OPTIMAL CONTROL OF A CONTINUUM OF AGENTS IN MDPS: THE POWER OF FINITELY MANY CONTRACTS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study how a principal can control the state distribution of a continuum of agents through dynamic contracts, implementing a prescribed trajectory at minimum expected payment. In a finite-horizon Markov model with n states, agents of finitely many observable types privately choose actions from compact sets, with known, future payoff. The principal commits to randomized offers of bounded, nonnegative payments contingent on the next state. Agents are repooled by type and current state after each transition, and ties between best responses are resolved in the principal's favor.

We show that finitely many contracts suffice to attain the minimum expected payment for every implementable trajectory, yielding an exact finite-dimensional formulation of the infinite-dimensional design problem. Combining backward incentive conditions with forward population flows, we obtain an optimal implementation using at most $n + 1$ payment-action pairs per type, current state, and decision date. Joint reoptimization across current states reduces total support per type to at most $2n - 1$ pairs at the first date and $3n - 2$ at every later date, while retaining the local bound and minimum cost. This improves the support bound across current states from quadratic to linear in n by preserving aggregate population flows and relative continuation values. Both per-date joint support bounds are tight, already with two decision dates and continuous scalar actions.

1 INTRODUCTION

When a platform delegates sequential decisions to agents with their own objectives, controlling outcomes requires designing incentives for their behavior. Principal-agent Markov decision processes (MDPs) formalize this interaction between a designer who chooses payments and an agent who responds to them (Ben-Porat et al., 2024; Wu et al., 2024; Ivanov et al., 2024). Across a large population, a natural design objective is the evolution of the state distribution: how many agents occupy each state at each date. This objective makes the allocation of incentives across agents central to the control problem.

As a stylized example, consider a platform commissioning many independent service providers, each responsible for a continuing processing or verification task. At successive checkpoints, the task's status is observable, while the provider's computational or verification effort is private. A prescribed trajectory specifies the fraction of tasks in each status at every checkpoint. Providers facing the same status may receive different offers of payment for the next observed outcome, allowing the platform to combine the transitions induced by those offers. The design problem is to determine whether the prescribed trajectory can be implemented, its minimum expected payment, and how much randomization an optimal implementation requires.

We study this problem in a finite-horizon model with an atomless population, m observable types, and n states. Agents privately choose actions from compact sets with known, continuous payoff and transition functions. The principal commits to distributions of bounded, nonnegative payments contingent on the next state. After each transition, agents are repooled by type and current state: future offer distributions depend only on type, state, and date. Agents maximize discounted payoffs, participation is exogenous, and ties between best responses are resolved in the principal's favor. Individual primitives do not depend on population masses; the prescribed trajectory couples the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

groups through aggregate flow requirements. All implementation and optimality claims refer to this repooled contract class.

The design problem is infinite-dimensional: the principal chooses probability distributions over a continuum of payment vectors, coupled with agents' responses. Reducing this search to finitely many variables must preserve dynamic incentives, even though best responses can change discontinuously with the offer. In particular, preserving the population transition induced by a future contract need not preserve the agent's payoff from that contract. A replacement that reproduces later state masses can therefore change earlier actions. A useful finite representation must account for both population flows and the continuation-value information on which earlier incentives depend.

Our approach retains the transition and payoff quantities needed for population flows and dynamic incentives. Local convexification gives a finite representation of these quantities at minimum payment. A population target also leaves more freedom than a collection of separately specified transition policies: for a fixed type and date, it fixes the aggregate flow into each destination, while the conditional transitions from individual origin states remain choices. Earlier incentives depend on differences between continuation values, since adding the same constant to all next-state values shifts every action's payoff equally. Preserving aggregate flows and relative continuation values therefore allows us to reduce support jointly across current states. The resulting representation uses a number of payment-action pairs linear in the number of states, with matching lower bounds.

1.1 Our contributions

Our results characterize the minimum cost of implementing a prescribed population trajectory and the randomization required to attain it. We first obtain an exact finite-dimensional representation and then establish tight support bounds across current states.

Exact finite-dimensional reduction. We reduce the search over contract and response distributions to an optimization over finitely many payment vectors, actions, and mixing weights. The reduction preserves implementability and minimum expected payment, and every feasible target admits an optimizer that specifies an admissible contract. At most $n + 1$ payment-action pairs are needed per type, current state, and decision date (Theorem 3.1). This local support bound is independent of the action dimension and horizon; a refinement depends on the affine dimension of the attainable transition-payoff set. The reduction combines local convexification with backward consistency of continuation values and forward consistency of population flows.

Tight linear support bounds. Choosing assignment probabilities jointly across current states reduces the support per type and date from the separate bound of $n(n + 1)$ pairs to at most $2n - 1$ at the first date and $3n - 2$ at every later date, while retaining the local bound and the globally minimum payment (Theorem 3.2). For T decision dates, total support is therefore $O(mTn)$. These bounds sum support sizes across groups and also bound the sum of payment-support sizes, since several pairs may share a payment vector. Both per-date joint support bounds are tight (Proposition 3.3): for every $n \geq 2$, a single instance with one type, two decision dates, and continuous scalar actions attains both bounds. Every implementation requires at least $2n - 1$ joint pairs at the first date, and every minimum-cost implementation requires at least $3n - 2$ at the second date. At the second date, every group has positive mass, and the lower bound also holds for the sum of payment-support sizes. Thus the need for randomization persists even with one-dimensional actions and fully occupied state groups.

The resulting formulation characterizes feasibility and minimum cost for the original compact action spaces. It uses finitely many decision variables, with incentive constraints that require optimality against all actions in those spaces.

1.2 Related work

Our work connects dynamic contract design with population-level control objectives and finite-support representations. The discussion below relates our incentive structure to principal-agent models and our support bounds to convex methods for randomized policies.

Dynamic incentives and principal-agent MDPs. Recent work studies budget-constrained state-action reward shaping (Ben-Porat et al., 2024), planning and learning with next-state-contingent

payments (Wu et al., 2024), and subgame-perfect contract design with reinforcement learning methods (Ivanov et al., 2024). History dependence can improve contract performance: Bollini et al. (2024) use promised utilities to represent history-dependent commitments, while Gan et al. (2025) use inducible value sets under hindsight observability, including disclosure of past actions. Our model keeps actions private and treats repooling as a restriction on the contract class. Within this class, we study exact implementation of prescribed population state trajectories and the support required to minimize payment. The use of continuation values builds on dynamic moral hazard (Spear & Srivastava, 1987; Phelan & Townsend, 1991; Sannikov, 2008), including contracts for controlled Markov processes (Plambeck & Zenios, 2000) and hidden maintenance effort (Tian et al., 2021).

Randomized contracts and population objectives. Randomization over contracts has a long history (Fellingham et al., 1984). Castiglioni et al. (2023a) study randomized menus for privately informed agents, and Castiglioni et al. (2023b) study contracting with multiple agents whose individual outcomes jointly determine the principal's reward. Carmona & Wang (2021) study finite-state contracts with mean-field interactions in continuous time, allowing agents' transition rates and costs to depend on the population distribution and designing contracts to induce a mean-field equilibrium. In our model, individual payoff and transition primitives are independent of population masses. Types are observable, and random offers allocate incentives within each type-state group. The prescribed population trajectory links these allocations across groups, while continuation values link them across dates. Our support bounds quantify how many payment-action pairs suffice for exact implementation at minimum cost.

Constrained MDPs and finite support. Occupation measures and linear programming are standard tools for constrained MDPs (Altman, 1999), with randomization bounds determined by constraint structure (Ross, 1989; Feinberg & Shwartz, 1996). Extreme-point results for moment sets and randomized strategies provide related finite-support representations (Winkler, 1988; Pinelis, 2016; González-Hernández & Hernández-Lerma, 2005). Our analysis uses this convex geometry with moments that also preserve dynamic incentives. The local reduction fixes a group's transition and payoff coordinates; the joint reduction fixes aggregate flows and, after the first date, relative continuation values. The choice of these moments yields the linear support bounds, and the lower-bound construction establishes their tightness within the contract model.

2 Model

In this section, we present a dynamic principal-agent model in which privately chosen actions determine transitions between publicly observed states. The principal influences these actions through payments contingent on the next state. Starting from a given initial population distribution, the principal seeks to implement a prescribed population trajectory at minimum expected cost. We first describe the controlled Markov dynamics and the contract structure. We then formulate agents' incentives, the induced population dynamics, and the principal's design problem.

Notation. For any positive integer j, write $[j] := \{1, \dots, j\}$. For a metric space X, let $\Delta(X)$ denote the set of Borel probability measures on X. For a finite set Ω, we identify $\Delta(\Omega)$ with the probability vectors on Ω, and let $\mathbb{R}^{\Omega}$ denote the real vectors indexed by Ω.

2.1 Population and controlled Markov dynamics

The population is atomless and has unit mass. Let m be a positive integer denoting the number of observable types. For each type $k \in [m]$, let $\mu_k > 0$ denote its population mass, with $\sum_{k\in[m]} \mu_k = 1$. An agent's type remains fixed throughout the model.

