^k, a^k)\}_{k=1}^K$ be any linear dynamic contract. Assume that for some k, $a^k = a^{k+1} = a$. Then the linear dynamic contract π' formed by replacing the two segments (α^k, τ^k, a) and $(\alpha^{k+1}, \tau^{k+1}, a)$ with the single segment $((\alpha^k \tau^k + \alpha^{k+1}\tau^{k+1})/(\tau^k + \tau^{k+1}), \tau^k + \tau^{k+1}, a)$ has the property that $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$. To see this, observe that the same action a is played throughout the entire interval, and the average payout to the agent is the same. Therefore, in fact we have $\mathsf{Util}(\pi') = \mathsf{Util}(\pi)$. It only remains to confirm that this is still a valid dynamic contract (i.e., that each prescribed action is still a best response in the corresponding segment).

To see this, observe first that the cumulative contract at the start (respectively, end) of the merged segment in π' is the same as the start of segment k (respectively, end of segment $k+1$) in π. Therefore, all segments before and after the merged segment are still correct. To confirm the merged segment, we need only confirm that a is a best response on the merged segment in π' using the fact that it was a best response in both segments k and $k+1$ in π.

For this, let α_0 denote the cumulative contract after the first $k-1$ segments, α_1 denote the cumulative contract after the first $k+1$ segments, $\alpha'(t)$ denote the cumulative contract of π' during a time t in the merged segment, and $\alpha(t)$ denote the cumulative contract of π during a time t in segments k or $k+1$. Observe first that for every x between α_0 and α_1, there is some t such that $\alpha(t) = x$. Because a is a best response on the entire segments k and $k+1$, this means that a is a best response to x for all x between α_0 and α_1. Moreover, observe that $\alpha'(t)$ lies between α_0 and α_1 for all t. Therefore, a is indeed a best response to $\alpha'(t)$ for all t in the merged segment, and the dynamic contract is valid.

By repeatedly applying this merging of segments, we can obtain a linear dynamic contract π' satisfying the constraints of the lemma. $\square$

Figure 3 illustrates the above lemma graphically. The figure displays the cumulative contract over time for the contract game depicted in Figure 1. The blue curve represents the trajectory of an arbitrary dynamic contract strategy under which the agent's best response is to take action 3 until time $t/T \approx 0.425$, and then take action 2 in the remaining time. The crossing point between the best response regions is marked with a red dot. Lemma B.2 demonstrates that we can replace the blue trajectory with the simpler trajectory depicted in red. In this red trajectory, every region between two consecutive α values is crossed by a single linear segment (i.e., a piecewise-stationary trajectory), resulting in the same behavior by the agent and the same revenue.

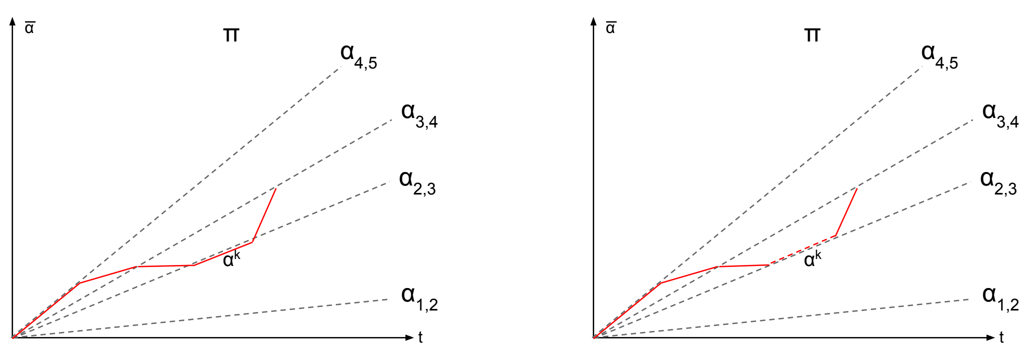
Our second rewriting rule is specific to linear contracts. It shows that for every linear contract in which the agent is indifferent between two actions, it is beneficial for the principal to shift the contract infinitesimally so that the agent prefers the action with the higher expected reward.

Lemma B.3. *Let $\pi = \{(\alpha^k, \tau^k, a^k)\}_{k=1}^K$ be a dynamic linear contract where during segment k the agent is indifferent between actions i and $i+1$ (i.e., $\mathsf{BR}(\overline{\alpha}^{k-1}) \cap \mathsf{BR}(\overline{\alpha}^k) \supseteq \{i, i+1\}$), but $a^k = i$. If we form π' by replacing a^k with $i+1$, then $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$ (the principal always prefers that the agent plays the action with higher expected reward).*

Proof. Since actions in the linear contract setting are sorted by increasing value of expected reward, we have that $\mathsf{Util}(\pi') - \mathsf{Util}(\pi) = \frac{\tau^k}{T^K}\left(u_P(p^k, i+1) - u_P(p^k, i)\right) = \frac{\tau^k}{T^K}\left(R_{i+1} - R_i)(1-\alpha^k\right) \geq 0$. $\square$

Note that the principal can implement the change in the agent's action in Lemma B.3 by simply increasing their payment to the agent by an arbitrarily small amount – this incentivizes the agent to break ties in favor of the action with larger expected reward (which is the action labeled with a larger number). The fact that the principal can implement this change also follows as a direct consequence of the discrete-to-continuous reduction of Theorem 2.4.

By applying the above two rewriting rules (Lemmas B.2 and B.3) along with our observation in Lemma B.1, we can establish our third rewriting rule: it is always possible to rewrite a dynamic contract so that the sequence of actions is a consecutively decreasing sequence.



(a) Base policy π. In this example, the kth segment is the first segment where the action increases immediately after the segment ($a_k = 2$, $a_{k+1} = a_k + 1 = 3$).

(b) Since $a_{k-1} = a_{k+1}$, the kth lies along the best response boundary $\alpha_{2,3}$, and its existence violates Lemma B.3 (we can rewrite it as a segment with $a_k = 3$, and then further collapse these segments via Lemma B.2).

Figure 4: Illustrations for the proof of Lemma B.4.

Lemma B.4. *Let π be any dynamic linear contract. Then there exists a dynamic linear contract $\pi' = \{(\alpha^k, \tau^k, a^k)\}_{k=1}^K$ with $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$ and where $a^1, a^2, \dots, a^K$ is a decreasing sequence of consecutive actions (i.e., $a^k = a^1 - (k-1)$).*

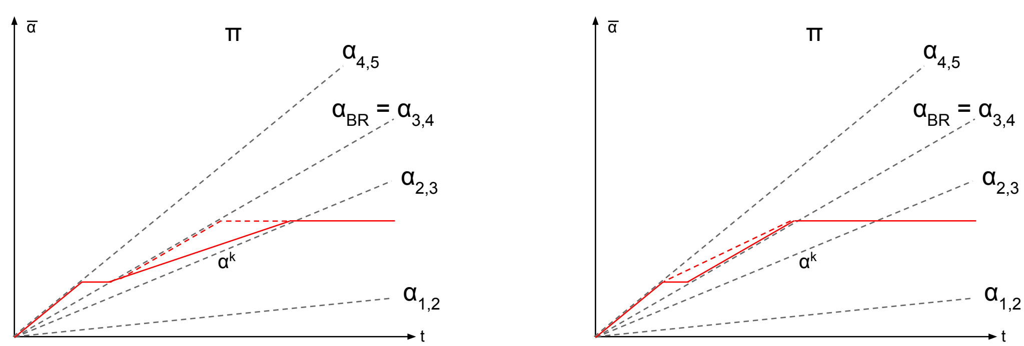
Proof. Apply the two rewriting rules in Lemmas B.2 and B.3 to π until it satisfies the post-conditions of both lemmas (so, no two consecutive segments incentivize the same action, and any segment on a best-response boundary incentivizes the higher-reward action). Since Lemma B.1 implies that consecutive segments cannot skip over an action, this means that every two consecutive actions under π are consecutive: either the agent switches to the next higher action or the next lower action each time step. We therefore just must show that any dynamic contract where the agent increases their action can be rewritten as a decreasing contract with at least same payoff.

Consider the first segment in π where the agent switches to a larger action, that is, the smallest k such that $a^{k+1} = a^k + 1$. Let $a^k = j$ (so $a^{k+1} = j+1$). Note that the agent must be indifferent between actions j and $j+1$ at the end of the jth segment (i.e., $\{j, j+1\} \subseteq \mathsf{BR}(\overline{\alpha}^k)$).

There are two cases: either segments k and $k+1$ are the first two segments of the dynamic contract π (i.e., $k=1$), or there exists a $(k-1)$st segment. In the first case, the agent is indifferent between actions j and $j+1$ for the entire first segment (from time 0 to τ^1), but plays action j. This contradicts the fact that π cannot be reduced further by Lemma B.3.

In the second case, the $(k-1)$st segment must incentivize action $j+1$ for the learner (since the sequence of actions is decreasing up until segment k). But this means that the agent must be indifferent between actions j and $j+1$ also after the $(k-1)$st segment, and thus for the entirety of the kth segment ($\{j, j+1\} \subseteq \mathsf{BR}(\overline{\alpha}^{k-1})$). Since the agent plays j during the kth segment, this also contradicts the fact that π cannot be reduced further by Lemma B.3 (see Figure 4 for an example of this reduction). □

We are now almost done – Lemma B.4 shows we can rewrite any dynamic contract so that the agent descends through their action space. We now need only show that the principal should abruptly switch to offering the zero contract after the first segment (instead of slowing the rate of descent through these regions by offering a positive contract). We do this in our final rewriting lemma.



(a) Base policy π. The kth segment is the first non-free-fall segment; we decompose it into a segment $\alpha = \alpha_{BR} = \alpha_{3,4}$ and a free-fall segment with $\alpha = 0$.

(b) The segment with $\alpha = \alpha_{BR}$ can be rewritten via Lemma B.3 to lie in region 4, where it can then be further combined with earlier segments via Lemma B.2. This moves the first non-free-fall action earlier in π.

Figure 5: Illustrations for the proof of Lemma B.5.

Lemma B.5. *Let π be a dynamic linear contract where the agent plays a decreasing sequence of actions. Then there exists a free-fall linear contract π' with $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$.*

Proof. Let $\pi = \{(\alpha^k, \tau^k, a^k)\}_{k=1}^K$ (with a^k decreasing), and consider the last non-free-fall segment (α^k, τ^k, a^k), i.e., k is the maximal k for which $\alpha^k \neq 0$. Assume that $k > 1$ (if not, then π is already a free-fall contract).

Let $\alpha_{BR} = \alpha_{a^k,a^k+1}$ be the indifference contract for the best-response boundary separating the current action from the previously incentivized action. Consider replacing this segment with the two consecutive segments $(\alpha_{BR}, (\alpha^k/\alpha_{BR})\tau^k, a^k)$, $(0, (1 - \alpha^k/\alpha_{BR})\tau^k, a^k))$. In doing so we essentially are doing the inverse of the first rewriting rule in Lemma B.2 – replacing a single segment with two segments that average to the original segment – and because of this, the resulting dynamic contract is valid and has the same utility as our original contract (the construction also guarantees both segments stay within this region). But now we have a segment $(\alpha_{BR}, (\alpha^k/\alpha_{BR})\tau^k, a^k)$ that lies along the best-response boundary α_{BR}, so by Lemma B.3 we can replace it with the segment $(\alpha_{BR}, (\alpha^k/\alpha_{BR})\tau^k, a^k + 1)$ and strictly increase the utility of our dynamic contract (see Figure 5).

We can then merge this segment with the previous segment in (which also incentivizes action $a^k + 1$) to obtain a new dynamic contract with strictly greater utility than π and whose first non-free-fall action occurs strictly earlier. Repeating this process, we obtain a free-fall contract π' with at least the same utility as π. □

We can now prove the main theorem of this section.

Proof of Theorem 3.1. From Lemmas B.4 and B.5 the first part of this theorem (that there exists a free-fall linear contract π' with $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$) immediately follows.

To show that we can efficiently compute this free-fall contract, note that the optimal free-fall linear contract might as well start with a segment of the form $(\alpha_{i-1,i}, \tau, i)$ for some indifference contract $\alpha_{i-1,i}$ (if it does not start by offering some indifference contract, we can apply the rewriting rule of Lemma B.2 to merge this segment with the following segment, which would incentivize the same action).

It is also true that the optimal free-fall linear contract might as well *end* at an indifference contract: that is, $\overline{\alpha}^K = \alpha_{j-1,j}$ for some j. To see this, consider a free-fall linear contract π that does not end on an indifference contract. It ends with a segment of the form $(0, \tau^K, a)$ for some agent action a. Consider the contract $\pi(\tau)$ formed by replacing the duration of the last segment with τ; this operation is valid for all τ in some interval $[0, \tau_{\max}]$. Note that $\mathsf{Util}(\pi(\tau))$ is a convex function of τ (it is of the form $(\mathsf{Util}(\pi(0))\mathcal{T}^{K-1} + u_P(0, a)\tau)/(\mathcal{T}^{K-1} + \tau)$) so it is maximized when τ equals one of the endpoints of this interval. But at both endpoints, $\overline{\alpha}^K$ lies on a best-response boundary (for $\tau = 0$, $\alpha_{a,a+1}$, for $\tau = \tau_{\max}$, $\alpha_{a-1,a}$).