Let Ω be a finite nonempty set of individual states, and let T be a positive integer denoting the number of decision periods. Decisions occur at dates $t \in [T]$, and date $T+1$ is terminal. For an individual agent, let $s(t) \in \Omega$ denote her state at date $t \in [T+1]$. For a decision date $t \in [T]$, her state after the next transition is $s(t+1)$. Whenever the date is clear from context, we write s and s' for the current state $s(t)$ and the next state $s(t+1)$, respectively. In population arrays, sums, and state-indexed vectors, $s(t)$ ranges over all state values at date t. Type and current state are publicly observed. Actions are private, and the realized next state is verifiable.

For each type $k \in [m]$, date $t \in [T+1]$, and state $s(t) \in \Omega$, let $\mu_{k,s(t)}(t)$ denote the mass of type-k agents in state $s(t)$ at date t. The masses $\mu_{k,s(1)}(1)$ are given and specify the initial population distribution. They satisfy

$$\mu_{k,s(1)}(1) \geq 0, \qquad \sum_{s(1)\in\Omega} \mu_{k,s(1)}(1) = \mu_k, \qquad k \in [m]. \tag{2.1}$$

Subsequent population distributions depend on the contract and agents' responses.

Actions and transitions. Fix a positive integer d, common to all types, states, and dates. For each type $k \in [m]$, date $t \in [T]$, and current state $s(t) \in \Omega$, let $A_{k,s(t)} \subseteq \mathbb{R}^d$ be the nonempty compact set of feasible actions. The feasible set depends only on the type and current state. Subscripts identify the type and state; for time-dependent quantities, the date appears in parentheses.

A type-k agent in state $s(t)$ at date t chooses an action vector

$$a_{k,s(t)}(t) = \big(a_{k,s(t)}(t,1), \ldots, a_{k,s(t)}(t,d)\big) \in A_{k,s(t)}.$$

Its ℓth coordinate is denoted by $a_{k,s(t)}(t,\ell)$ for $\ell \in [d]$. Thus, the pair of arguments records the date first and the action coordinate second. The action is scalar when $d = 1$.

The action determines both the agent's current payoff and her next-state distribution. For each type $k \in [m]$ and current state $s(t) \in \Omega$, define the current payoff function

$$b_{k,s(t)} : A_{k,s(t)} \longrightarrow \mathbb{R}.$$

It maps each feasible action $a_{k,s(t)}(t)$ to the payoff $b_{k,s(t)}(a_{k,s(t)}(t))$, including the action's benefits and costs but excluding contract payments.

Let $P_{k,s(t)}(\cdot) : A_{k,s(t)} \to \Delta(\Omega)$ be the controlled transition map. For each action $a_{k,s(t)}(t) \in A_{k,s(t)}$, the coordinate $P_{k,s(t)}(s(t+1) \mid a_{k,s(t)}(t))$ is the probability of reaching state $s(t+1)$ at the next date. Thus,

$$P_{k,s(t)}(a_{k,s(t)}(t)) = \big(P_{k,s(t)}(s(t+1) \mid a_{k,s(t)}(t))\big)_{s(t+1)\in\Omega},$$
$$P_{k,s(t)}(s(t+1) \mid a_{k,s(t)}(t)) \geq 0, \qquad \sum_{s(t+1)\in\Omega} P_{k,s(t)}(s(t+1) \mid a_{k,s(t)}(t)) = 1.$$

For each type and state, both $b_{k,s(t)}(\cdot)$ and $P_{k,s(t)}(\cdot)$ are continuous in the action. These primitives are common knowledge and depend on type, individual state, and action. The map $P_{k,s(t)}(\cdot)$ specifies the transition probabilities associated with each feasible action; the probabilities induced for agents of that type in that state at that date depend on the actions they take.

2.2 Contracts and timing

The principal offers agents a contract specifying payments contingent on their next observed states. We define such a contract as follows.

Definition 2.1. *Fix a finite payment cap $\bar{f} > 0$ and let $F := [0, \bar{f}]^{\Omega}$. A payment offered at date $t \in [T]$ is a function, identified with its payment vector,*

$$f(t+1,\cdot) : \Omega \longrightarrow [0, \bar{f}], \qquad f(t+1) \;\; := \big(f(t+1, s(t+1))\big)_{s(t+1)\in\Omega} \in F.$$

Its realized payment is $f(t+1, s(t+1))$ at date $t+1$.

A state-contingent contract specifies distributions

$$\Phi_{k,s(t)}(t) \in \Delta(F), \qquad k \in [m], \quad s(t) \in \Omega, \quad t \in [T].$$

The contract specifies one distribution of payment vectors $f(t+1)$ for each type k, date t, and possible current state $s(t)$, including groups with zero mass.

Under exact aggregation, the probability assigned to a measurable set of payment vectors equals the fraction of a positive-mass group receiving a vector in that set. Thus, $f(t+1)$ may differ across agents in the same group.

Timeline. The principal commits to the contract before the first decision date. Agents' participation is exogenous. At each decision date t, an agent first receives and observes her payment vector $f(t+1)$. She then chooses her action $a_{k,s(t)}(t)$ privately. Her next state $s(t+1)$ is drawn according to the controlled transition map, and the transfer $f(t+1, s(t+1))$ is paid. Payment dates belong to $\{2, \dots, T+1\}$, and transfers are nonnegative. Agents are subsequently repooled by type and new state. Conditional on type, current state, and date, the next offer distribution is independent of earlier states, offers, and actions.

2.3 Agent utility and optimal behavior

Agent utility. Agents are risk neutral and have quasilinear utility. The function $b_{k,s(t)}(\cdot)$ introduced in Section 2.1 gives a type-k agent's current payoff excluding transfers. Given action $a_{k,s(t)}(t)$ and realized next state $s(t+1)$, her date-t payoff is

$$b_{k,s(t)}(a_{k,s(t)}(t)) + f(t+1, s(t+1)).$$

The transfer is paid at date $t+1$ and counted as part of the date-t payoff. For each type $k \in [m]$, let $\lambda_k \in (0,1]$ be the discount factor. Apart from the payment associated with the final transition, agents receive no additional payoff at date $T+1$.

Histories and general strategies. For a fixed type k, the agent's history at date t, after observing the current offer and before choosing her action, is

$$h(t) := \Big(\big(s(\tau)\big)_{\tau=1}^{t},\ \big(f(\tau+1)\big)_{\tau=1}^{t},\ \big(a_{k,s(\tau)}(\tau)\big)_{\tau=1}^{t-1} \Big).$$

Thus, the history records all observed states and payment vectors, and all past actions; the action record is empty at $t=1$. Let $\mathcal{H}_k(t)$ be the space of these records, with states in Ω, payment vectors in F, and each past action in its corresponding feasible set. A history-dependent strategy σ specifies Borel action kernels

$$\sigma_{k,s(t)}(t, \cdot \mid h(t)) \in \Delta(A_{k,s(t)}), \qquad k \in [m], \quad t \in [T], \quad h(t) \in \mathcal{H}_k(t).$$

A strategy is Markov if its action distribution depends on the history only through the current state and offer, given type and date. Fix a contract. The expected discounted payoff from history $h(t)$ under strategy σ is

$$J_k^{\sigma}(t, h(t)) := \mathbb{E}\left[\sum_{\tau=t}^{T} \lambda_k^{\tau-t} \big(b_{k,s(\tau)}(a_{k,s(\tau)}(\tau)) + f(\tau+1, s(\tau+1)) \big) \right].$$

The expectation starts from the fixed history $h(t)$, including the current offer. At each date $\tau = t, \dots, T$, the action is drawn from $\sigma_{k,s(\tau)}(\tau, \cdot \mid h(\tau))$ and the next state from $P_{k,s(\tau)}(\cdot \mid a_{k,s(\tau)}(\tau))$. At later dates $\tau > t$, a new offer is first drawn from $\Phi_{k,s(\tau)}(\tau)$.

The following lemma shows that using earlier history cannot improve on an optimal Markov strategy.

Lemma 2.1. *Fix a contract. There exists a deterministic Borel Markov strategy $\sigma^\star$ such that, for every type $k \in [m]$, date $t \in [T]$, history $h(t) \in \mathcal{H}_k(t)$, and Borel history-dependent strategy σ,*

$$J_k^{\sigma}(t, h(t)) \le J_k^{\sigma^\star}(t, h(t)) = U^\star_{k,s(t)}(t, f(t+1)).$$

The optimal expected payoff $U^\star_{k,s(t)}(t, f(t+1))$ therefore depends only on type, current state, date, and current offer.

The proof is given in Appendix A.

Let $\overline{U}^\star_{k,s(t)}(t)$ denote the expectation of $U^\star_{k,s(t)}(t, f(t+1))$ over the current offer drawn from $\Phi_{k,s(t)}(t)$, and set $\overline{U}^\star_{k,s(T+1)}(T+1) = 0$.

Best-response distributions. For a fixed contract, let $\mathrm{BR}_{k,s(t)}(t, f(t+1))$ be the set of actions maximizing current expected payoff plus discounted optimal future payoff:

$$\mathrm{BR}_{k,s(t)}(t, f(t+1)) := \underset{a_{k,s(t)}(t)\in A_{k,s(t)}}{\arg\max} \Bigg\{ b_{k,s(t)}(a_{k,s(t)}(t)) + \sum_{s(t+1)\in\Omega} P_{k,s(t)}(s(t+1) \mid a_{k,s(t)}(t)) \Big[f(t+1, s(t+1)) + \lambda_k \overline{U}^{\star}_{k,s(t+1)}(t+1) \Big] \Bigg\}.$$

A Markov best-response rule specifies Borel action kernels

$$\Lambda_{k,s(t)}(t, \cdot \mid f(t+1)) \in \Delta(\mathrm{BR}_{k,s(t)}(t, f(t+1))),$$

for all $k \in [m]$, $s(t) \in \Omega$, $t \in [T]$, and observed payment vectors $f(t+1) \in F$. Thus, each action distribution assigns probability one to best responses and may mix among tied maximizing actions.

Given the contract and these response distributions, define the agent's expected discounted payoff from date t onward, with current state $s(t)$ and observed payment vector $f(t+1)$ fixed:

$$U_{k,s(t)}(t, f(t+1)) := \mathbb{E}\left[\sum_{\tau=t}^{T} \lambda_k^{\tau-t} \big(b_{k,s(\tau)}(a_{k,s(\tau)}(\tau)) + f(\tau+1, s(\tau+1)) \big) \right].$$

The expectation starts from the given state and offer. At each date $\tau = t, \ldots, T$, the action is drawn from $\Lambda_{k,s(\tau)}(\tau, \cdot \mid f(\tau+1))$ and the next state from $P_{k,s(\tau)}(\cdot \mid a_{k,s(\tau)}(\tau))$. At later dates $\tau > t$, a new offer is first drawn from $\Phi_{k,s(\tau)}(\tau)$.

By Lemma 2.1, these best-response distributions attain the optimal payoff:

$$U_{k,s(t)}(t, f(t+1)) = U^{\star}_{k,s(t)}(t, f(t+1)).$$

Assumption 1. *For a fixed contract, the principal may select the distribution over tied best-response actions at each type, state, date, and observed payment vector. Agents follow the selected distribution.*

The contract distributions together with their associated best-response distributions constitute an admissible outcome. The principal's design problem considers only such outcomes.

2.4 Population trajectories and the principal's objective

Fix the contract distributions $\Phi_{k,s(t)}(t)$ and best-response distributions $\Lambda_{k,s(t)}(t, \cdot \mid f(t+1))$ for all types, states, dates, and current offers. Let $\mu_{k,s(t)}(t)$ denote the induced population masses, collected in the trajectory μ. Starting from the initial masses in (2.1), exact aggregation gives the following population dynamics for every type $k \in [m]$, destination state $s(t+1) \in \Omega$, and decision date $t \in [T]$:

$$\mu_{k,s(t+1)}(t+1) = \sum_{s(t)\in\Omega} \mu_{k,s(t)}(t)\, \mathbb{E}\big[P_{k,s(t)}(s(t+1) \mid a_{k,s(t)}(t)) \big]. \tag{2.2}$$

For each fixed type and date, the expectation at each origin state averages over an offer drawn from $\Phi_{k,s(t)}(t)$ and then an action drawn from $\Lambda_{k,s(t)}(t, \cdot \mid f(t+1))$. The destination state $s(t+1)$ is fixed.

A population trajectory is an array $\mu = (\mu_{k,s(t)}(t))_{k\in[m],\, s(t)\in\Omega,\, t\in[T+1]}$ of state masses. Let $\mathcal{T}$ be the set of nonnegative such arrays that share the given date-one masses specified in Section 2.1 and preserve each type's total mass:

$$\sum_{s(t)\in\Omega} \mu_{k,s(t)}(t) = \mu_k, \qquad k \in [m],\ t \in [T+1].$$

Membership in this set imposes only the initial condition and population accounting. Incentives and the controlled transition maps determine which of these trajectories can be induced by admissible outcomes.

Fix a prescribed target trajectory $\bar{\mu} \in \mathcal{T}$, with masses $\bar{\mu}_{k,s(t)}(t)$. An admissible outcome implements $\bar{\mu}$ if its induced trajectory satisfies $\mu = \bar{\mu}$. The target is *implementable* if such an outcome exists. The principal seeks to implement this trajectory at minimum total expected payment.

Principal payment. The principal minimizes total expected payments. For the fixed distributions above, let W denote the total expected payment:

$$W := \sum_{t\in[T]} \sum_{k\in[m]} \sum_{s(t)\in\Omega} \mu_{k,s(t)}(t)\, \mathbb{E}[f(t+1, s(t+1))]\,. \tag{2.3}$$

For each type, origin state, and date in the sum, the expectation averages over the offer, action, and next state drawn successively from $\Phi_{k,s(t)}(t)$, $\Lambda_{k,s(t)}(t, \cdot \mid f(t+1))$, and $P_{k,s(t)}(\cdot \mid a_{k,s(t)}(t))$, respectively.

Let $\mathrm{OPT}(\bar{\mu})$ denote the infimum of this cost over admissible outcomes implementing $\bar{\mu}$. The principal's problem is

$$\mathrm{OPT}(\bar{\mu}) := \inf \left\{ W : \begin{array}{l} \Phi_{k,s(t)}(t) \in \Delta(F), \\ \Lambda_{k,s(t)}(t, \cdot \mid f(t+1)) \in \Delta(\mathrm{BR}_{k,s(t)}(t, f(t+1))), \\ \mu = \bar{\mu} \end{array} \right\}.$$

The infimum is taken jointly over the indexed contract distributions and Borel response kernels for all $k \in [m]$, $s(t) \in \Omega$, and $t \in [T]$, with the response condition holding for every $f(t+1) \in F$. The best-response sets are computed under the chosen contract distributions; the trajectory μ and payment W are determined by (2.2) and (2.3). We set $\mathrm{OPT}(\bar{\mu}) = +\infty$ if $\bar{\mu}$ is not implementable.

3 Implementable population trajectories and optimal contracts

This section characterizes which population trajectories can be induced by state-contingent contracts and the minimum total expected payment required to induce them. For a prescribed trajectory, we combine the population dynamics with agents' best-response conditions to obtain a finite-dimensional formulation and support bounds both within each group and jointly across current states.

We maintain the model assumptions of Section 2, including principal-favorable selection (Assumption 1). The principal chooses the contract distributions $\Phi_{k,s(t)}(t)$ and selects action distributions $\Lambda_{k,s(t)}(t, \cdot \mid f(t+1))$ supported on the corresponding best-response sets. These choices determine the population trajectory μ and total expected payment W through (2.2) and (2.3), respectively.

Let $n := |\Omega|$ and fix a target trajectory $\bar{\mu} \in \mathcal{T}$. We use the notion of implementability and the implementation cost $\mathrm{OPT}(\bar{\mu})$ defined in Section 2.4.

For each type, current state, and decision date, let $\Gamma_{k,s(t)}(t) \in \Delta(F \times A_{k,s(t)})$ be the joint distribution of the offered payment vector and the chosen action. It is induced by the contract and action distributions through

$$\Gamma_{k,s(t)}(t, \mathrm{d}f(t+1), \mathrm{d}a_{k,s(t)}(t)) := \Phi_{k,s(t)}(t, \mathrm{d}f(t+1))\, \Lambda_{k,s(t)}(t, \mathrm{d}a_{k,s(t)}(t) \mid f(t+1)).$$

Write supp for the support of a distribution and use vertical bars for its cardinality. A finitely supported joint distribution is specified by its payment-action pairs and their probabilities.

Theorem 3.1. *A target trajectory $\bar{\mu} \in \mathcal{T}$ is implementable if and only if it admits an admissible implementation using at most $n+1$ payment-action pairs for each type, current state, and decision date. Whenever $\bar{\mu}$ is implementable, its least total expected payment is attained by an implementation of this form. Such an optimal implementation satisfies $W = \mathrm{OPT}(\bar{\mu})$ and*

$$\left|\mathrm{supp}\, \Phi_{k,s(t)}(t)\right| \le \left|\mathrm{supp}\, \Gamma_{k,s(t)}(t)\right| \le n+1, \qquad k \in [m], \quad s(t) \in \Omega, \quad t \in [T].$$

The proof is given in Appendix B.

Implications of Theorem 3.1. The theorem turns the search over distributions into a finite-dimensional design problem. For each type, state, and date, one can allocate $n+1$ candidate pairs and optimize their payment vectors, actions, and probabilities, allowing zero weights. Subject

to the population-flow and best-response conditions, this restriction preserves both feasibility and minimum payment for the prescribed trajectory. Attainment ensures that an optimizer specifies an admissible optimal contract.