Since our optimal contract is completely characterized by its start and end points, it can be computed in polynomial time in n by testing all the pairs of indifference points $\{\alpha_{i-1,i}, \alpha_{j-1,j}\}$ with $j \leq i$ as candidates for the start and end points of the optimal initial contract (this pair of indifference points also uniquely specifies the fraction of time that must be spent in free-fall). Note that in the case where in the optimal free-fall contract $i = j$, the optimal contract is the best static contract. □

In Appendix D (see Theorem D.1), we generalize Theorem 3.1, showing that free-fall contracts are optimal for a much broader family of dynamic contracts with "single-dimensional scaling," where the principal is using an arbitrary non-linear contract and dynamically rescales it during the interaction with the agent.

Our proof of Theorem D.1 is parallel to the proof of Theorem 3.1 in the sense that we demonstrate how to gradually transform a general single-dimensional-scaling contract into a free-fall contract, while increasing utility for the principal. The main difficulty in applying the proof of Theorem 3.1 directly is that the rewriting rule in Lemma B.3 no longer holds – for general contracts with single-dimensional scaling, it is not the case that segments along a best-response boundary should always incentivize the higher action for the agent. In the proof of Theorem D.1, we forego the use of this rewriting rule and instead using the weaker condition that there cannot be two consecutive segments along a best-response boundary.

C Proof of Theorem 3.2 (Win-Win)

Proof. Consider the following contract game.[10] There are $n > 2$ actions, with expected reward $R_i = v^i$ for some $v > 0$. Concretely, we let $v = 2$. The cost of the action are specified recursively by $c_1 = 0$ and $c_i = c_{i-1} + R_{i-1} - \frac{1}{2}$ for $i > 1$, yielding $c_{i>1} = \sum_{k=2}^{i}(2^{k-1} - \frac{1}{2})$. The resulting indifference contracts are thus $\alpha_i = 1 - 2^{-i}$ for $1 < i \leq n$. In this game, the principal has the same utility (of one unit) for all the indifference contracts. The agent's utility under the contract α_i, as the reader can verify, is given by $2^i - \frac{3}{2} - \sum_{k=1}^{i}(2^{k-1} - \frac{1}{2}) = \frac{1}{2}(1+i)$. Notably, this utility is higher for the higher actions. The welfare of action i is thus $w_i = \frac{1}{2}(3+i)$. Next, we slightly alter this game by increasing the payoff of action 2 by a small amount $\epsilon > 0$ such that the optimal static contract is now α_2, which yields a utility of $1 + \mathcal{O}(\epsilon)$ for the principal, and the agent is still indifferent under this contract between action 2 and the null action. In the following analysis, we are mainly interested in large (but finite) n. Notice that the optimal static contract is extremely inefficient for large n, getting an arbitrarily low (independent of n) fraction of the optimal welfare.

Now consider an optimal dynamic strategy; by Theorem 3.1, there is an optimal strategy of a free-fall form. We will construct a free-fall contract **p** that starts at α_n, so action n is played initially, where the duration λT of that stage is chosen such that the final action at time T is action $\lceil \frac{1}{2}\log(n) \rceil$. Specifically, we require $\lambda\alpha_n + (1-\lambda)\cdot 0 = \alpha_{\lceil \frac{1}{2}\log(n) \rceil}$, and so $\lambda = \frac{2^n}{2^n - 1}\left(1 - \frac{1}{\sqrt{n}}\right)$. We show that this free-fall strategy bounds the utilities of both players from below.

Claim C.1. *In an optimal free-fall contract, the utilities for both players are at least those obtained in the contract described above.*

For ease of presentation, the proof for this claim is deferred to the following subsection. It consists of three parts: first, we show that an optimal free-fall contract must start at α_n. This is done directly by way of contradiction. Then, the proof shows that the last action that is played by the agent in an optimal free-fall contract must be higher than $\frac{1}{2}\log n$. The intuition for this part of the proof is that as the principal continues to free fall through lower and lower actions, the marginal gain from each action (which is the expected reward of that action because we are free falling) continues to diminish. At some point, the marginal gain is outweighed by the current average principal utility, which we show should occur at action $\Theta(\log n)$ (since we know the principal can get an average utility of $\Theta(n)$ and the expected reward of action i is 2^i). Lastly, we compare the utilities of both players in a free-fall strategy that begins at action n and ends at action $\lceil \frac{1}{2}\log n \rceil$ to those of the optimal free-fall strategy and observe that the utilities in the former case bound the respective utilities in the latter case from below. The principal's utility is clearly bounded from below by her utility in our strategy due to optimality. For the agent, the total utility is determined by the stopping point. Since the agent's utility at α_i is increasing with i in our game, we conclude that the agent's utility in an optimal contract is at least $\frac{1}{2}\log n$.

Now let us calculate the average utilities for the players under our dynamic strategy, averaged over the whole sequence of play. The agent's average utility at the last step is the same as the utility that would have been obtained under the average contract at that time, which is $\frac{1}{2}(1 + \lceil \frac{1}{2}\log(n) \rceil)$. To calculate the utility for the principal, we define t_i to be the time when the agent switches from action i to action $i-1$. We know that the transition from action n to $n-1$ happens at time $t_n = \lambda T$, and until that time the principal gains a utility of one per time unit. After that time, the average contract at time t is the weighted average until t of the contract α_n with weight λT and zero contract with weight $t - \lambda T$. Therefore, the transition times from each action i are given by $t_i = \lambda T \frac{\alpha_n}{\alpha_i}$. After time λT, the principal pays zero and extracts the full welfare from the agents actions, and so the overall utility for the principal is $\lambda T + \sum_{i=\lceil \frac{1}{2}\log(n) \rceil}^{n}(t_{i-1} - t_i)R_{i-1}$.

Claim C.2. *The utility for the principal in the free-fall contract (α_n, λ) is $\mathcal{O}(n)$.*

[10] In this example we shift the rewards with an additive constant such that the reward for the principal when the agent takes the null action equals some constant instead of zero. This simplifies the following analysis and is without loss of generality.

Proof. The utility from region i is $\lambda T + \sum_{i=\lceil \frac{1}{2}\log(n)\rceil}^{n}(t_{i-1}-t_i)R_{i-1}$. The time intervals are

$$(t_{i-1}-t_i) = \lambda T\alpha_n\Big(\frac{1}{\alpha_{i-1}}-\frac{1}{\alpha_i}\Big) = \frac{\lambda T\alpha_n 2^i}{(2^i-2)(2^i-1)}.$$

The utility for the principal from region $i > \frac{1}{2}\log(n)$ and large n is thus:

$$\frac{\lambda T\alpha_n 2^{2i}}{2(2^i-2)(2^i-1)} = \Theta(1).$$

Summing over $n - \frac{1}{2}\log(n)$ such terms yields a utility of $\mathcal{O}(n)$. □

The above arguments hold similarly also for perturbed versions of this game. For example, shifting the rewards by arbitrary and independent values in the range $[-1, 1]$, as well as re-scaling the reward parameter v, yielding a positive measure in the parameter space. □

C.1 Proof of Claim C.1

Proof. We execute this proof in two parts. In the first part of the proof, we will show that any optimal dynamic (free fall) contract must begin at α_n. In the second part of the proof, we show that an optimal dynamic (free fall) contract that begins at α_n must end at $\alpha_{\lceil \frac{1}{2}\log n\rceil}$ or higher, if n is sufficiently large. This is enough to imply the claim because if the optimal free fall contract stops at a higher action than $\lceil \frac{1}{2}\log n\rceil$, then the principal has higher utility due to optimality and the agent has higher utility since their utility is increasing in actions.

We now prove that any optimal dynamic contract must begin at α_n. For the sake of contradiction, suppose that it instead begins at α_i for some action $i \in [1, n-1]$. In particular, it begins with the segment $(p^1 = \alpha_i R, \tau^1, a^1 = i)$ for some $i \in [1, n-1]$. To achieve a contradiction, we will show that this dynamic contract is not optimal by producing a better dynamic contract.

In particular, let us consider replacing this first segment with the following two segments: $(\alpha_{i+1}R, x \triangleq \frac{\alpha_i}{\alpha_{i+1}}\tau^1, i+1), (0, y \triangleq \left[1-\frac{\alpha_i}{\alpha_{i+1}}\right]\tau^1, i)$ (and re-indexing all subsequent segments appropriately). We claim that this will achieve strictly greater principal utility, while leaving the total time unaffected. We first show how we solved for the appropriate time-split (x, y).

$$\begin{aligned}
x + y &= \tau^1 \qquad ((x, y) \text{ is a time split}) \\
x\alpha_{i+1} &= \tau^1\alpha_i \qquad (\text{At time } \tau^1\text{, the cumulative linear contract is still } \alpha_i) \\
x &= \frac{\alpha_i}{\alpha_{i+1}}\tau^1 \\
y &= \left[1-\frac{\alpha_i}{\alpha_{i+1}}\right]\tau^1
\end{aligned}$$

By construction, our choice of x and y keeps the total time invariant, so it remains to prove that this results in strictly more principal utility. Since all subsequent segments are the same and generate the same amount of principal utility, we only need to compare the principal utility of these three segments.

The (cumulative) principal utility of the original segment $(\alpha_i R, \tau^1, i)$ is just τ^1 since the contract problem is designed so that all indifference contracts α_i result in one unit of utility to the principal. The exception is action one, which was adjusted to have $1 + O(\varepsilon)$ principal utility and therefore has cumulative principal utility $\tau^1(1 + O(\varepsilon))$.

Next, we consider the cumulative principal utility of our two new segments $(a_{i+1}R, x, i+1)$ and $(0, y, i)$. The first segment has (cumulative) principal utility equal to just x for the same reason as above (but now $i+1$ cannot be the first action). The second segment has (cumulative) principal utility equal to $y(R_{i+1})$ where R_{i+1} is the expected reward from action $i+1$, due to the fact that this segment offers the zero contract. Together, these two segments generate (cumulative) principal utility equal to the following.

$$\begin{aligned}
x + y(R_{i+1}) &= (x + y) + y(R_{i+1} - 1) \\
&= \tau^1 + y(2^{i+1} - 1)
\end{aligned}$$

However, we can see from our choice of y that $y > 0$ and $(2^{i+1} - 1) > 0$ since $i \geq 1$. Hence this strictly beats the cumulative principal utility of the original segment as long as ε is sufficiently small. This completes our contradiction, since the original dynamic contract was assumed to be optimal but we found a strictly better one. Hence the optimal dynamic contract must free fall from α_n (which there is no higher action to start from instead), completing the first part of the proof.