The finite list also describes how to implement the contract: take its payment marginal as the offer distribution and, conditional on each assigned offer, mix among the actions paired with it. Several pairs may share one payment vector, so the number of distinct offers can be smaller than the joint support. The bound per group is independent of the action dimension d and horizon T, although the pairs themselves contain d-dimensional actions. The incentive constraints still compare each selected action with all feasible deviations in the original action space.

Applying Theorem 3.1 separately at each current state gives at most $n(n+1)$ payment-action pairs per type and date. The target trajectory, however, fixes only the aggregate population flow into each next state. The individual conditional transition vectors remain choice variables. Reoptimizing the assignment probabilities jointly across current states gives a smaller total support. Earlier incentives can also be preserved while allowing a common additive shift in all continuation values of a type.

Theorem 3.2. *Every implementable target $\bar{\mu} \in \mathcal{T}$ admits an optimal implementation with $W = \mathrm{OPT}(\bar{\mu})$ such that*

$$\sum_{s(t)\in\Omega} \left|\operatorname{supp} \Gamma_{k,s(t)}(t)\right| \le \begin{cases} 2n-1, & t=1, \\ 3n-2, & t \in \{2,\dots,T\}, \end{cases} \qquad k \in [m]. \tag{3.1}$$

The same implementation satisfies $\left|\operatorname{supp} \Phi_{k,s(t)}(t)\right| \le \left|\operatorname{supp} \Gamma_{k,s(t)}(t)\right| \le n+1$ for every type, current state, and decision date. Each group with $\bar{\mu}_{k,s(t)}(t) = 0$ uses a single payment-action pair.

The proof is given in Appendix C.

Implications of Theorem 3.2. Equation (3.1) reduces the support bound per type and date from quadratic to linear in n, while retaining every local bound. In the resulting optimal implementation, the unweighted average support size across all current states is below two at the first date and below three at every later date. Individual states may still use larger supports: beyond one pair per state, there are at most $n-1$ additional pairs at date 1 and $2n-2$ at any later date. These counts sum support sizes separately across states, including zero-mass states. The offers and weights may differ across groups, and the payment marginals inherit the same upper bounds.

The distinction between the first and later dates reflects dynamic incentives. At later dates, preserving continuation-value differences keeps earlier actions optimal: a common shift in all states' values leaves those actions' rankings unchanged. Alongside n normalization equations, the joint reoptimization therefore preserves $n-1$ aggregate-flow coordinates and, at later dates, $n-1$ value differences. At the first date there are no earlier incentives to preserve. This explains the two constants and why the bound at a given date does not grow with the horizon.

The next proposition shows that neither $2n-1$ nor $3n-2$ in Theorem 3.2 can be decreased uniformly over the model class. At the second date, the lower bound also applies to the sum of payment-support sizes across groups, and every group has positive population mass.

Proposition 3.3. *For every $n \ge 2$, there is an instance with one type, $T = 2$, action sets $A_{1,s} = [0,1]$, and continuous, time-independent payoff and transition functions, together with an implementable target $\bar{\mu}$, such that*

$$\sum_{s\in\Omega} |\operatorname{supp} \Gamma_{1,s}(1)| \ge 2n-1 \qquad \textit{for every admissible implementation}, \tag{3.2}$$

$$\sum_{s\in\Omega} |\operatorname{supp} \Gamma_{1,s}(2)| \ge 3n-2 \qquad \textit{for every optimal implementation}. \tag{3.3}$$

One optimal implementation attains both bounds simultaneously. Moreover, $\bar{\mu}_{1,s}(2) > 0$ for every state, and every optimal implementation satisfies

$$|\operatorname{supp} \Phi_{1,s}(2)| \ge 3 \quad \textit{for each } s \in [n-1], \qquad \sum_{s\in\Omega} |\operatorname{supp} \Phi_{1,s}(2)| \ge 3n-2.$$

The same implementation attains this payment-support bound.

The proof is given in Appendix D.

Implications of Proposition 3.3. The proposition makes both per-date constants in Theorem 3.2 exact. The first-date lower bound is a requirement for implementing the trajectory; the second-date lower bound is a requirement for minimizing payment. Both are attained in one instance, so neither constant can be decreased uniformly even with one type, two dates, continuous primitives, and a scalar action.

The example also separates randomization over offers from mixing among tied actions. At the first date, the attaining contract uses a single offer in each group; its joint support count includes one pair for each zero-mass group. At the second date, every state has positive mass and every optimum requires at least three distinct payment vectors in each of $n-1$ states. Restricting these groups to at most two offers would exclude every minimum-cost implementation, even with principal-favorable tie-breaking. Thus the finite representation must retain payment randomization to attain the optimum in this example.

4 Conclusion

We showed how to replace the infinite-dimensional search for a dynamic contract with a finite-dimensional formulation of minimum-cost population implementation. The reduction preserves both the set of implementable trajectories and their minimum expected payment. For every feasible target, a minimizer exists and specifies a contract through finitely many payment-action pairs and their probabilities. Backward incentive conditions and forward population flows ensure that this representation is consistent across dates.

The support bounds control the size of this representation. With n states, at most $n+1$ payment-action pairs are needed within each type-state group and date. Coordinating assignments across current states reduces the sum of support sizes per type to at most $2n-1$ at the first date and $3n-2$ at every later date, while preserving the local bound and minimum cost. The key is to preserve aggregate population flows and relative continuation values: a common shift in continuation values leaves earlier incentives unchanged. Both per-date joint bounds are tight, already with two decision dates and continuous scalar actions.

These results concern an atomless population with known primitives and observable types, under repooling and principal-favorable tie-breaking. Developing efficient algorithms for the finite-dimensional formulation remains an important direction, since its incentive constraints still encode best responses over continuous action spaces. Another direction is to study approximate implementation when payoff and transition functions are estimated from data or agents' choices among tied best responses cannot be controlled. Such extensions would clarify how the support guarantees interact with uncertainty in population outcomes and incentives.

AI use statement

We used OpenAI Codex to assist with drafting and revising the manuscript, identifying and summarizing related literature, refining mathematical statements, checking mathematical arguments, and drafting and revising proofs. We also used it to edit LaTeX source files and format references. The authors take full responsibility for the correctness and originality of the mathematical results and proofs, the accuracy of the references, and the final content of the paper.

References

Eitan Altman. *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999. ISBN 9780849303821. URL https://www.routledge.com/Constrained-Markov-Decision-Processes/Altman/p/book/9780849303821.

Omer Ben-Porat, Yishay Mansour, Michal Moshkovitz, and Boaz Taitler. Principal-agent reward shaping in MDPs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(9):9502–9510, 2024. doi: 10.1609/aaai.v38i9.28805.

Matteo Bollini, Francesco Bacchiocchi, Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. Contracting with a reinforcement learning agent by playing trick or treat. *arXiv preprint*

arXiv:2410.13520, 2024. doi: 10.48550/arXiv.2410.13520. URL https://arxiv.org/abs/2410.13520.

René Carmona and Peiqi Wang. Finite-state contract theory with a principal and a field of agents. *Management Science*, 67(8):4725–4741, 2021. doi: 10.1287/mnsc.2020.3760. URL https://doi.org/10.1287/mnsc.2020.3760.

Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. Designing menus of contracts efficiently: The power of randomization. *Artificial Intelligence*, 318:103881, 2023a. doi: 10.1016/j.artint.2023.103881. URL https://doi.org/10.1016/j.artint.2023.103881.

Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. Multi-agent contract design: How to commission multiple agents with individual outcomes. In *Proceedings of the 24th ACM Conference on Economics and Computation*, EC '23, pp. 412–448. Association for Computing Machinery, 2023b. doi: 10.1145/3580507.3597793. URL https://doi.org/10.1145/3580507.3597793.

Eugene A. Feinberg and Adam Shwartz. Constrained discounted dynamic programming. *Mathematics of Operations Research*, 21(4):922–945, 1996. doi: 10.1287/moor.21.4.922. URL https://doi.org/10.1287/moor.21.4.922.

John C. Fellingham, Young K. Kwon, and D. Paul Newman. Ex ante randomization in agency models. *The RAND Journal of Economics*, 15(2):290–301, 1984. doi: 10.2307/2555682. URL https://doi.org/10.2307/2555682.

Jiarui Gan, Rupak Majumdar, Debmalya Mandal, and Goran Radanovic. Stochastic principal-agent problems: Computing and learning optimal history-dependent policies. In *Advances in Neural Information Processing Systems*, volume 38, pp. 156800–156825. Curran Associates, Inc., 2025. doi: 10.52202/085713-5239. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/e5fefcfcb7288181562b67162c7d5e3a-Abstract-Conference.html.

Juan González-Hernández and Onésimo Hernández-Lerma. Extreme points of sets of randomized strategies in constrained optimization and control problems. *SIAM Journal on Optimization*, 15(4):1085–1104, 2005. doi: 10.1137/040605345. URL https://doi.org/10.1137/040605345.