We now use this fact to prove that the optimal dynamic (free fall) contract must end at $\alpha_{\lceil \frac{1}{2}\log n \rceil}$ or higher, if n is sufficiently larger. The proof plan is to consider the effect of free falling through an additional action, and determining when that might improve the free fall contract. As a first step, we observe that the objective function of the continuous setting, $\mathsf{Util}(\pi)$, is invariant when we equally scale all times τ^k. As a result, we can assume without loss of generality that the first segment of free-fall $(p^1 = \alpha_n R, \tau^1, a^1 = n)$ uses $\tau^1 = 1$. We can also assume without loss of generality that the other segments $\{(p^k = 0, \tau^k, a^k = n - k + 1)\}_{k=2}^K$ begin and end at region boundaries, which is enough to work out their durations τ^k based on when the average linear contract reaches a particular indifference point.

$$
\begin{aligned}
\tau^k &= \frac{\alpha_n}{\alpha_{n-k+1}} - \frac{\alpha_n}{\alpha_{n-k+2}} = \frac{1-2^{-n}}{1-2^{-n+k-1}} - \frac{1-2^{-n}}{1-2^{-n+k-2}} \\
&= [1-2^{-n}]\frac{2^{-n+k-1} - 2^{-n+k-2}}{(1-2^{-n+k-1})(1-2^{-n+k-2})} \\
&= [1-2^{-n}]\frac{2^{-n+k-2}}{(1-2^{-n+k-1})(1-2^{-n+k-2})}
\end{aligned}
$$

Hence segment $k \in [2, K]$ contributes the following (cumulative) principal utility.

$$
\begin{aligned}
\tau^k u_P(p^k, a^k) &= \tau^k 2^{n-k+1} = 2^{n-k+1} \cdot \left[1-2^{-n}\right] \frac{2^{-n+k-2}}{(1-2^{-n+k-1})(1-2^{-n+k-2})} \\
&= \frac{1}{2}\left[1-2^{-n}\right] \frac{1}{(1-2^{-n+k-1})(1-2^{-n+k-2})}
\end{aligned}
$$

Let π_K be the trajectory that uses K segments. We can compute its objective value to be the following.

$$
\begin{aligned}
\mathsf{Util}(\pi_K) &= \frac{1 + \frac{1}{2}[1-2^{-n}]\sum_{k=2}^K \frac{1}{(1-2^{-n+k-1})(1-2^{-n+k-2})}}{(1-2^{-n})/(1-2^{-n+K-1})} \\
&= (1-2^{-n+K-1})\left[1/(1-2^{-n}) + \frac{1}{2}\sum_{k=2}^K \frac{1}{(1-2^{-n+k-1})(1-2^{-n+k-2})}\right]
\end{aligned}
$$

We can take the difference of two such expressions to decide whether π_{K+1} is better than π_K. For $n - \lceil \frac{1}{2} \log n \rceil \leq K \leq n-1$:

$$\begin{aligned}
\mathsf{Util}(\pi_{K+1}) - \mathsf{Util}(\pi_K) &= (1-2^{-n+K}) \left[1/(1-2^{-n}) + \frac{1}{2} \sum_{k=2}^{K+1} \frac{1}{(1-2^{-n+k-1})(1-2^{-n+k-2})} \right] \\
&\quad - (1-2^{-n+K-1}) \left[1/(1-2^{-n}) + \frac{1}{2} \sum_{k=2}^{K} \frac{1}{(1-2^{-n+k-1})(1-2^{-n+k-2})} \right] \\
&= (1-2^{-n+K}) \frac{1}{2} \frac{1}{(1-2^{-n+K})(1-2^{-n+K-1})} \\
&\quad - 2^{-n+K-1} \left[1/(1-2^{-n}) + \frac{1}{2} \sum_{k=2}^{K} \frac{1}{(1-2^{-n+k-1})(1-2^{-n+k-2})} \right] \\
&\leq \frac{1}{2} \frac{1}{(1-2^{-n+(n-1)})(1-2^{-n+(n-1)-1})} \\
&\quad - 2^{-n+(n-\frac{1}{2}\log n)-1} \left[\frac{1}{2} \sum_{k=2}^{n-\lceil \frac{1}{2} \log n \rceil} 1 \right] \\
&= \frac{1}{2} \frac{1}{(1/2)(3/4)} - \frac{1}{2\sqrt{n}} \left[\frac{1}{2} (n - \left\lceil \frac{1}{2} \log n \right\rceil - 1) \right]
\end{aligned}$$

Since the positive term has magnitude $O(1)$ and the negative term has magnitude $O(\sqrt{n})$, this bound will always be negative when n is sufficiently large. Hence it is strictly not worth it to free fall below $\alpha_{\lceil \frac{1}{2} \log n \rceil}$, as desired. This completes the proof. $\square$

D General Contracts with Single-Dimensional Scaling

Here we consider general contracts, and in Theorem D.1 generalize the result of Theorem 3.1 to families of one-dimensional (yet non-linear) dynamic contracts for which free-fall contracts are optimal.

Given any contract $\mathbf{p}$, the set of $\mathbf{p}$-scaled contracts are the one-dimensional family of contracts of the form $\alpha \mathbf{p}$ for some $\alpha \geq 0$. We will consider a principal that is restricted to only play $\mathbf{p}$-scaled contracts. In the continuous-time formulation of Section 2.2, this means that each contract p^k must be $\mathbf{p}$-scaled. We will let $p^k = \alpha^k \mathbf{p}$, and we will often abuse notation and write α^k as shorthand for this contract (e.g., we will specify segments of the trajectory π in the form (α^k, τ^k, a^k)). Recall that a free-fall contract denotes such a dynamic contract for the principal where $\alpha^k = 0$ for all $k > 1$.

As with linear contracts, note that as α increases from 0, the contract $\alpha \mathbf{p}$ incentivizes the agent to play an action in $\mathsf{BR}_{\mathbf{p}}(\alpha)$ (which is unique except for at most n "breakpoint" values of α, where the agent is indifferent between two actions). This induces an ordering over the actions; we will relabel the actions so that actions 1 (the null action), 2, 3, $\ldots$ are incentivized for increasing values of α. Formally, if the agent has n actions, we have n "breakpoints" $0 = \alpha_{0,1} < \alpha_{1,2} < \alpha_{2,3} < \cdots < \alpha_{n-1,n}$, where action i belongs to $\mathsf{BR}_{\mathbf{p}}(\alpha)$ iff $\alpha \in [\alpha_{i-1,i}, \alpha_{i,i+1}]$ (with $\alpha_{n,n+1} = \infty$).

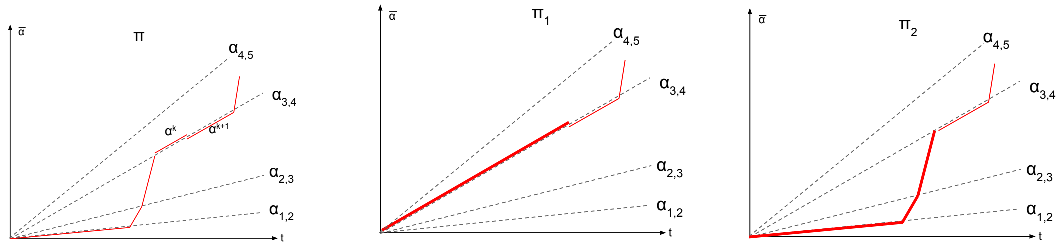
Our main result in this section is the following theorem, by which free-fall $\mathbf{p}$-scaled contracts are optimal $\mathbf{p}$-scaled dynamic contracts.

Theorem D.1. *Let π be any $\mathbf{p}$-scaled dynamic contract. Then there exists a free-fall $\mathbf{p}$-scaled contract π' where $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$.*

To prove Theorem D.1, we will establish a sequence of lemmas constraining the potential geometry of an optimal $\mathbf{p}$-scaled dynamic contract. Note that since linear contracts are a specific case of $\mathbf{p}$-scaled contracts, this also provides an alternate proof of Theorem 3.1.

We begin our proof with the observation that, similar to linear contracts, $\mathbf{p}$-scaled contracts cannot "skip over" actions for the agent (c.f. Lemma B.1, which has an essentially identical proof).

Lemma D.2. *If $\pi = \{(\alpha^k, \tau^k, a^k)\}$ is a $\mathbf{p}$-scaled dynamic contract, then $\forall k$, $|a^k - a^{k+1}| \leq 1$.*



(a) Base policy π. In this example the kth and $(k+1)$st segments both lie along the α_3 indifference boundary, but the two segments incentivize different actions ($a^k = 4$, $a^{k+1} = 3$).

(b) Policy π_1 formed by replacing the first k segments of π with an enlarged version of the kth segment.

(c) Policy π_2 formed by replacing the first k segments of π with an enlarged version of the first $k-1$ segments.

Figure 6: Figures for the proof of Lemma D.3.

Proof. Note that a^k and a^{k+1} both must be best-responses to the average historical contract $\overline{\mathbf{p}}^k$ after segment k, which is a $\mathbf{p}$-scaled contract with parameter $\overline{\alpha}^k = \sum_{k'=1}^{k} \alpha^{k'} \tau^{k'} / \sum_{i=1}^{k} \tau^{k'}$. In other words, a^k and a^{k+1} belong to $\mathsf{BR}_{\mathbf{p}}(\overline{\alpha}^k)$. Since $\mathsf{BR}_{\mathbf{p}}(\alpha)$ is always of the form $\{i\}$ or $\{i, i+1\}$ the conclusion follows. $\square$

Now, recall that in Lemma B.2 we show that (for general contract problems) we can restrict our attention to trajectories where no two consecutive segments have the same agent best response (i.e., $a^k \neq a^{k+1}$ for any k). The following lemma proves a strengthening of this fact specific to $\mathbf{p}$-scaled contracts.

Lemma D.3. *Let π be any $\mathbf{p}$-scaled dynamic contract. Then there exists a $\mathbf{p}$-scaled dynamic contract $\pi' = \{(\alpha^k, \tau^k, a^k)\}$ with the property that for all k, $a^k \neq a^{k+1}$ and $a^k \neq a^{k+2}$, and that* $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$.

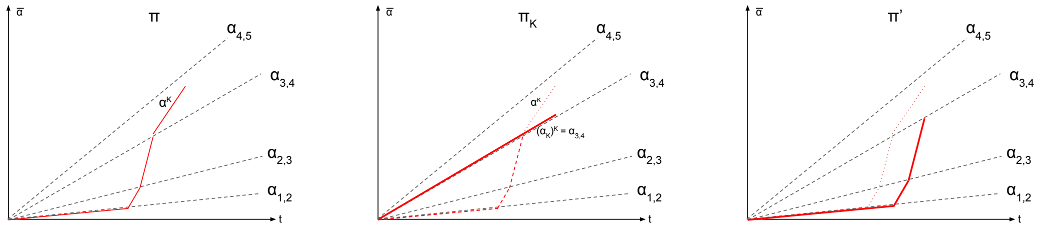
Proof. The fact that we can rewrite π into an equivalent contract where $a^k \neq a^{k+1}$ follows from the proof of Lemma B.2. Therefore, assume without loss of generality that π already has this form. We will show how to rewrite it into a new dynamic contract π' with the additional property that $a^k \neq a^{k+2}$.

We will induct on the number of segments in the path (it is obviously true when there is only $K = 1$ segment). Assume that for some k, π has the property that $a^k = a^{k+2} \neq a^{k+1}$. This implies that $\mathsf{BR}_{\mathbf{p}}(\overline{\alpha}^k) = \mathsf{BR}_{\mathbf{p}}(\overline{\alpha}^{k+1}) = \{a^k, a^{k+1}\}$. Since there is a unique value of α for which $\mathsf{BR}_{\mathbf{p}}(\alpha) = \{a^k, a^{k+1}\}$ (namely, one of the breakpoints $\alpha_{i,i+1}$), this can only happen if $\overline{\alpha}^k = \overline{\alpha}^{k+1}$, which in turn means that $\alpha^{k+1} = \overline{\alpha}^k = \overline{\alpha}^{k+1}$. Pictorially, this is because if a dynamic contract spends only one segment in a best-response region, this segment must lie along the boundary of the best-response region (see Figure 6).

Now, consider the following two modifications of π:

1. In π_1, we replace the first $k+1$ segments of π with a scaled up version of the $(k+1)$st segment. That is, remove the first $k+1$ segments of π and replace them with $(\alpha^{k+1}, \mathcal{T}^{k+1}, a^{k+1})$. To see that this is a valid contract, note that since $\alpha^{k+1} = \overline{\alpha}^{k+1}$, α^{k+1} incentivizes action a^{k+1} so the first segment of this contract is valid. Moreover, after $\mathcal{T}^{k+1}$ time units have elapsed, both π and π_1 resume the same sequence of segments from the same state $\overline{\alpha}^{k+1}$.

2. In π_2, we replace the first $k+1$ segments of π with a scaled up version of the first k segments. That is, remove the segment $(\alpha^{k+1}, \tau^{k+1}, a^{k+1})$, and scale up $\tau^{k'}$ (for each $1 \leq k' \leq k$) to $\tau^{k'}(\mathcal{T}^{k+1}/\mathcal{T}^k)$. Again, this is a valid dynamic contract because scaling up a (prefix of a) dynamic contract results in a valid dynamic contract, and π and π_2 both resume the remainder of segments at the same time and from the same state $\overline{\alpha}^{k+1}$.



(a) Base policy π with consecutively increasing actions.

(b) Policy π_K formed by setting α^K to the minimum α required to remain in the same best-response region (in this case, α_3).

(c) Policy π' formed by scaling up the first $K-1$ segments of π (i.e., all but the last segment).

Figure 7: Figures for the proof of Lemma D.4.

Finally, note that $\mathsf{Util}(\pi)$ is a convex combination of $\mathsf{Util}(\pi_1)$ and $\mathsf{Util}(\pi_2)$ – specifically, $\mathsf{Util}(\pi) = (\tau^{k+1}\pi_1 + \mathcal{T}^k\pi_2)/\mathcal{T}^{k+1}$ – and so is less than or equal to one of them. But both π_1 and π_2 have strictly fewer segments than π, so by applying the inductive hypothesis, we are finished. □

As a consequence of Lemmas D.2 and D.3, we can restrict ourselves to dynamic contracts whose sequences of actions are either consecutively increasing ($a^{k+1} = a^k + 1$) or consecutively decreasing ($a^{k+1} = a^k - 1$). We show that we can ignore the first case – such contracts can never be better than static contracts.