Dima Ivanov, Paul Dütting, Inbal Talgam-Cohen, Tonghan Wang, and David C. Parkes. Principal-agent reinforcement learning: Orchestrating AI agents with contracts. *arXiv preprint arXiv:2407.18074*, 2024. doi: 10.48550/arXiv.2407.18074. URL https://arxiv.org/abs/2407.18074.

Christopher Phelan and Robert M. Townsend. Computing multi-period, information-constrained optima. *The Review of Economic Studies*, 58(5):853–881, 1991. doi: 10.2307/2297941. URL https://doi.org/10.2307/2297941.

Iosif Pinelis. On the extreme points of moments sets. *Mathematical Methods of Operations Research*, 83(3):325–349, 2016. doi: 10.1007/s00186-015-0530-0. URL https://doi.org/10.1007/s00186-015-0530-0.

Erica L. Plambeck and Stefanos A. Zenios. Performance-based incentives in a dynamic principal-agent model. *Manufacturing & Service Operations Management*, 2(3):240–263, 2000. doi: 10.1287/msom.2.3.240.12345. URL https://doi.org/10.1287/msom.2.3.240.12345.

Keith W. Ross. Randomized and past-dependent policies for Markov decision processes with multiple constraints. *Operations Research*, 37(3):474–477, 1989. doi: 10.1287/opre.37.3.474. URL https://doi.org/10.1287/opre.37.3.474.

Yuliy Sannikov. A continuous-time version of the principal-agent problem. *The Review of Economic Studies*, 75(3):957–984, 2008. doi: 10.1111/j.1467-937X.2008.00486.x. URL https://doi.org/10.1111/j.1467-937X.2008.00486.x.

Stephen E. Spear and Sanjay Srivastava. On repeated moral hazard with discounting. *The Review of Economic Studies*, 54(4):599–617, 1987. doi: 10.2307/2297484. URL https://doi.org/10.2307/2297484.

Feng Tian, Peng Sun, and Izak Duenyas. Optimal contract for machine repair and maintenance. *Operations Research*, 69(3):916–949, 2021. doi: 10.1287/opre.2020.2018. URL https://doi.org/10.1287/opre.2020.2018.

Gerhard Winkler. Extreme points of moment sets. *Mathematics of Operations Research*, 13(4): 581–587, 1988. doi: 10.1287/moor.13.4.581. URL https://doi.org/10.1287/moor.13.4.581.

Jibang Wu, Siyu Chen, Mengdi Wang, Huazheng Wang, and Haifeng Xu. Contractual reinforcement learning: Pulling arms with invisible hands. *arXiv preprint arXiv:2407.01458*, 2024. doi: 10.48550/arXiv.2407.01458. URL https://arxiv.org/abs/2407.01458.

A Optimality of Markov strategies

We prove that a deterministic Markov strategy attains the optimal payoff after every history under a fixed contract. The proof constructs a measurable maximizing action at each state and offer and then applies backward induction.

Proof of Lemma 2.1. Set $\overline{U}^{\star}_{k,s(T+1)}(T+1) = 0$. For $t = T, \dots, 1$, define recursively

$$U^{\star}_{k,s(t)}(t, f(t+1)) := \max_{a_{k,s(t)}(t) \in A_{k,s(t)}} \Bigg\{ b_{k,s(t)}(a_{k,s(t)}(t)) + \sum_{s(t+1)\in\Omega} P_{k,s(t)}(s(t+1) \mid a_{k,s(t)}(t)) \Big[f(t+1, s(t+1)) + \lambda_k \overline{U}^{\star}_{k,s(t+1)}(t+1) \Big] \Bigg\},$$

$$\overline{U}^{\star}_{k,s(t)}(t) := \int_F U^{\star}_{k,s(t)}(t, f(t+1))\, \Phi_{k,s(t)}(t, \mathrm{d}f(t+1)).$$

Here $\overline{U}^{\star}$ averages the value over the current offer. All values are bounded. The maximand is jointly continuous in the action and offer, and the action set is compact, so a maximizer exists and admits a Borel selection. Let $\sigma^{\star}$ assign probability one to such a maximizer at every type, state, date, and offer.

Fix any history-dependent strategy σ. We prove the payoff bound by backward induction. At $t = T$, the future payoff is zero. At an earlier date, the induction hypothesis bounds the payoff after every next history by $U^{\star}_{k,s(t+1)}(t+1, f(t+2))$. Conditional on the next state, the new offer has distribution $\Phi_{k,s(t+1)}(t+1)$ regardless of the earlier history, so its average is bounded by $\overline{U}^{\star}_{k,s(t+1)}(t+1)$. Hence, at every date,

$$\begin{aligned} J^{\sigma}_{k}(t, h(t)) &\le \int_{A_{k,s(t)}} \Bigg\{ b_{k,s(t)}(a_{k,s(t)}(t)) \\ &\quad + \sum_{s(t+1)\in\Omega} P_{k,s(t)}(s(t+1) \mid a_{k,s(t)}(t)) \Big[f(t+1, s(t+1)) + \lambda_k \overline{U}^{\star}_{k,s(t+1)}(t+1) \Big] \Bigg\} \\ &\quad \cdot \sigma_{k,s(t)}(t, \mathrm{d}a_{k,s(t)}(t) \mid h(t)) \\ &\le U^{\star}_{k,s(t)}(t, f(t+1)). \end{aligned}$$

For $\sigma^{\star}$, the selected action attains the maximum and the induction hypothesis holds with equality at every next history. Both inequalities are therefore equalities, which proves the claim. □

B LOCAL CONVEXIFICATION AND OPTIMAL IMPLEMENTATION

This appendix proves Theorem 3.1. Fix a target trajectory $\bar{\mu} \in \mathcal{T}$. We first express population flows and agents' expected payoffs using finitely many variables. We then show that each feasible transition-payoff pair has a minimum-payment representation using at most $n+1$ payment-action pairs. Combining these representations gives an optimal contract for the target trajectory.

Definition B.1. *For a fixed contract, a continuation value is an agent's optimal expected discounted payoff from date t onward, before observing the current offer. Let Ψ collect the candidate values $\Psi_{k,s(t)}(t)$, with state vectors*

$$\Psi_k(t) := (\Psi_{k,s(t)}(t))_{s(t)\in\Omega}.$$

For fixed k and t, $\psi \in \mathbb{R}^{\Omega}$ denotes a candidate next-date vector; when specified by Ψ, $\psi = \Psi_k(t+1)$. Terminal values are zero.

For each type $k \in [m]$, current state $s(t) \in \Omega$, and date $t \in [T]$, let

$$q_{k,s(t)}(t) := \big(q_{k,s(t)}(t, s(t+1))\big)_{s(t+1)\in\Omega} \in \Delta(\Omega)$$

be a candidate next-state probability vector. Each coordinate $q_{k,s(t)}(t, s(t+1)) \in [0,1]$ is the probability of reaching the indicated next state. Collect all mnT such vectors in the full transition array

$$\boldsymbol{q} := \big(q_{k,s(t)}(t)\big)_{k\in[m],\, s(t)\in\Omega,\, t\in[T]} \in \Delta(\Omega)^{mnT}.$$

Define $\mathcal{Q}(\bar{\mu}) \subseteq \Delta(\Omega)^{mnT}$ as the set of full arrays $\boldsymbol{q}$ satisfying the population flow equations

$$\bar{\mu}_{k,s(t+1)}(t+1) = \sum_{s(t)\in\Omega} \bar{\mu}_{k,s(t)}(t) q_{k,s(t)}(t, s(t+1)), \qquad k \in [m], \quad s(t+1) \in \Omega, \quad t \in [T]. \tag{B.1}$$

These equations require the candidate transitions to reproduce the target population trajectory. The set $\mathcal{Q}(\bar{\mu})$ is compact. It is nonempty: assigning $q_{k,s(t)}(t, s(t+1)) = \bar{\mu}_{k,s(t+1)}(t+1)/\mu_k$ at every origin satisfies the equations. Whether these transitions can be induced by contracts is the remaining incentive question.

To bound the continuation values, choose $\bar{b} \geq 0$ such that $|b_{k,s(t)}(a_{k,s(t)}(t))| \leq \bar{b}$ for every type, state, and feasible action, and set

$$\bar{\Psi} := T(\bar{b} + \bar{f}), \qquad K := [-\bar{\Psi}, \bar{\Psi}]^{\Omega}.$$

Every actual continuation-value vector lies in K: there are at most T remaining payoffs, each bounded in absolute value by $\bar{b} + \bar{f}$, and discount factors do not exceed one.