Lemma D.4. *Let $\pi = \{(\alpha^k, \tau^k, a^k)\}$ be a $\mathbf{p}$-scaled dynamic contract where the a^k are consecutively increasing. Then there exists a static $\mathbf{p}$-scaled contract π' (i.e., a single segment dynamic contract of the form $(\alpha', 1, a')$) where $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$.*

Proof. As in the proof of Lemma D.3, we will again induct on the number of segments of π. If π has one segment, we are done.

Now consider a π with K segments, whose last segment is (α^K, τ^K, a^K). Recall that for any i, $\alpha_{i-1,i}$ is the smallest value of α for which $\alpha\mathbf{p}$ incentivizes action i. Note that if $\alpha^K > \alpha_{a^K-1,a^K}$, we can improve the utility of the principal by decreasing α^K to α_{a^K-1,a^K} (this pays strictly less to the agent but still incentivizes the same action a^K). We'll therefore assume the last segment is of the form $(\alpha_{a^K-1,a^K}, \tau^K, a^K)$; note that this segment by itself is a valid static $\mathbf{p}$-scaled contract, as α_{a^K-1,a^K} incentivizes action a^K. Call this contract π_K.

Let π' be the dynamic contract formed by the first $K-1$ segments of π (see Figure 7 for examples of π_K and π'). But now, $\mathsf{Util}(\pi)$ is a convex combination of $\mathsf{Util}(\pi')$ and $\mathsf{Util}(\pi^K)$, so it is at most the maximum of these two quantities. If this maximum is $\mathsf{Util}(\pi^K)$, we are done (π^K is a static contract); if it is $\mathsf{Util}(\pi')$, we are also done by the inductive hypothesis (π' has $K-1$ segments). □

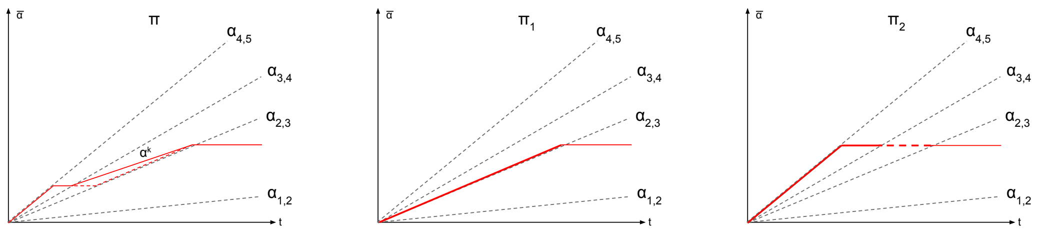
Finally, we show that in the case where the sequence of actions are consecutively decreasing, such a contract is no better than some free-fall contract.

Lemma D.5. *Let $\pi = \{(\alpha^k, \tau^k, a^k)\}$ be a $\mathbf{p}$-scaled dynamic contract where the a^k are consecutively decreasing. Then there exists a free-fall $\mathbf{p}$-scaled contract π' where $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$.*

Proof. Assume that π is not a free-fall contract. We will show we can rewrite π in a way so that either the first agent action a^1 strictly decreases or the first non-free-fall occurs strictly later. Since the number of segments in π is bounded (by n, since the actions are consecutively decreasing), this implies the theorem statement.

Let (α^k, τ^k, a^k) be the first segment in π with $k \geq 2$ where $\alpha^k > 0$ (so, the dynamic contract is not free-falling here). Note that since this segment ends on the boundary between the best-response regions for a^k and $a^{k+1} = a^k - 1$, $\overline{\alpha}^k = \alpha_{a^k-1,a^k}$.

The main observation of this proof is that we can rewrite this segment as a combination of a free-fall segment (with $\alpha = 0$) and a segment along the boundary of these two best-response regions (with



(a) Base policy π with consecutively decreasing actions. The kth segment is the first non-free-fall segment: we form π' (the dashed trajectory) by decomposing this segment into a free-fall segment followed by a boundary segment.

(b) Policy π_1 formed by replacing the first k segments of π by a scaled up version of the boundary dashed segment.

(c) Policy π_2 formed by replacing the first k segments of π by a scaled up version of the prefix of π up to (but not including) the boundary dashed segment.

Figure 8: Figures for the proof of Lemma D.5.

$\alpha = \overline{\alpha}^k$). Specifically, form a new dynamic contract π' by replacing (α^k, τ^k, a^k) in π with the two consecutive segments $(0, \lambda\tau^k, a^k)$ and $(\alpha_{a^k-1,a^k}, (1-\lambda)\tau^k, a^k)$, where λ is chosen so that $(1-\lambda)\alpha_{a^k-1,a^k} = \alpha^k$. Note that by doing this π' now has $K+1$ segments, where segments k and $k+1$ are this new free-fall and boundary segment respectively. Note that we will let quantities like τ^k, $\mathcal{T}^k$, and $\overline{\alpha}^k$ still refer to the relevant quantities for π, not π'.

This allows us to proceed via a similar technique as in Lemma D.3. Consider the following two modifications of π' (see Figure 8 for examples):

1. In π_1, we replace the first $k+1$ segments of π' with a scaled-up version of the boundary segment of the form $(\overline{\alpha}^k, \mathcal{T}^k, a^k)$.
2. In π_2, we replace the first $k+1$ segments of π' with a scaled up version of the first k segments (the first $k-1$ segments of π and the free-fall segment, but not the boundary segment). Specifically, let $C = \mathcal{T}^k/(\mathcal{T}^{k-1} + \lambda\tau^k)$. Then the first $k-1$ segments of π_2 are of the form $(\alpha^{k'}, C\tau^{k'}, a^{k'})$, and the kth segment of π_2 is of the form $(0, C\lambda\tau^k, a^k)$.

As in the proof of Lemma D.3, we can check that both π_1 and π_2 are valid dynamic contracts: in particular, after $\mathcal{T}^k$ units of time, they are both in the state $\overline{\alpha}^k$, so the remaining suffix of π is a valid extension for both contracts.

Again, $\mathsf{Util}(\pi)$ can be written as a convex combination of $\mathsf{Util}(\pi_1)$ and $\mathsf{Util}(\pi_2)$, specifically,

$$\mathsf{Util}(\pi) = \frac{(1-\lambda)\tau^k \mathsf{Util}(\pi_1) + (\mathcal{T}^{k-1} + \lambda\tau^k)\mathsf{Util}(\pi_2)}{\mathcal{T}^k}.$$

But π_1 starts at a later action than π (since $a^k = a^1 - (k-1)$), and π_2 is a free-fall contract for one further step than π (since $\alpha^2 = \alpha^3 = \cdots = \alpha^{k-1} = 0$, and the kth segment in π_2 also has $\alpha = 0$). This completes the proof. $\square$

We can now conclude the proof of Theorem D.1.

Proof of Theorem D.1. Because of Lemmas D.2 and D.3, we can assume without loss of generality that the actions a^k in π are either consecutively increasing or decreasing. The conclusion now immediately follows from Lemmas D.4 and D.5. $\square$

E Proof of Thorem 4.2 (Unknown Time Horizon)

Proof. Due to Theorem H.3, it suffices to show that there exists some $\underline{\gamma}$ such that $U^{\star}_{\underline{\gamma}} < (1+\varepsilon)R_{\star}$. This lets us focus on the continuous-time setting.

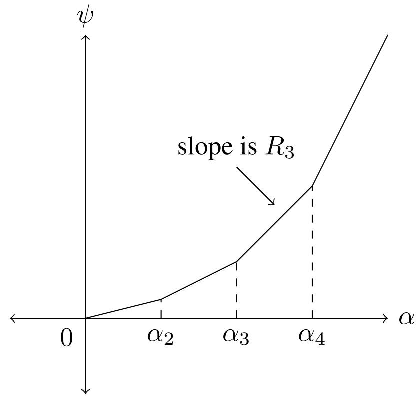


Figure 9: We use a "raw" potential function $\psi(\alpha)$ which maps time-averaged linear contracts α to (raw) potentials.

The high-level plan from here is to focus on a particular continuous trajectory $\pi = \left\{(p^k, \tau^k, a^k\right\}_k$ apply a potential argument to it. We will then show our analysis extends to distributions $\mathcal{D}$ for free. We will define a potential function $\psi(\alpha)$ that maps time-averaged linear contracts α to potentials in $\mathbb{R}_{\geq 0}$. This potential is based only on the principal-agent problem (c, F, r). There are some peculiarities about our potential argument, relating to the passage of time. Consider a principal managing to produce a time-averaged linear contract of α after t units of time, and compare that with a principal that has managed to arrive a time-averaged linear contract of α after $2t$ units of time instead, i.e. twice the time. In terms of absolute (not time-averaged) units of profit we can extract from this point, it is twice as good to be in the latter situation. With this in mind, our proof will carefully distinguish between the *raw potential* $\psi(\alpha)$ and the *time-weighted potential* $\psi(\alpha) \cdot t$. If a principal maintains a steady time-averaged linear contract, then the raw potential will remain constant while the time-weighted potential will grow.

The purpose of the time-weighted potential is to model the ability of a principal to extract additional profit by gradually lowering time-averaged linear contract. It will be used to demonstrate that this extra profit produced by using up a finite resource, which will imply the desired theorem result.

We now give our raw potential function $\psi(\alpha)$. We begin by writing down the linear contract breakpoints of (c, F, r); without loss of generality[11] they are $0 < \alpha_2 < \alpha_3 < \cdots < \alpha_n$, where the linear contract α_i leaves the agent indifferent between actions $i-1$ and i. For notational convenience, we also define an $\alpha_1 \triangleq 0$ as the minimum linear contract to incentivize the first action. We also denote the expected reward of action i with R_i. With this notation in place, our raw potential function $\psi : [0, \alpha_n] \to \mathbb{R}_{\geq 0}$ is the following piecewise-linear function. Note that we can assume without loss of generality that the average linear contract never exceeds α_n, because capping it to this quantity only improves principal utility at all moments in time.

$$\psi(\alpha) \triangleq \begin{cases} \sum_{i=1}^{i'-1} (\alpha_{i+1} - \alpha_i) R_i + (\alpha - \alpha_{i'}) R_{i'} & \text{if } \alpha \in [\alpha_{i'}, \alpha_{i'+1}) \\ \sum_{i=1}^{n-1} (\alpha_{i+1} - \alpha_i) R_i & \text{if } \alpha = \alpha_n \end{cases}$$

The potential above is depicted in Figure 9 and can be seen as the product of the following thought experiment: what if the principal was allowed to offer unbounded payments (in particular, payments can be negative and can exceed the payment bound P)? In our continuous-time setting, this gives the principal the ability to produce segments of play (p^k, τ^k, a^k) which have near-instantaneous times $\tau^k \to 0$ while using large-magnitude cumulative contracts $p^k \tau^k$ to move between the boundaries between actions. If these near-instantaneous actions are used at time t, then the time-weighted potential $\psi(\alpha) \cdot t$ captures the necessary payments to alter the time-averaged linear contract. One interesting aside about this thought experiment is that the necessary payment to near-instantaneously move up from α_i to the next α_{i+1}, namely $[\psi(\alpha_{i+1}) - \psi(\alpha_i)]\, t$, is equal to the payout received for near-instantaneously using a negative contract to move down from α_{i+1} to α_i.

[11] Implicitly, this step prunes away all actions which cannot be incentivized by a linear contract.

Potential function in hand, we return to the original problem where payments are bounded and nonnegative. Let us consider the k^{th} segment of play $(p^k = \alpha^k R, \tau^k, a^k)$ and relate the total profit generated during this segment of play with the change in potential.