For a candidate continuation-value vector $\psi \in K$, define the optimized payoff after observing the current offer by

$$V_{k,s(t)}(t, f(t{+}1); \psi) := \max_{a_{k,s(t)}(t)\in A_{k,s(t)}} \Big\{ b_{k,s(t)}(a_{k,s(t)}(t)) + P_{k,s(t)}(a_{k,s(t)}(t)) \cdot \big(f(t{+}1) + \lambda_k \psi\big) \Big\}. \tag{B.2}$$

The dot product sums over the next states $s(t+1) \in \Omega$: the payment coordinate is $f(t+1, s(t+1))$, and the future-payoff coordinate is $\psi(s(t+1))$. Let $\mathrm{BR}_{k,s(t)}(t, f(t+1); \psi)$ denote the set of maximizing actions in (B.2). Compactness and continuity make this set nonempty, give it a closed graph, and provide a Borel selection; the optimized value is continuous in the offer and ψ.

This is the Bellman expression of Lemma 2.1, with a candidate next-date vector. In particular, if the selected actions maximize this expression and

$$\Psi_{k,s(T+1)}(T+1) = 0, \qquad \Psi_{k,s(t)}(t) = \int_F V_{k,s(t)}(t, f(t+1); \Psi_k(t+1))\, \Phi_{k,s(t)}(t, \mathrm{d}f(t+1)), \tag{B.3}$$

backward induction shows that $\Psi_{k,s(t)}(t)$ is the agent's continuation value under the contract and that the selected actions are best responses. Conversely, every admissible outcome satisfies these conditions. They apply for all types, states, and dates, including states with zero prescribed mass, and optimality is required after every feasible offer.

To determine which candidate transitions and expected payoffs can be implemented locally, fix a type k, current state $s(t)$, date t, and next-date vector $\psi \in K$. Let

$$Z_{k,s(t)}(t;\psi) := \big\{\big(f(t+1), a_{k,s(t)}(t)\big) \in F \times A_{k,s(t)} : a_{k,s(t)}(t) \in \mathrm{BR}_{k,s(t)}(t, f(t+1);\psi)\big\}.$$

Each such pair determines a next-state distribution, an expected payoff, and an expected payment. Collect these triples in

$$\begin{aligned} G_{k,s(t)}(t;\psi) := \Big\{\Big(&P_{k,s(t)}(a_{k,s(t)}(t)),\, V_{k,s(t)}(t, f(t+1);\psi), \\ &P_{k,s(t)}(a_{k,s(t)}(t)) \cdot f(t+1)\Big) : \big(f(t+1), a_{k,s(t)}(t)\big) \in Z_{k,s(t)}(t;\psi)\Big\}. \end{aligned} \tag{B.4}$$

Randomizing over these pairs averages the three coordinates. Write $\widehat{G}_{k,s(t)}(t;\psi)$ for the convex hull of $G_{k,s(t)}(t;\psi)$, consisting of all finite convex combinations of its elements. For the fixed type, current state, and date, let $q \in \Delta(\Omega)$ denote a single candidate next-state distribution. For an expected payoff $u \in \mathbb{R}$ and expected payment $p \in \mathbb{R}$, define

$$M_{k,s(t)}(t;\psi) := \big\{(q,u) : \text{ there exists } p \text{ with } (q,u,p) \in G_{k,s(t)}(t;\psi)\big\},$$

$$\kappa_{k,s(t)}(t,\psi;q,u) := \min\big\{p : (q,u,p) \in \widehat{G}_{k,s(t)}(t;\psi)\big\}.$$

The minimum is $+\infty$ if the constraint set is empty. Thus κ is the least payment per agent that implements the specified transition distribution and expected payoff.

For a nonempty set, affdim denotes the dimension of the smallest affine subspace containing that set. Because the transition coordinates sum to one,

$$\mathrm{affdim}\, M_{k,s(t)}(t;\psi) \le (n-1) + 1 = n. \tag{B.5}$$

Lemma B.1. *For every type, current state, and decision date, the following statements hold.*

(i) The triples induced by joint distributions of payment vectors and best-response actions are exactly $\widehat{G}_{k,s(t)}(t;\psi)$. The set of all (ψ,q,u,p) satisfying this membership with $\psi \in K$ is compact.

(ii) The function $\kappa_{k,s(t)}(t,\psi;q,u)$ is jointly lower semicontinuous in (ψ,q,u) and lies in $[0,\bar{f}]$ whenever finite. For fixed ψ, it is convex in (q,u).

(iii) Every feasible (q,u) admits a minimum-payment implementation supported on at most $\mathrm{affdim}\, M_{k,s(t)}(t;\psi) + 1 \le n+1$ payment-action pairs, with best responses specified for every feasible payment vector.

Proof. Fix a type k, current state $s(t)$, and date t. We first prove compactness while allowing ψ to vary over K. The set

$$\Big\{\big(\psi, f(t+1), a_{k,s(t)}(t)\big) : \psi \in K, \quad \big(f(t+1), a_{k,s(t)}(t)\big) \in Z_{k,s(t)}(t;\psi)\Big\}$$

is closed: by continuity, the best-response inequalities remain valid when the payment vector, action, and future-payoff vector converge. This set lies in the compact product $K \times F \times A_{k,s(t)}$, so it is compact. Applying the continuous triple map in (B.4), while retaining ψ, gives a compact set of quadruples (ψ,q,u,p).

For each fixed ψ, the transition vector has at most $n-1$ independent coordinates, while the payoff and payment each contribute one. Thus the triples lie in an affine space of dimension at most $n+1$. Carathéodory's theorem gives a representation with $r \le n+2$ terms,

$$(q,u,p) = \sum_{j=1}^{r} \theta^{(j)} \cdot (q^{(j)}, u^{(j)}, p^{(j)}), \qquad \theta^{(j)} > 0, \quad \sum_{j=1}^{r} \theta^{(j)} = 1,$$

for every point of $\widehat{G}_{k,s(t)}(t;\psi)$, with all component triples in $G_{k,s(t)}(t;\psi)$. For $j \in [r]$, the superscript (j) labels the jth payment-action pair. Its triple $(q^{(j)}, u^{(j)}, p^{(j)})$ records the next-state distribution, expected payoff, and expected payment, and $\theta^{(j)}$ is its assignment probability. In particular, all

components use the same ψ. For the compactness argument, allow zero weights and pad shorter representations to $n+2$ terms. The common ψ, the $n+2$ optimal pairs, and their nonnegative weights summing to one form a closed subset of a compact product. Retaining ψ and averaging the triples is continuous, so its image, the full set of feasible (ψ, q, u, p) after convexification, is compact as well.

We now show that this convex hull describes exactly the outcomes of random assignment. For fixed ψ, any joint distribution supported on optimal payment-action pairs averages points of $G_{k,s(t)}(t;\psi)$. Its expectation therefore belongs to the closed convex set $\widehat{G}_{k,s(t)}(t;\psi)$, even if the distribution has infinite support.

Conversely, take a point of this convex hull and a representation as above. For each component triple, choose a corresponding optimal pair $(f^{(j)}(t+1), a^{(j)}_{k,s(t)}(t))$ and assign it probability $\theta^{(j)}$. Denote the resulting joint distribution by $\Gamma_{k,s(t)}(t)$. To recover the payment distribution, add the weights of all pairs with the same payment vector:

$$\Phi_{k,s(t)}(t, \{f(t+1)\}) = \sum_{j:\, f^{(j)}(t+1)=f(t+1)} \theta^{(j)}.$$

For every payment vector with positive probability, define the conditional action probabilities by

$$\Lambda_{k,s(t)}(t, \{a_{k,s(t)}(t)\} \mid f(t+1)) = \frac{\Gamma_{k,s(t)}(t, \{f(t+1)\}, \{a_{k,s(t)}(t)\})}{\Phi_{k,s(t)}(t, \{f(t+1)\})}.$$

These formulas describe a two-stage draw: first draw a payment vector from $\Phi_{k,s(t)}(t)$, then draw an action according to the conditional probabilities above. The resulting joint distribution is $\Gamma_{k,s(t)}(t)$, so it yields the chosen triple (q, u, p). Every action used after a given payment vector is a best response to that vector.

At unassigned payment vectors, choose best responses using a measurable rule; these choices do not change (q, u, p). The full response rule remains measurable because it differs from this rule only at finitely many assigned vectors, proving (i).

For (ii), the set of feasible payments for fixed (ψ, q, u) is compact and contained in $[0, \bar{f}]$, by (i). If nonempty, it has a minimum in this interval. To see lower semicontinuity, take a convergent sequence of parameter triples $(\psi^{[N]}, q^{[N]}, u^{[N]}) \to (\psi, q, u)$, where the superscript $[N]$ labels the Nth sequence term, for $N = 1, 2, \ldots$. If the lower limit of its minimum costs is finite, choose a subsequence attaining that limit. Compactness of the joint graph makes the limiting payment feasible, and hence

$$\kappa_{k,s(t)}(t, \psi; q, u) \le \liminf_{N\to\infty} \kappa_{k,s(t)}(t, \psi^{[N]}; q^{[N]}, u^{[N]}).$$

If the lower limit is infinite, the inequality is immediate.