For notational convenience we define shorthand for the cumulative linear contract offered.

$$\mathcal{A}^k \triangleq \sum_{k'=1}^{k} \tau^{k'} \alpha^k$$

We will also use u_P^k to denote the (time-weighted) principal utility for segment k and $u_\star^k$ to denote the corresponding amount of principal utility that the optimal static contract obtains over τ^k time. Using this notation, we can compute an upper bound on how much additional principal utility this segment manages to achieve over the optimal static contract.

$$\begin{aligned}
u_P^k &= \left[(1-\alpha^k)R_{a^k}\right]\tau^k \\
u_\star^k &= \left[\max_a (1-\alpha_a)R_a\right]\tau^k \\
&\geq \left[(1-\alpha_{a^k})R_{a^k}\right]\tau^k \\
(u_p^k - u_\star^k) &\leq \left[(\alpha_{a^k}-\alpha^k)R_{a^k}\right]\tau^k
\end{aligned}$$

At the same time, this contract has shifted the time-averaged linear contract and hence altered the time-weighted potential.

$$\begin{aligned}
&\psi\left(\mathcal{A}^k/\mathcal{T}^k\right)\mathcal{T}^k - \psi\left(\mathcal{A}^{k-1}/\mathcal{T}^{k-1}\right)\mathcal{T}^{k-1} \\
&= \left[\sum_{i=1}^{\alpha^k}(\alpha_i-\alpha_{i-1})R_{i-1} + \left(\mathcal{A}^k/\mathcal{T}^k - \alpha_{a^k}\right)R_{a^k}\right]\mathcal{T}^k \\
&\quad - \left[\sum_{i=1}^{\alpha^k}(\alpha_i-\alpha_{i-1})R_{i-1} + \left(\mathcal{A}^{k-1}/\mathcal{T}^{k-1} - \alpha_{a^k}\right)R_{a^k}\right]\mathcal{T}^{k-1} \\
&= \left[\mathcal{T}^k\sum_{i=1}^{\alpha^k}(\alpha_i-\alpha_{i-1})R_{i-1} + \left(\mathcal{A}^k - \mathcal{T}^k\alpha_{a^k}\right)R_{a^k}\right] \\
&\quad - \left[\mathcal{T}^{k-1}\sum_{i=1}^{\alpha^k}(\alpha_i-\alpha_{i-1})R_{i-1} + \left(\mathcal{A}^{k-1} - \mathcal{T}^{k-1}\alpha_{a^k}\right)R_{a^k}\right] \\
&= \tau^k\sum_{i=1}^{\alpha^k}(\alpha_i-\alpha_{i-1})R_{i-1} + \left[\left(\alpha^k\tau^k - \tau^k\alpha_{a^k}\right)R_{a^k}\right]
\end{aligned}$$

Interestingly, the expression for time-weighted potential has a term that perfectly cancels with our bound for how much additional principal utility this segment produces over the optimal static contract.

$$\begin{aligned}
(u_p^k - u_\star^k) + \psi\left(\mathcal{A}^k/\mathcal{T}^k\right)\mathcal{T}^k - \psi\left(\mathcal{A}^{k-1}/\mathcal{T}^{k-1}\right)\mathcal{T}^{k-1} &\leq \tau^k\sum_{i=1}^{\alpha^k}(\alpha_i-\alpha_{i-1})R_{i-1} \\
&\leq \int_{\mathcal{T}^{k-1}}^{\mathcal{T}^k}\psi\left(\frac{\mathcal{A}^{k-1}+(\mathcal{T}-\mathcal{T}^{k-1})\alpha^k}{\mathcal{T}}\right)d\mathcal{T}
\end{aligned}$$

The right-hand side expression above is just the integral of the current raw potential as this segment advances the time from $\mathcal{T}^{k-1}$ to $\mathcal{T}^k$. Conveniently, this upper bound still works out to the same amount even if we subdivide our segment (p^k, τ^k, a^k) into two sub-segments $(p^k, x, a^k), (p^k, y, a^k)$ such that $x, y \in [0, \tau^k]$ and $x + y = \tau^k$ (and re-index the other segments appropriately). This means we can sum this bound to get an overall bound for any time $t \in [0, \overline{T}]$, just by splitting the last segment appropriately. To formalize this, we introduce some more parenthetical superscript notation to denote the corresponding objects when considering time from zero to t. In particular, $u_\star^{(t)}$ denotes the optimal static contract's principal utility for t units of time, $\mathcal{A}^{(t)}$ denotes the cumulative linear contract for t units of time.

$$(u_p^{(t)}(\pi) - u_\star^{(t)}) + \psi\left(\mathcal{A}^{(t)}/t\right) t \leq \int_0^t \psi\left(\frac{\mathcal{A}^{(\mathcal{T})}}{\mathcal{T}}\right) d\mathcal{T}$$

Recall our notation where $R_\star$ denotes the optimal static contract's principal utility. For $t \in [\underline{T}, \overline{T}]$, we know that the excess principal utility needs to be at least $\varepsilon R_\star t$, which implies the following.

$$\begin{aligned} \varepsilon R_\star t + \psi\left(\mathcal{A}^{(t)}/t\right) t &\leq \int_0^t \psi\left(\frac{\mathcal{A}^{(\mathcal{T})}}{\mathcal{T}}\right) d\mathcal{T} \\ \psi\left(\mathcal{A}^{(t)}/t\right) &\leq -\varepsilon R_\star + \frac{1}{t}\int_0^t \psi\left(\frac{\mathcal{A}^{(\mathcal{T})}}{\mathcal{T}}\right) d\mathcal{T} \end{aligned}$$

With this bound in mind, we can view every trajectory π that manages to successfully beat the optimal static contract by $(1+\varepsilon)$ in terms of how much raw potential it has as a function of time. Note that this bound controls the current raw potential based on the average raw potential up to this point (minus a constant). As a result, if we just consider trajectories π that obey this bound, the worst case for us would be a function that satisfies it with equality everywhere since greedily picking the maximum value for the function early on allow for higher values later on (greedy stays ahead). We now solve for this function $f(t)$ which simultaneously maximizes raw potential everywhere.

$$\begin{aligned} \varepsilon R_\star t + f(t)t &= \int_0^t f(\mathcal{T}) d\mathcal{T} \\ \varepsilon R_\star + f(t) + f'(t)t &= f(t) \\ f'(t) &= -\varepsilon R_\star / t \end{aligned}$$

At time $\underline{T}$, we know the raw potential can be at most $\psi(\alpha_n)$. We want to choose $\underline{\gamma}$ and hence $\overline{T}$ so that $f(\overline{T})$ is negative in order to create a contradiction. Because f yields the maximum possible function value attainable at time $\overline{T}$, this means that our actual raw potential will also be negative at $\overline{T}$. We now solve for the largest value of γ that does not actually create a contradiction.

$$\begin{aligned} f(\overline{T}) - f(\underline{T}) &= -\psi(\alpha_n) \\ \int_{\underline{T}}^{\overline{T}} f'(t) dt &= -\psi(\alpha_n) \\ -\varepsilon R_\star \left[\ln t\right]_{\underline{T}}^{\overline{T}} &= -\psi(\alpha_n) \\ \ln(\overline{T}/\underline{T}) &= \frac{\psi(\alpha_n)}{\varepsilon R_\star} \\ \gamma &= e^{\phi(\alpha_n)/(\varepsilon R_\star)} \end{aligned}$$

Hence it suffices to pick a $\underline{\gamma} > e^{\phi(\alpha_n)/(\varepsilon R_\star)}$. This demonstrates that it is impossible for a *deterministic* trajectory π to beat the optimal static contract by a $(1+\varepsilon)$ multiplicative factor.

What about *randomized* dynamic contracts $\mathcal{D}$? We can just take the appropriate convex combination of our bounds according to drawing $\pi \sim \mathcal{D}$. In particular, this yields:

$$\mathbb{E}_{\pi\sim\mathcal{D}}\left[(u_p^{(t)}(\pi) - u_\star^{(t)})\right] + \mathbb{E}_{\pi\sim\mathcal{D}}\left[\psi\left(\mathcal{A}^{(t)}/t\right)\right] t \leq \int_0^t \mathbb{E}_{\pi\sim\mathcal{D}}\left[\psi\left(\frac{\mathcal{A}^{(\mathcal{T})}}{\mathcal{T}}\right)\right] d\mathcal{T}$$

We can then re-execute the remainder of the proof in the same way, replacing the deterministic additional principal utility with expected additional principal utility and deterministic raw potential with expected raw potential. The expected potential function is still bounded everywhere by the same function $f(T)$ and we reach the same conclusions about $\underline{\gamma}$. This completes the proof. $\square$

Remark. *Due to Yao's minimax principle, Theorem 4.2 implies that there exists an adversarial distribution over times in $[\underline{T}, \overline{T}]$ such that for any randomized principal strategy, the ratio between expected principal utility and the principal utility of the optimal static contract for that duration of time is strictly less than $(1+\varepsilon)$. In order to apply Yao's minimax principle, we need the set of relevant principal strategies and the set of relevant adversary strategies to be finite. We already do this in our proof of Theorem H.3: the latter can just be an ε-net since principal utility is Lipschitz with Lipschitz constant depending on the contract problem, and after that the former then follows from Carathéodory's Theorem by treating each deterministic trajectory as a vector with one coordinate for every point in our ε-net.*

F Proof of Theorem 4.3 (Unknown Time Horizon – Converse)

Proof. We prove this by proving the contrapositive. Suppose for any fixed time T there is a dynamic contract that can achieve an expected utility of $(1+\epsilon)u_\star T$ for some $\epsilon > 0$. By Theorem 3.1, we can assume without loss of generality that this is a free-fall linear contract. We will show that for any γ we will construct a dynamic contract such that for all $\underline{T} \in \mathbb{R}$ and all $t \in [\underline{T}, \gamma \cdot \underline{T}]$, we can achieve an expected utility of $(1 + f(\epsilon, \gamma)) \cdot u_\star \cdot t$ where $f(\epsilon, \gamma) \geq \Omega\left(\min\left((\frac{\varepsilon}{4})^{O(\log(1+\gamma))}, \frac{\varepsilon}{\gamma}\right)\right)$.

As a first step, we will show that if there is a free-fall linear contract that beats the optimal static contract, then there is a free-fall linear contract that beats the optimal static contract but also either (1) ends at or above the optimal static contract or (2) begins at the optimal static contract. Afterwards, we plan to analyze case (1) and (2) separately.

If our free-fall linear contract does not already satisfy case (1) or (2), then it must do one of the following; (a) begin at a higher breakpoint than the optimal static contract and end at a lower breakpoint than the optimal static contract or (b) being and end at lower breakpoints than the optimal static contract. We now analyze these two cases. In the process, we will lose a constant factor which is folded into our Ω notation.

Case A: Dynamic contracts beginning above $\alpha_\star$ and ending below $\alpha_\star$. We write our free-fall linear contract in the usual form $\pi = \{(\mathbf{p}^k, \tau^k, a^k)\}_{k=1}^K$. By virtue of being in this case, we know there is some index $2 \leq i < K$ such that the average linear contract after i segments, $\overline{\mathbf{p}}^i$, is exactly $\alpha_\star$. We "cut" the trajectory π at this point to produce two new trajectories π' and π''. Specifically, $\pi' = \{(\mathbf{p}^k, \tau^k, a^k\}_{k=1}^i$ and $\pi'' = \{(\alpha_\star, \mathcal{T}^i, a^i)\} \circ \{(\mathbf{p}^k, \tau^k, a^k\}_{k=i+1}^K$ where $\circ$ denotes concatenation. In other words, we construct π' by ending at this point and we construct π'' by taking the optimal static contract to this point and continuing as normal. Observe that the combined performance of π' and π'' is equal to the combined performance of π and just playing the single segment $\{(\alpha_\star, \mathcal{T}^i, a^i)\}$: $(1+\epsilon)u_\star\mathcal{T}^K + u_\star\mathcal{T}^i$. This results in a combined time-averaged performance of

$$\begin{aligned}\frac{(1+\epsilon)u_\star\mathcal{T}^K + u_\star\mathcal{T}^i}{\mathcal{T}^K + \mathcal{T}^i} &= u_\star\left[(1+\epsilon)\frac{\mathcal{T}^K}{\mathcal{T}^K + \mathcal{T}^i} + (1)\frac{\mathcal{T}^i}{\mathcal{T}^K + \mathcal{T}^i}\right] \\ &\geq (1+\epsilon/2)u_\star\end{aligned}$$

since $\mathcal{T}^K \geq \mathcal{T}^i$. Since π' and π'' have this combined average, one of them must have at least this average (and we only lost a factor $1/2$ on our ϵ, which is indeed a constant. Since π' matches case (1) and π'' matches case (2), this completes the analysis of case (a).

Case B: Dynamic contracts beginning and ending below $\alpha_\star$. We take the obvious approach and choose to begin at $\alpha_\star$ instead. Specifically, we replace the first segment with a sequence of segments that begins at $\alpha_\star$ and then undergoes the appropriate number of free-fall segments to arrive at the same endpoint as before (same total time and average linear contract). We argue that each new segment has at least as much principal utility per unit time as the original segment. Since the total time is the same, this is a direct improvement over the original dynamic contract, both in terms of total principal utility and time-averaged principal utility. The argument that each new segment does at least as well per unit time is similar to before. The first new segment just hovers at the optimal static contract, which by definition is better than any other static contract (which our original segment must be). The remaining new segments are freefall segments, and achieve principal utility per unit time equal to the expected revenue of the actions they fall through. We observe that we fall through segments in order of decreasing expected utility, meaning all of these segments have higher expected utility than the action we originally began with, and expected revenue is at least the principal utility of the static contract that achieves a particular action. We finish this case by noting that we did not diminish ϵ at all, which trivially a constant factor.

This completes our analysis of cases (a) and (b). In all cases, we managed to reduce to either case (1) or (2), which we now consider.

Case 1: Dynamic contracts ending at or above $\alpha_\star$. First, we consider the case where for any fixed T there is a dynamic contract $\pi(T) = ((\alpha^1, \tau^1, a^1), \dots, (\alpha^k, \tau^k, a^k))$ which ends at or above the optimal static action: $a^k \geq a_\star$. Given any γ and time period $[\underline{T}, \overline{T} = \gamma \cdot \underline{T}]$, consider the dynamic contract which starts with $\pi(\underline{T})$, free falls to the optimal static contract, and then plays the optimal static contract for the remaining time period. We again observe (as we did for case (b)) that free falling through actions that are at least the $a_\star$ results in at least $u_\star$ principal profit per unit time. Hence the total revenue for any time $t \in [\underline{T}, \overline{T}]$ for the principal is $(1+\epsilon) \cdot u_\star \underline{T} + (t - \underline{T}) \cdot u_\star$, which is at least $(1 + \epsilon/\gamma) R_\star t$.

Case 2: Dynamic contracts starting in $\alpha_\star$. By Theorem 3.1, we know any dynamic contract can be transformed into a free-fall dynamic contract with no loss in revenue. Therefore, we assume that for any fixed time horizon T, there is a dynamic contract form $\pi(T) = \big((\alpha_\star, \tau^1, a^1), (0, \tau^2, a^2), \dots, (0, \tau^k, a^k)\big)$ which achieves a total revenue of $(1+\varepsilon) R_\star T$. Since this is a free-fall contract, the optimal revenue from this contract can be characterized as $(1-\alpha_\star) R_\star \tau^1 + \sum_{i=2}^k \tau^i R_{a_i}$ which is at least $(1+\varepsilon) R_\star t > (1+\varepsilon)(1-\alpha_\star) R_\star$. Let μ be the minimum fraction of time such that for any time T, the dynamic contract $\pi(T)$ achieves revenue at least $(1+\varepsilon/2)\mu u_\star T$. Since we know that $\pi(T)$ achieves a total revenue of $(1+\varepsilon) u_\star T$ and starts out at the optimal static contract, we know that $\mu \geq \tau^1 / \sum_{i=1}^k \tau^i$ and it is a constant bounded away from 1. Let $S_i = \lceil \mu^i \overline{T} \rceil$ and let p be the first index where S_p is less than $\underline{T}$ (i.e., $p = \lceil \frac{\log(1+\gamma)}{\log(\mu)} \rceil$) . By construction, S_i satisfy two properties:

1. $S_p \leq \underline{T} \leq S_{p-1} \leq \dots S_1 \leq \overline{T}$.
2. If the principal runs dynamic contract ending at S_i, namely $\pi(S_i)$, then they are guaranteed revenue $(1 + \varepsilon/2) t R_\star$ for any $t \in [S_{i+1}, S_i]$.

We will construct a sequence of dynamic contracts π^i which have the property that for any $t \in [S_i, \gamma \underline{T}]$ achieves revenue that is at least $(1 + (\varepsilon/4)^i) R_\star t$. We do this via induction. For the base case, let $\pi^1 = \pi(\overline{T})$. By construction, we know that for all $t \in [S_1, \overline{T}]$, the principal will get revenue $(1+\varepsilon/2) u_\star t$. Now suppose we have such a dynamic contract π^i, then we construct π^{i+1} by taking a convex combination of π^i and the optimal dynamic contract ending at $\pi(S_i)$. In particular, let

$$\pi^{i+1} = \frac{1+\varepsilon/2}{1+\varepsilon/2 + (\varepsilon/4)^i} \pi^i + \frac{(\varepsilon/4)^i}{1+\varepsilon+(\varepsilon/4)^i} \pi(S_i).$$

For any $t \in [S_i, \overline{T}]$, we have that revenue we attain is at least the revenue from the contract

$$\frac{1+\varepsilon/2}{1+\varepsilon/2+(\varepsilon/4)^i} \text{Revenue}(\pi^i(t)) \geq \frac{1+\varepsilon/2}{1+\varepsilon/2+(\varepsilon/4)^i} (1 + (\varepsilon/4)^i) u_\star t \geq$$
$$1 + \frac{\varepsilon^{i+1}/2 \cdot 4^i}{1+\varepsilon/2+(\varepsilon/4)^i} \geq (1 + (\varepsilon/4)^{i+1}) u_\star t.$$

For any $t \in [S_{i+1}, S_i]$, observe that we get at least $u_\star t$ from the first contract π^i and at least $(1+\varepsilon/2) u_\star t$ in the second contract. Therefore we get at least

$$\frac{1+\varepsilon/2}{1+\varepsilon/2+(\varepsilon/4)^i} u_\star t + \frac{(\varepsilon/4)^i}{1+\varepsilon/2+(\varepsilon/4)^i} (1 + (\varepsilon/2)) u_\star t \geq (1 + (\varepsilon/4)^i) u_\star t.$$

□

G General Contracts

In this section, we give a general contract instance with $n = 4$ actions (3 non-null actions) and $m = 4$ outcomes (3 non-null outcomes), where the best dynamic contract provably outperforms the best free-fall dynamic contract. The instance in question is defined as follows:

- The cost vector $c = (c_1, c_2, c_3, c_4) = (0, 0.2, 0.4, 0.5)$.
- The reward vector $r = (r_1, r_2, r_3) = (0, 1.0, 1.6, 2.0)$.
- The forecast matrix is given by $F = \begin{pmatrix} 1.00 & 0.00 & 0.00 & 0.00 \\ 0.45 & 0.20 & 0.25 & 0.10 \\ 0.35 & 0.05 & 0.25 & 0.35 \\ 0.15 & 0.30 & 0.30 & 0.25 \end{pmatrix}$

This instance was found by a programmatic search over a large collection of instances. For this instance, we can (again, programmatically) compute that the best free-fall dynamic contract achieves a net asymptotic utility for the principal of at most 0.753 per round. At the same time, we can exhibit a non-free-fall dynamic contract for this instance that achieves a utility of at least 0.764 per round. For conciseness, we present the details of our approach in Appendix G.1, where we construct well-tailored linear programs that provide the aforementioned intricate instance.

G.1 Programmatic LP Search for Sub-Optimal Free Fall Against Non-Linear Contracts

At a high level, the verification of the example of section 3.3 relies on the following fact: given a sequence of actions $(a^1, a^2, \ldots, a^K)$, we can construct a polynomial-sized linear program to find the optimal continuous-time dynamic (general or free-fall) contract $\{(\mathbf{p}^k, \tau^k, a^k)\}_{k=1}^K$ with this specific action sequence.

The variables of this LP are the τ^k and $\mathbf{p}^k$ corresponding to each action a^k. The constraints follow from the definition of a valid trajectory of play in Section 2.2 and are as follows:

- **(Non-negativity)** $\mathbf{p}^k, \tau^k \geq 0$.
- **(Time normalization)** $\sum_{k=1}^K \tau^k = 1$. We normalize the total duration of the trajectory to 1.
- **(Beginning of segment is best response)** $\sum_{k'=1}^{k-1} \tau^{k'} u_L \left(\mathbf{p}^{k'}, a^k\right) \geq \sum_{k'=1}^{k-1} \tau^{k'} u_L \left(\mathbf{p}^{k'}, a'\right)$ for any $a' \in [n]$. This represents the constraint $a^k \in \mathsf{BR}(\overline{\mathbf{p}}^{k-1})$.
- **(End of segment is best response)** $\sum_{k'=1}^{k} \tau^{k'} u_L \left(\mathbf{p}^{k'}, a^k\right) \geq \sum_{k'=1}^{k} \tau^{k'} u_L \left(\mathbf{p}^{k'}, a'\right)$ for any $a' \in [n]$. This represents the constraint $a^k \in \mathsf{BR}(\overline{\mathbf{p}}^{k})$.

The objective of the LP is the optimizer utility $\sum_{k=1}^K \tau^k u_O(\mathbf{p}^k, a^k)$. If we want to further impose that the contract is a free-fall contract, we can add the constraint that $\mathbf{p}^k = 0$ for $k > 1$.

For free-fall contracts, we have an additional constraint on what sequences of actions are possible. Note that a free-fall contract will never repeat an action – in particular, after the initial segment, the cumulative utility of each action $i \in [n]$ decreases by c_i per round, so the sequence of actions $(a^1, a^2, \ldots, a^K)$ a free-fall contract passes through must be sorted in *decreasing order of cost*. This means there are at most 2^n sequences of actions to check, and by checking all of them we can provably compute the optimal free-fall contract for a given general contract instance.

On the other hand, it's not clear if there are any constraints on how complex the sequence of actions for the optimal general dynamic contract can be – it is an interesting open question whether there exists any efficient (or even computable) algorithm for computing $U^\star$ for a general contract instance. Luckily, in order to show this separation, we need only exhibit a single general contract which outperforms the best free-fall contract. In the example above, we compute the best general contract for the same action sequence that the optimal free-fall contract passes through, and observe that the general contract obtains strictly larger utility.

H Simplifying Tool: Reductions from Discrete to Continuous Time

H.1 Proof of Theorem 2.4

In this section we prove Theorem 2.4, showing that instead of working with discrete-time learning algorithms, it instead suffices to work with the set of continuous-time trajectories piecewise-linear trajectories described in Section 2.2. Our proof will generally follow the proof structure of [26] (which proves a similar reduction in the case of two-player bi-matrix games), with a few slight additional complexities due to some differences in notation (namely, we do not insist that every segment lies in the interior of a best-response region).

Before we begin the proof, it will be useful to establish a helpful auxiliary lemma about trajectories. Call a segment (p^k, τ^k, a^k) of a trajectory π *degenerate* if it lies on the boundary of two best-response regions (i.e., $|\mathsf{BR}(\overline{\mathbf{p}}^{k-1}) \cap \mathsf{BR}(\overline{\mathbf{p}}^k)| \geq 2$), and *non-degenerate* otherwise. Let $\mathsf{Util}_0(\pi)$ be the utility contributed by just non-degenerate segments. We begin by showing that starting with any trajectory π, we can construct a mostly non-degenerate trajectory π_0 with $\mathsf{Util}_0(\pi_0)$ almost as large as $\mathsf{Util}(\pi)$.

Lemma H.1. *For any trajectory π and any $\varepsilon > 0$, there exists a trajectory π_0 such that* $\mathsf{Util}_0(\pi_0) \geq (1-\varepsilon)\mathsf{Util}(\pi)$.

Proof. Let $\pi = \{(p^k, \tau^k, a^k)\}$. We will produce π_0 by interleaving a sequence of small perturbations (q^k, δ^k) into π for some $q^k \in \mathbb{R}^m_{\geq 0}$ and $\delta^k > 0$; that is, we will let π_0 be defined by the sequence of segments $(q^1, \delta^1), (p^1, \tau^1, a^1), (q^2, \delta^2), \dots, (q^k, \delta^k), (p^k, \tau^k, a^k)$. Note that we have not specified the best-response of the learner for the perturbation segments (q^k, δ^k), because we will not count the utility from these segments (in fact, these perturbation segments might cross best-response boundaries, in which case we can split them into smaller segments). We will show that if we choose q^i and δ^i correctly, a^k is the unique best-response for each of the shifted (p^k, τ^k, a^k) segments.

Without loss of generality, assume $\sum_k \tau^k = 1$. For any $t \in [0,1]$, we will let $\overline{\mathbf{p}}(t)$ be the average contract at time t under trajectory π. That is, if $t = \tau^1 + \tau^2 + \cdots + \tau^{i-1} + \tau$ with $0 \leq \tau < \tau^i$, then

$$\overline{\mathbf{p}}(t) = \frac{\tau^1 p^1 + \tau^2 p^2 + \cdots + \tau^{i-1} p^{i-1} + \tau p^i}{t}.$$

For each $i \in [k]$, we will also let $\Delta^i = \delta^1 + \delta^2 + \cdots + \delta^i$, and $Q^i = (\delta^1 q^1 + \delta^2 q^2 + \cdots + \delta^i q^i)/\Delta^i$. Now, if $t = \tau^1 + \tau^2 + \cdots + \tau^{i-1} + \tau$ with $0 \leq \tau < \tau^i$, we will let $\overline{\mathbf{p}}_0(t)$ be the average contract under trajectory π_0 at time $\Delta^i + \tau$ (i.e., time τ into segment (p^i, τ^i, a^i)). It is the case that for such t,

$$\overline{\mathbf{p}}_0(t) = \frac{Q^i \Delta^i + t\overline{\mathbf{p}}(t)}{\Delta^k + t} = \overline{\mathbf{p}}(t) + \frac{\Delta^i}{\Delta^i + t}(Q^i - \overline{\mathbf{p}}(t)).$$

We would like to choose Q^i and Δ^i such that for each $i \in [k]$, for a large sub-interval of $\tau \in [0, \tau^i)$, the unique best response to $\overline{\mathbf{p}}_0(t)$ is exactly a^i. To begin, note that for any sequence of *strictly positive contracts* $Q^i \in \mathbb{R}^m_{>0}$, there is a sequence of q^i and δ^i that implements it (because we can make each Q^i any convex combination of Q^{i-1} and q^i). Moreover, we can make Δ^k arbitrarily small, because scaling all the δ^i simultaneously does not affect the values of the Q^i.