For fixed ψ, the set $\widehat{G}_{k,s(t)}(t;\psi)$ is convex by definition. Take two of its triples attaining the minimum payments for their respective transition-payoff pairs. Every convex combination of these triples remains in this set and is implementable by (i). By the definition of κ, the minimum payment at the combined transition-payoff pair is no greater than the combined payment. This is the convexity inequality for κ, completing (ii).

For the support bound, take a finite representation of a minimum-payment triple and keep its r payment-action pairs fixed. Reoptimizing their probabilities gives the linear program

$$\begin{aligned} \min_{\theta^{(j)} \ge 0} \quad & \sum_{j=1}^{r} \theta^{(j)} p^{(j)} \\ \text{s.t.} \quad & \sum_{j=1}^{r} \theta^{(j)} = 1, \qquad \sum_{j=1}^{r} \theta^{(j)} \cdot (q^{(j)}, u^{(j)}) = (q, u). \end{aligned}$$

The original weights are feasible, and every feasible choice implements the same (q, u), so the minimum remains $\kappa_{k,s(t)}(t, \psi; q, u)$. The feasible weight set is a nonempty compact polytope, so an optimal extreme point exists. Including normalization, there are at most $\operatorname{affdim} M_{k,s(t)}(t;\psi) + 1 \le n + 1$ linearly independent equality constraints, by (B.5). An extreme point has at most that many positive weights. Discarding zero weights and using the construction in (i) proves (iii). □

When replacing a local distribution by one with finite support, we preserve the agent's expected payoff u. Agents deciding at earlier dates therefore face the same future payoffs and have the same best-response sets.

We now combine these local representations with the population equations and payoff recursion. Let $\mathcal{E}(\bar{\mu})$ be the set of pairs $(\Psi, \boldsymbol{q})$ of candidate value and transition arrays satisfying

$$\begin{cases} \Psi \in [-\bar{\Psi}, \bar{\Psi}]^{mn(T+1)}, & \boldsymbol{q} \in \mathcal{Q}(\bar{\mu}), \\ \Psi_{k,s(T+1)}(T+1) = 0, & \\ \kappa_{k,s(t)}(t, \Psi_k(t+1); q_{k,s(t)}(t), \Psi_{k,s(t)}(t)) < +\infty, & k \in [m], \quad s(t) \in \Omega, \quad t \in [T]. \end{cases}$$

The terminal condition holds for every type and terminal state. Local feasibility is imposed at every decision state, including those with zero prescribed mass. The following characterization completes the proof.

Lemma B.2. *A target $\bar{\mu} \in \mathcal{T}$ is implementable if and only if $\mathcal{E}(\bar{\mu})$ is nonempty. Its least expected payment is*

$$\mathrm{OPT}(\bar{\mu}) = \inf_{(\Psi,\boldsymbol{q}) \in \mathcal{E}(\bar{\mu})} \sum_{t \in [T]} \sum_{k \in [m]} \sum_{s(t) \in \Omega} \bar{\mu}_{k,s(t)}(t) \cdot \kappa_{k,s(t)}(t, \Psi_k(t+1); q_{k,s(t)}(t), \Psi_{k,s(t)}(t)). \quad \text{(B.6)}$$

The infimum is attained when feasible and equals $+\infty$ otherwise. Each feasible $(\Psi, \boldsymbol{q})$ admits an implementation at its objective value satisfying

$$\left|\operatorname{supp} \Gamma_{k,s(t)}(t)\right| \leq \operatorname{affdim} M_{k,s(t)}(t; \Psi_k(t+1)) + 1 \leq n + 1 \quad \text{(B.7)}$$

for every type, current state, and decision date.

Proof. First take an admissible outcome generating $\bar{\mu}$. Set its candidate values equal to the actual pre-offer values from Lemma 2.1. Define its transition and expected-payment coordinates by

$$q_{k,s(t)}(t) := \mathbb{E}\left[P_{k,s(t)}(a_{k,s(t)}(t))\right], \qquad p_{k,s(t)}(t) := \mathbb{E}\left[P_{k,s(t)}(a_{k,s(t)}(t)) \cdot f(t+1)\right].$$

Both expectations average over the offer and action under $\Gamma_{k,s(t)}(t)$. Population accounting gives $\boldsymbol{q} \in \mathcal{Q}(\bar{\mu})$. Best-response support and (B.3) give

$$\left(q_{k,s(t)}(t), \Psi_{k,s(t)}(t), p_{k,s(t)}(t)\right) \in \widehat{G}_{k,s(t)}(t; \Psi_k(t+1)).$$

Thus $(\Psi, \boldsymbol{q}) \in \mathcal{E}(\bar{\mu})$, and the local minimum payment is no larger than $p_{k,s(t)}(t)$. Multiplying by the prescribed masses and summing proves that the right side of (B.6) is at most the payment of any admissible implementation.

Conversely, take $(\Psi, \boldsymbol{q}) \in \mathcal{E}(\bar{\mu})$. At every type, state, and date, Lemma B.1 gives a joint distribution with at most $n+1$ payment-action pairs realizing

$$\left(q_{k,s(t)}(t), \Psi_{k,s(t)}(t), \kappa_{k,s(t)}(t, \Psi_k(t+1); q_{k,s(t)}(t), \Psi_{k,s(t)}(t))\right) \in \widehat{G}_{k,s(t)}(t; \Psi_k(t+1)).$$

Recover the indexed offer and response distributions as in that lemma, specifying best responses also at unassigned offers. Their expected-payoff coordinates satisfy (B.3). Starting from the zero terminal values, backward induction therefore identifies the candidate values with the actual optimal payoffs and makes every selected action a best response. The resulting outcome is admissible by Lemma 2.1. Starting from the given initial masses, forward induction using (B.1) gives the target trajectory. Its payment equals the objective in (B.6), proving the reverse inequality and the reconstruction claim.

Finally, $\mathcal{E}(\bar{\mu})$ is compact: values and transition arrays are bounded, the terminal and flow equations are closed, and Lemma B.1 makes each local feasible set closed. The objective in (B.6) is a finite sum of lower-semicontinuous functions with fixed nonnegative weights. At zero-mass states the corresponding term is zero, while the local cost remains finite. The minimum is therefore attained whenever the feasible set is nonempty. Applying the local support bound in the reconstruction gives (B.7). □

Theorem 3.1. *A target trajectory $\bar{\mu} \in \mathcal{T}$ is implementable if and only if it admits an admissible implementation using at most $n+1$ payment-action pairs for each type, current state, and decision date. Whenever $\bar{\mu}$ is implementable, its least total expected payment is attained by an implementation of this form. Such an optimal implementation satisfies $W = \mathrm{OPT}(\bar{\mu})$ and*

$$\left|\operatorname{supp} \Phi_{k,s(t)}(t)\right| \le \left|\operatorname{supp} \Gamma_{k,s(t)}(t)\right| \le n+1, \qquad k \in [m], \quad s(t) \in \Omega, \quad t \in [T].$$

Proof. Suppose $\bar{\mu}$ is implementable. By Lemma B.2, $\mathcal{E}(\bar{\mu})$ contains a minimizer $(\Psi, \boldsymbol{q})$ of (B.6). Its reconstruction gives an admissible implementation with $W = \mathrm{OPT}(\bar{\mu})$ and

$$\left|\operatorname{supp} \Gamma_{k,s(t)}(t)\right| \le \operatorname{affdim} M_{k,s(t)}(t; \Psi_k(t+1)) + 1 \le n+1$$

for every type, current state, and decision date. Each $\Phi_{k,s(t)}(t)$ is the payment marginal of $\Gamma_{k,s(t)}(t)$, so its support has no more points than the joint support. This proves attainment and the stated support bounds. Conversely, an admissible implementation with these bounds implements $\bar{\mu}$, proving the equivalence. □

C JOINT SUPPORT REDUCTION ACROSS CURRENT STATES

This appendix proves Theorem 3.2, using the local representation developed in Appendix B. We study the total support required across current states for a fixed type and decision date. Starting from an optimal implementation, we jointly reoptimize assignment probabilities while preserving the target population flow. At dates after the first, we also preserve relative continuation values so that earlier incentives are unchanged. An extreme point of the resulting linear program yields a total support bound linear in n while retaining the minimum payment. We restate the theorem before giving its proof.

Theorem 3.2. *Every implementable target $\bar{\mu} \in \mathcal{T}$ admits an optimal implementation with $W = \mathrm{OPT}(\bar{\mu})$ such that*

$$\sum_{s(t)\in\Omega} \left|\operatorname{supp} \Gamma_{k,s(t)}(t)\right| \le \begin{cases} 2n-1, & t = 1, \\ 3n-2, & t \in \{2, \dots, T\}, \end{cases} \qquad k \in [m]. \tag{3.1}$$

The same implementation satisfies $\left|\operatorname{supp} \Phi_{k,s(t)}(t)\right| \le \left|\operatorname{supp} \Gamma_{k,s(t)}(t)\right| \le n+1$ for every type, current state, and decision date. Each group with $\bar{\mu}_{k,s(t)}(t) = 0$ uses a single payment-action pair.