Now, for each i, we will set Q^i to a positive contract that uniquely incentivizes action a^i. Note that a non-negative contract exists by our assumption in Section 2; but since infinitesimal perturbations maintain the property that the contract uniquely incentivizes a^i, there must also be a positive contract with this property. We claim that if $\mathsf{BR}(Q^k) = \{a^k\}$ and $a^k \in \mathsf{BR}(\overline{\mathbf{p}}(t))$, then for any $\lambda \leq 1$, $\mathsf{BR}(\overline{\mathbf{p}}(t) + \lambda(Q^k - \overline{\mathbf{p}}(t))) = \{a^k\}$. To see this, note that we can write $\overline{\mathbf{p}}(t) + \lambda(Q^k - \overline{\mathbf{p}}(t)) = (1-\lambda)\overline{\mathbf{p}}(t) + \lambda Q^k$. Since the utility of the agent is an affine linear function in the contract they are offered, for any action $a' \neq a$ we have that $u_A((1-\lambda)\overline{\mathbf{p}}(t) + \lambda Q^k, a^k) = (1-\lambda)u_A(\overline{\mathbf{p}}(t), a^k) + \lambda u_A(Q^k, a^k) > (1-\lambda)u_A(\overline{\mathbf{p}}(t), a') + \lambda u_A(Q^k, a') = u_A((1-\lambda)\overline{\mathbf{p}}(t) + \lambda Q^k, a^k)$.

It follows that if we choose the Q^i in this way, $\mathsf{BR}(\overline{\mathbf{p}}_0(t)) = \{a^k\}$, and therefore each of the segments (p^i, τ^i, a^i) is non-degenerate. We will set Δ^k equal to ε. Doing so, we have that:

$$\mathsf{Util}_0(\pi_0) \geq \frac{\sum_{i=1}^k u_P(p^i, a^i)\tau^i}{1 + \Delta_k} \geq (1-\varepsilon)\mathsf{Util}(\pi)$$

□

Equipped with Lemma H.1, we can now prove Theorem 2.4.

Proof of Theorem 2.4. We follow the proof structure of [26] and prove both parts separately.

Part 1. Let $\pi = \{(p^k, \tau^k, a^k)\}_{k=1}^K$ represent a valid strategy for the principal in the continuous-time problem. Without loss of generality, assume $\sum_k \tau^k = 1$ (if not, we can divide all τ^k through by $\sum_k \tau_k$ without changing the value of this strategy). We will convert π into the following discrete-time strategy for the principal: for each $i \in [K]$ (in order), the principal offers the contract p^k for $\tau^k T$ rounds.

Our goal is to show that for any $\delta > 0$ and any mean-based algorithm $\mathcal{A}$, the above strategy results in at least $(\mathsf{Util}_0(\pi) - \delta)T - o(T)$ utility for the optimizer. The conclusion then follows by choosing a trajectory π for which $\mathsf{Util}_0(\pi) > U^\star - \varepsilon/2$ (such a π exists by Lemma H.1 and the definition of $U^\star$) and some $\delta < \varepsilon/2$. In the remainder of this proof, we will fix a specific mean-based algorithm $\mathcal{A}$ that is $\gamma(T)$-mean-based for some $\gamma(T) = o(1)$.

As in Definition 2.2, let $\sigma_{i,t}$ denote the aggregate utility of action i to the agent over the first t rounds. Let $T^k = \sum_{j=1}^k \tau^j T$, and consider the values of σ_t for rounds $t \in [T^{k-1}, T^k]$ corresponding to the kth segment. Note that σ_t is linear in this interval and so we can interpolate

$$\sigma_t = \frac{(t - T^{k-1})\sigma_{T^{k-1}} + (T^k - t)\sigma_{T^k}}{\tau^k T}. \tag{1}$$

Furthermore, assume that segment k is non-degenerate, and so $\mathsf{BR}(\overline{\mathbf{p}}^{k-1}) \cap \mathsf{BR}(\overline{\mathbf{p}}^k) = \{a^k\}$. In particular, for any $t \in [T^{k-1}, T^k]$ and $a' \neq a_k$, either $\sigma_{T^{k-1},a^k} > \sigma_{T^{k-1},a'}$ or $\sigma_{T^k,a^k} > \sigma_{T^k,a'}$. As a consequence of (1), this means that for any $\varepsilon_k > 0$, there exists a $\delta_k > 0$ such that for $t \in [T^{k-1} + \varepsilon_k \tau^k, T^k - \varepsilon_k \tau^k]$, $\sigma_{t,a^k} \geq \sigma_{t,a'} + \delta_k T$. For sufficiently large T, $\delta_k T > \gamma(T)T$, and so the learner will put weight at least $(1 - n\gamma(T))$ on action a^k. The total utility of the principal from these rounds is therefore at least

$$(1 - n\gamma(T))(1 - 2\varepsilon_k)\tau^k u_P(p^k, a^k) \geq \tau^k u_P(p^k, a^k) - (n\gamma(T) + 2\varepsilon_k)T. \tag{2}$$

Summing over all non-degenerate segments k, we find the total utility of the principal is at least

$$\sum_k \tau^k u_P(p^k, a^k) - \sum_k k(n\gamma(T) + 2\varepsilon_k)T = \mathsf{Util}_0(\pi) - \sum_k k(n\gamma(T) + 2\varepsilon_k)T.$$

By choosing ε_k sufficiently small, we can guarantee that this is at least $\mathsf{Util}_0(\pi) - \delta T$ for sufficiently large T, as desired.

Part 2. Fix any $\varepsilon > 0$. Assume that for some sufficiently large T_0, there exists a (possibly adaptive) dynamic strategy for the principal that guarantees utility at least $(U^\star + \varepsilon)T_0$ against every mean-based agent. We will show that this implies the existence of a continuous trajectory π and $\mathsf{Util}(\pi) \geq U^\star + \varepsilon$, contradicting the definition of $U^\star$. Fix $\gamma(T) = T^{-1/2}$ and at any time t, let $J_t = \{j \in [n] | (\max_i \sigma_{t,i}) - \sigma_{t,j} < \gamma(T)T\}$ be the set of actions for the learner whose historical performance are within $\gamma(T)T$ of the optimally performing action. The set J_t contains exactly the set of actions that the mean-based guarantee implies the agent must play with high probability. Our agent will do the following: if the principal is about to play contract p^t, the agent will play the action $j \in J_t$ that minimizes $u_L(p^t, j)$ (note that because we are tailoring the agent to this principal, we can do this).

Assume that this results in the principal playing the sequence of contracts $p^1, p^2, \ldots, p^{T_0}$. Consider the trajectory π defined by the sequence of tuples $(p^1, 1/T_0), (p^2, 1/T_0), \ldots, (p^{T_0}, 1/T_0)$. In this description of the trajectory, we've omitted the response action for the agent, which can be any best-response action for that segment. In fact, some segments may not be valid, as they start in one best response region and end in another; for those, we can subdivide them into however many parts are necessary to form a valid trajectory.

Now, note that the sub-segments corresponding to the step $(p^t, 1/T_0)$ only contain agent actions in the set J_t. This is since the agent utility at the start of this segment is σ_t, the agent utility at the end of this segment is σ_{t+1}, each component of $\sigma_{t+1} - \sigma_t$ is at most 1 (since the problem is bounded), but any action j not in J_t is at least $\gamma(T)$ away from optimal. The principal's utility contributed by this segment is therefore at least $\frac{1}{T_0} \min_{j \in J_t} u_P(p^t, j)$. But this is exactly the utility the principal obtained in round t of the discrete-time game. Therefore the total utility $\mathsf{Util}(\pi)$ of this trajectory is at least $U^\star + \varepsilon$ – but this contradicts the definition of $U^\star$, as desired. $\square$

We will need the following lemma which says we can restrict our attention to finite-support $\mathcal{D}$ without loss of generality.

Lemma H.2. *Fix a principal-agent problem, a $\gamma > 1$, and an $\varepsilon > 0$. Let $\mathcal{D}$ be any distribution over trajectories. Then there exists a finite-support distribution $\mathcal{D}'$ over trajectories with the property that $\mathsf{Util}_\gamma(\mathcal{D}') \geq \mathsf{Util}_\gamma(\mathcal{D})$.*

Proof. We first claim the following: if a distribution $\mathcal{D}$ has the property that $\mathbb{E}_{\pi\sim\mathcal{D}}[u_P^{(t)}(\pi)] \geq Ut$ for each t in the *discretized* set of time-intervals $S_{\varepsilon,\gamma} = \{1/\gamma, 1/\gamma + \varepsilon/\gamma, 1/\gamma + 2\varepsilon/\gamma, \ldots, 1 - \varepsilon/\gamma, 1\}$, then it is the case that $\mathbb{E}_{\pi\sim\mathcal{D}}[u_P^{(t)}(\pi)] \geq (U - \varepsilon)t$ for all $t \in [1/\gamma, 1]$. This follows from the fact that the principal's profit per round is bounded above by 1, so $|u_P^{(t')}(\pi) - u_P^{(t)}(\pi)| \leq |t' - t|$. In particular, if t' is the closest element of $S_{\gamma,\varepsilon}$ to a $t \in [1/\gamma, 1]$, it is the case that $|u_P^{(t')}(\pi) - u_P^{(t)}(\pi)| \leq \varepsilon/\gamma \leq \varepsilon t$.

Now, associate to each trajectory π the $|S_{\varepsilon,\gamma}|$-tuple of real numbers $f(\pi) = \{u_P^{(t)}(\pi)\}_{t\in S_{\varepsilon,\gamma}}$; define $f(\mathcal{D}) = \mathbb{E}_{\pi\sim\mathcal{D}}[f(\pi)]$. Define $\mathcal{X} = \{f(\pi) \mid \pi \text{ is a trajectory}\} \subset \mathbb{R}^{|S|}$ to be the set of all such tuples. By Caratheodory's theorem, we can construct a distribution over at most $|S| + 1$ elements of $\mathcal{X}$ that (is arbitrarily close to) $f(\mathcal{D})$, for any $\mathcal{D}$. If we let $\mathcal{D}'$ be the corresponding distribution over trajectories, this satisfies the constraints of the theorem statement. $\square$

H.2 Reduction from Discrete to Continuous Time with Unknown Time Horizons

In this section, we extend the previous reduction to the case where the time horizon can belong to an interval. One of the biggest differences is the introduction of this parameter $\gamma \geq 1$ which equals the multiplicative ratio $(\overline{T}/\underline{T})$. Instead of a trajectory $\pi = \{(p^k, \tau^k, a^k\}_{k=1}^K$ being solely evaluated at its end time $\mathcal{T}^K$, we now care about its performance over its final interval of multiplicative width γ, namely $[\frac{1}{\gamma}\mathcal{T}^K, \mathcal{T}^K]$.

In order to quantify the performance of a trajectory at a certain time t, we will introduce some corresponding parenthetical superscript notation. In particular, $u_P^{(t)}(\pi)$ will denote the cumulative expected principal utility of trajectory π from time zero to t, and is formally defined as

$$u_P^{(t)}(\pi) \triangleq \begin{cases} \sum_{k=1}^{k'-1} \tau^k u_P(p^k, a^k) + (t - \mathcal{T}^{k'})u_P(p^{k'}, a^{k'}) & \text{if } t \in [\mathcal{T}^{k'}, \mathcal{T}^{k'+1}) \\ \sum_{k=1}^{k'-1} \tau^k u_P(p^k, a^k) & \text{if } t = \mathcal{T}^K. \end{cases}$$

Then, the worst-case (under possible time horizons) expected (under drawing from the distribution and actions producing random outcomes) utility of the principal for distribution $\mathcal{D}$ is given by

$$\mathsf{Util}_\gamma(\mathcal{D}) = \min_{x\in[1/\gamma,1]} \mathbb{E}_{\pi\sim\mathcal{D}} \frac{u_P^{(x\mathcal{T}^K)}(\pi)}{x\mathcal{T}^K},$$

where each $\mathcal{T}^K$ is according to the drawn trajectory π.

Finally, let $U_\gamma^\star = \sup_{\mathcal{D}} \mathsf{Util}_\gamma(\mathcal{D})$, where the sup runs over all distributions of valid trajectories of arbitrary finite length. We can think of $U_\gamma^\star$ as the maximum possible worst-case utility of the principal in the unknown time horizon continuous setting game.

Theorem H.3. *Fix any principal-agent problem and $\gamma \geq 1$. We have the following two results:*

1. *For any $\varepsilon > 0$, there exists an oblivious strategy for the principal that gets at least $(U_\gamma^\star - \varepsilon)t - o(t)$ utility for the principal for all $t \in [T, \lceil\gamma T\rceil]$ for sufficiently large T.*

2. *For any $\varepsilon > 0$, there exists a mean-based algorithm $\mathcal{A}$ such that no (even adaptive[12]) principal can get more than $(U_\gamma^\star + \varepsilon)t + o(t)$ utility against an agent running $\mathcal{A}$ for all $t \in [T, \lceil\gamma T\rceil]$ for any T.*

[12]As with the known time-horizon result, this holds against adaptive principals in the full-feedback setting, or if the principal is deterministic. See Appendix H.3 for details.

Proof of Theorem H.3. **Part 1.** Begin by picking a strategy $\mathcal{D}$ for the optimizer in the continuous-time game that achieves utility at least $U^\star_\gamma - \varepsilon/2$. This strategy $\mathcal{D}$ is a distribution over trajectories π; by Lemma H.2, we can assume (at the cost of losing an arbitrarily small $O(\varepsilon)$ term) that $\mathcal{D}$ has finite support. For each of these trajectories, we apply Lemma H.1 to transform π into a new trajectory π' which obtains at least $(1-\varepsilon)$ fraction of the utility of π on non-degenerate segments. We will also normalize the total duration of each π' to 1.

Now, note that since inequality (2) holds per segment (and indeed, even fractionally per segment), we can convert each resulting trajectory π' to a discrete-time strategy over $\overline{T}$ rounds, with the property that for sufficiently large values of $\overline{T}$, for any $t \in [\underline{T} = \overline{T}/\gamma, \overline{T}]$, the utility of this discrete-time strategy until time t is at least $u_P^{(t)}(\pi)$. Taking the corresponding distribution over these discrete-time strategies (choosing a sufficiently large $\overline{T}$ for all such strategies – note that we can do this because $\mathcal{D}$ has finite support), we obtain a discrete-time randomized (but otherwise oblivious) strategy for the principal that satisfies the theorem statement.

Part 2. As in the previous proof, fix an $\varepsilon > 0$, assume to the contrary there exists a T_0 along with a discrete-time (possibly randomized / adaptive) dynamic strategy which achieves at least $(U^\star_\gamma + \varepsilon)t$ utility for the principal for all $t \in [T_0/\gamma, T_0]$ against any mean-based bidder. Construct the same mean-based bidder as in the proof of part 2 of Theorem 2.4, which always picks the action in the set of approximate best-responses that least to the minimum expected utility for the principal.

When this principal plays against this agent, this leads to a distribution over sequences of contracts $(p^1, p^2, \dots, p^{T_0})$. Each such sequence can be converted to a trajectory π of the form $\{(p^1, 1/T_0), (p^2, 1/T_0), \dots, (p^{T_0}, 1/T_0)\}$. This trajectory π not only has the property that $\mathsf{Util}(\pi)T$ upper bounds the utility of the discrete-time agent (as in the proof of part 2 of Theorem 2.4), but in fact $u_P^{(t)}(\pi)$ is at least the utility of the agent discrete-time agent at time tT_0 (by exactly the same logic). It follows that if we let $\mathcal{D}$ be the distribution over such trajectories, it is the case that $\mathsf{Util}_\gamma(\mathcal{D}) \geq U^\star_\gamma + \varepsilon$. This contradicts the definition of $U^\star_\gamma$, as desired. □

Finally, we conclude this supplementary section with the proof of a preliminary lemma exploited in Section 3.1

Lemma H.4. *(Restated Lemma B.2) Consider any dynamic contract. For any time interval in which a mean-based agent plays a single action, we can replace the contracts in this interval with their average and obtain overall a revenue-equivalent dynamic contract.*

Proof. The result follows since the utility for the principal u_P is affine in its first argument. Formally, let $\pi = \{(p^k, \tau^k, a^k)\}_{k=1}^K$ be a dynamic contract, with $a^k = a^{k+1} = a$ for some k. Consider a different contract π' where we replace the consecutive pair of segments (p^k, τ^k, a^k) and $(p^{k+1}, \tau^{k+1}, a^{k+1})$ with the their average segment, i.e., $(\overline{p}, \tau^k + \tau^{k+1}, a)$, where $\overline{p} = (p^k\tau^k + p^{k+1}\tau^{k+1})/(\tau^k + \tau^{k+1})$, and all other segments remain the same as in π. Then, we have $\mathsf{Util}(\pi') - \mathsf{Util}(\pi) = \frac{1}{\sum_{k=1}^K \tau^k}\left(\tau^k u_P(p^k, a) + \tau^{k+1} u_P(p^{k+1}, a) - (\tau^k + \tau^{k+1})\overline{p}\right) = 0$. That is, both contracts give same utility for the principal. A similar argument holds for the discrete formulation of the model as well. □

H.3 Mean-Based Algorithms in the Partial-Feedback Setting

We conclude with some clarifying remarks on the definition of a mean-based learning algorithm in a stochastic, partial-feedback setting (the bandits setting). The proofs of Theorems 2.4 and H.3 continue to hold essentially as written, but there are some subtleties that are worth pointing out.

We begin by clarifying the definition of mean-based in a partial-information setting. Formally, we write it as follows. Recall that $\sigma_i^t = \sum_{t'=1}^{t-1} u_i^{t'}$ is equal to the expected utility the learner would receive if they had played action i for the first $t-1$ rounds, *assuming the sequence of contracts the principal offers the learner remains static* (so in particular, for an adaptive / stochastic principal, σ^t is a random variable).

Definition H.5. *(Mean-based algorithms in partial-information settings) A learning algorithm* in a partial-information setting *is* $\gamma(T)$-mean-based *if the following conditions hold: Fix any adaptive dynamic strategy of the principal and let (for each round $t \in [T]$), X_t be the event that $\sigma_i^t <$*

$\sigma^t_{i'} - \gamma(T) \cdot T$, and Y_t be the event that the algorithm takes action i in round t. Then the algorithm is mean based if the probability $\Pr[X_t \wedge Y_t]$ (over all randomness in the learner's algorithm, principal's strategy, and problem setting) is at most $\gamma(T)$. We say an algorithm is mean-based *if it is $\gamma(T)$-mean-based for some $\gamma(T) = o(1)$.*

The above definition is very similar to Definition 2.2; the main reason for stating it like this is to avoid implying the slightly stronger constraint that event X_t deterministically implies that the probability of Y_t is small conditioned on the current history of play. This implication is fine in the full-information setting where algorithms like multiplicative weights will indeed deterministically place small weight on action i' if the event X_t holds; but in the partial-information setting, there is always a chance that the learner is unable to accurately observe whether X_t holds, and therefore no partial-information algorithm can achieve that guarantee. On the other hand, standard bandit algorithms with high-probability guarantees such as EXP3 (see [16]) satisfy the above definition of mean-based learning.

The proof of Theorem 2.4 works equally well with Definition H.5. The only subtlety is in Part 2, where to show a principal cannot do well against all mean-based agents, we design a mean-based agent that foils this specific principal. If the principal is randomized and adaptive, the agent cannot accurately predict the expected contract p^t the principal will play in round t (note that if the principal is adaptive but deterministic, the agent can still simulate the principal's behavior – likewise, if the principal is oblivious and randomized, the agent can compute the expected contract p^t at any round). The proof of Theorem H.3 is similar.

I Final Remarks

The following are observations about our repeated contract games with learning agents that arise from our analysis and from known results on learning agents in general games.

Observation I.1. *In the fixed contract setting, for any regret-minimizing agent in the limit $T \to \infty$ the support of the average empirical distribution of play includes only best-response actions with probability one. Therefore, the repeated game with a static contract against a regret-minimizing agent is essentially equivalent to the single-shot game against a rational agent.*

Proof. This follows directly from the regret-minimization property. Indeed, suppose, for the sake of contradiction, that there exists an action a in the support which is not a best response. Denote the best-response utility by OPT. Action a is played with probability $p > 0$. Notice that since there is only one player, the regret from any other action cannot be negative. Then we have that the regret is $\text{Regret} \geq p(OPT - u(a))T = \mathcal{O}(T)$, a contradiction. □

Observation I.2. *If the agent is using a no-swap-regret algorithm, then the optimal static contract played repeatedly is also optimal in the dynamic setting. As a corollary, this is the case also for general no-regret algorithms if the agent has at most two actions.*

Proof. The result follows from [26], who show that in any game between an optimizer and a no-swap-regret algorithm, the optimizer cannot extract higher payoff than the Stackelberg value of the game where the optimizer plays the first move. The corollary is since with (at most) two actions internal regret and external regret are equivalent. □

Below we show that in our analysis of dynamic linear contracts, it suffices to only examine linear contracts with $\alpha \in [0, 1]$. Note that although this is obvious in the static setting (offering $\alpha > 1$ requires the principal to suffer negative utility), it is not a priori clear that the principal cannot benefit via a dynamic strategy which offers a contract with $\alpha > 1$ for some fraction of the time horizon (perhaps counterbalancing it by offering a contract with a much smaller α later on). In fact, [41] show that when the agents have private information ("types") the principal *can* benefit by offering a randomized menu of linear contracts which possibly contains linear contracts with $\alpha > 1$.

Nonetheless, we show that the principal cannot benefit by doing this in the dynamic setting. The proof below follows from a slight modification of Lemma B.3 in our proof of Theorem 3.1.

Observation I.3. *Let $\pi = \{(\alpha^k, \tau^k, a^k)_{k=1}^K\}$ be any linear dynamic contract with some linear contract $\alpha^i > 1$. Then there exists a dynamic linear contract $\pi' = \{(\alpha^k, \tau^k, a^k)\}_{k=1}^K$ with $\mathsf{Util}(\pi') \geq \mathsf{Util}(\pi)$ and where $\alpha^k \leq 1$ for all k.*

Proof. We first observe that in Lemma B.3, when an agent is indifferent between actions i and $i+1$ then the change in utility for the principal by choosing an action $i+1$ over i is proportional to $(1-\alpha^i)$. This is negative if $\alpha^i > 1$ and therefore the principal will prefer to agent to play action i when $\alpha^i > 1$. However if $\alpha^i < 1$, then the principal will prefer that the agent play action $i+1$. Thus in this modified rewriting lemma, contracts with breakpoints greater than 1, will prefer the lower action and breakpoints lower than 1, will prefer the higher action. By modifying Lemma B.3, we can rewrite any linear contract π using the rewriting rules of Theorem 3.1 into a new linear contract π' with a breakpoints that are at most 1, without any loss in utility. □

J Conclusion

In this paper, we provide a clean and tractable answer to our main question. When the agent's choice among n actions can lead to the success or failure of a project, the principal's optimal dynamic contract is surprisingly simple. Specifically, the principal should offer a carefully designed contract for a certain fraction of the T rounds (both the contract and the fraction are poly-time computable), then switch to a zero contract (i.e., pay the agent nothing) for the remaining rounds. Our main result also generalizes to settings with a rich set of outcomes beyond success/failure, as long as the principal changes the contract dynamically by scaling it ("single-dimensional scaling"). However, we show that without this single-dimensional scaling restriction, there exist principal-agent instances where the optimal dynamic contract does not take this form. In these cases, with non-linear contracts, the principal can do strictly better than offering the same contract for several rounds before switching to a zero contract.

As our second main result, we address a significant gap in the current literature on optimizing against no-regret learners: the assumption that the optimizer knows the time horizon T . We show that when there is uncertainty about T , even if limited, the principal's ability to use dynamic contracts to guarantee more revenue than the optimal static contract diminishes. We characterize the optimal dynamic contract under uncertainty of T , demonstrating that the principal's added value from being dynamic sharply degrades with an appropriate measure of uncertainty.

Open Problems. The computational study of repeated contracts, particularly with learning agents, raises many open questions. These include determining the optimal dynamic contract when the principal is not restricted to one-dimensional dynamics, and the computational complexity of finding it. Additionally, it involves identifying the optimal dynamic contract against a learning agent with a hidden type, thereby unifying our contract model with the auction model of [14]. Another intriguing area is understanding what the optimal dynamic contract would be against a team of multiple learning agents. Finally, it is crucial to explore the effects on welfare and utilities when there are two learning players, rather than a learner and an optimizer.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims in the abstract and introduction are an accurate description of the proofs in the rest of the paper. The paper is theoretical and the proofs are self-contained.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Our paper focuses on the case of linear contracts; the general contracts setting is also discussed and much more difficult (e.g. in some cases it is not known whether these objects are even computable).

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Our paper fully provides the claimed proofs.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: We are a purely theoretical paper and do not include any experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: This paper does not include any experiments requiring code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: We are a purely theoretical paper and do not include any experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: We are a purely theoretical paper and do not include any experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The paper does not use humans or data. The authors do not believe the listed potential harmful societal impacts apply.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: The work is purely theoretical and works in an abstract setting and has no immediate or forseeable impact to society.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper is purely abstract and theoretical and has no foreseeable risk for misuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use existing assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The project neither involves crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The project neither involves crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.