Proof. Start from an optimal implementation supplied by Theorem 3.1, with at most $n+1$ pairs in every group, and let Ψ denote its actual continuation values. We first explain which changes in these values preserve earlier incentives. Write $\mathbf{1} \in \mathbb{R}^{\Omega}$ for the all-ones vector. Since every transition vector sums to one, (B.2) gives

$$\begin{aligned} V_{k,s(\tau)}(\tau, f(\tau+1); \psi + \eta\mathbf{1}) &= V_{k,s(\tau)}(\tau, f(\tau+1); \psi) + \lambda_k \eta, \\ \mathrm{BR}_{k,s(\tau)}(\tau, f(\tau+1); \psi + \eta\mathbf{1}) &= \mathrm{BR}_{k,s(\tau)}(\tau, f(\tau+1); \psi), \end{aligned} \tag{C.1}$$

whenever both next-date vectors lie in K. If a change at date $t \ge 2$ shifts $\Psi_k(t)$ by $\eta\mathbf{1}$, keep all earlier offer and response distributions unchanged. Backward induction in (B.3) shifts the date-τ value vector by $\lambda_k^{t-\tau}\eta\mathbf{1}$ for every $\tau < t$. Thus every earlier best-response set is unchanged, including at zero-mass states and after offers assigned zero probability.

Fix a type k and date t, leaving all other dates and types unchanged, and put $\psi := \Psi_k(t+1)$. Within this step, write s for a current state and s' for a next state, and abbreviate

$$\bar{\mu}_s := \bar{\mu}_{k,s}(t), \qquad \bar{\mu}'_{s'} := \bar{\mu}_{k,s'}(t+1), \qquad \Psi_s := \Psi_{k,s}(t).$$

At state s, list the distinct pairs in the current joint support as $(f_s^{(j)}, a_s^{(j)})$, $j \in [r_s]$. The payment vector is offered at date t and paid at date $t+1$, and the action is taken at date t. Define their transition, payoff, and payment coordinates by

$$q_s^{(j)} := P_{k,s}(a_s^{(j)}), \quad u_s^{(j)} := V_{k,s}(t, f_s^{(j)}; \psi), \quad p_s^{(j)} := P_{k,s}(a_s^{(j)}) \cdot f_s^{(j)}.$$

Each listed pair belongs to $Z_{k,s}(t; \psi)$. Let $q_s^{(j)}(s')$ be its transition probability to s'.

Choose a reference state $s_0 \in \Omega$. Over nonnegative assignment probabilities $\theta_s^{(j)}$, minimize

$$\sum_{s\in\Omega} \bar{\mu}_s \sum_{j=1}^{r_s} \theta_s^{(j)} p_s^{(j)} \tag{C.2}$$

subject to normalization and aggregate flow constraints,

$$\begin{aligned} \sum_{j=1}^{r_s} \theta_s^{(j)} &= 1, && s \in \Omega, \\ \sum_{s\in\Omega} \bar{\mu}_s \sum_{j=1}^{r_s} \theta_s^{(j)} q_s^{(j)}(s') &= \bar{\mu}'_{s'}, && s' \in \Omega \setminus \{s_0\}. \end{aligned} \tag{C.3}$$

When $t \geq 2$, also impose

$$\sum_{j=1}^{r_s} \theta_s^{(j)} u_s^{(j)} - \sum_{j=1}^{r_{s_0}} \theta_{s_0}^{(j)} u_{s_0}^{(j)} = \Psi_s - \Psi_{s_0}, \qquad s \in \Omega \setminus \{s_0\}. \tag{C.4}$$

The original probabilities are feasible. The omitted flow equation for s_0 follows from normalization, the unit sums of all transition vectors, and conservation of the type's total mass. Equation (C.4) preserves every continuation-value difference, so the new date-t value vector differs from the old one by a common constant. All states, including those with zero prescribed mass, enter the normalization and relative-value constraints. At date 1, there are no earlier incentives, so no value constraints are needed.

Every feasible choice of these weights gives an admissible implementation of the same target. Recover the offer distribution and its conditional action distributions as in Lemma B.1(i), completing responses at unassigned offers by a Borel best-response rule. The next-date vector ψ is fixed, so all retained actions remain optimal at date t. Equation (C.1) preserves all earlier best responses. Earlier population flows are unchanged, (C.3) preserves the date-t flow, and the unchanged later distributions generate the same later masses. Hence (C.2) is the only contribution to total payment that can change.

The feasible weight set is a nonempty compact polytope, so it has an optimal extreme point. Global optimality of the starting implementation implies that its minimum objective equals the original date-t, type-k payment contribution. There are n normalization equations and $n-1$ flow equations, together with $n-1$ relative-value equations when $t \geq 2$. An extreme point has no more positive coordinates than the rank of these equations: otherwise their columns on the positive coordinates would admit a nonzero null direction, and sufficiently small perturbations of both signs would contradict extremality. The total support is therefore at most $2n-1$ at date 1 and $3n-2$ at later dates. No payment equality is required, because payment is minimized. Only existing pairs are reweighted or removed, so each group's support remains at most $n+1$.

Apply this replacement to every type, processing dates $T, T-1, \ldots, 1$ and using the actual continuation values of the current implementation at each step. Every replacement preserves the target and the global minimum payment. Processing an earlier date leaves every later joint distribution unchanged, so all the bounds in (3.1) hold simultaneously.

Finally, consider a group with $\bar{\mu}_{k,s(t)}(t) = 0$. With $\psi := \Psi_k(t+1)$ fixed, put $v_0 := V_{k,s(t)}(t, 0; \psi)$. The payment bounds imply

$$v_0 \leq V_{k,s(t)}(t, f(t+1); \psi) \leq v_0 + \bar{f}.$$

Consequently, $\eta := \Psi_{k,s(t)}(t) - v_0$ lies in $[0, \bar{f}]$. Offer the constant vector $\eta\mathbf{1}$ with probability one and select any best response to the zero offer. Adding η to every payment coordinate adds exactly η to every action's payoff, so the selected action remains optimal and the group's continuation value is preserved. The group's zero mass makes its population-flow and payment contributions zero. Earlier values and incentives are unchanged; complete the response rule at other offers by a Borel best-response selection. Repeating this replacement at every zero-mass group leaves all support bounds intact. Taking payment marginals cannot increase support, proving the remaining marginal-support claim. □

D TIGHTNESS OF THE JOINT SUPPORT BOUNDS

This appendix proves Proposition 3.3, restated below. The construction attains both per-date bounds in a single two-decision instance, and the second-date lower bound also applies to the sum of payment-support sizes across groups.

Proposition 3.3. *For every $n \geq 2$, there is an instance with one type, $T = 2$, action sets $A_{1,s} = [0, 1]$, and continuous, time-independent payoff and transition functions, together with an implementable target $\bar{\mu}$, such that*

$$\sum_{s\in\Omega} |\operatorname{supp} \Gamma_{1,s}(1)| \geq 2n - 1 \qquad \textit{for every admissible implementation}, \tag{3.2}$$

$$\sum_{s\in\Omega} |\operatorname{supp} \Gamma_{1,s}(2)| \geq 3n - 2 \qquad \textit{for every optimal implementation}. \tag{3.3}$$

One optimal implementation attains both bounds simultaneously. Moreover, $\bar{\mu}_{1,s}(2) > 0$ for every state, and every optimal implementation satisfies

$$|\operatorname{supp} \Phi_{1,s}(2)| \geq 3 \quad \textit{for each } s \in [n-1], \qquad \sum_{s\in\Omega} |\operatorname{supp} \Phi_{1,s}(2)| \geq 3n - 2.$$

The same implementation attains this payment-support bound.

Proof. Let $\Omega = [n]$, let $\boldsymbol{e}_s$ denote the unit probability vector at state s, and set

$$\lambda_1 = \frac{49}{50}, \qquad \bar{f} = 5, \qquad c_\star = 5 + \lambda_1 \frac{23}{10} = \frac{3627}{500}.$$

We first specify finitely many basic actions at each state. At the end of the proof, we embed these actions into $[0, 1]$ using continuous primitives, without introducing any additional best responses. For each $s < n$, the four basic actions $a_s^{(j)}$, $j \in [4]$, have primitives

$$P_{1,s}(a_s^{(j)}) = x^{(j)}\boldsymbol{e}_s + (1 - x^{(j)})\boldsymbol{e}_n, \qquad b_{1,s}(a_s^{(j)}) = b^{(j)},$$

where

j	1	2	3	4
$x^{(j)}$	$1/5$	$2/5$	$3/5$	$4/5$
$b^{(j)}$	1	$4/5$	0	$-41/50$

At state n, there are n basic actions: g_s for each $s < n$, and h, with

$$P_{1,n}(g_s) = \boldsymbol{e}_s, \qquad b_{1,n}(g_s) = -c_\star,$$
$$P_{1,n}(h) = \boldsymbol{e}_n, \qquad b_{1,n}(h) = 0.$$

These primitives are the same at both decision dates. Write $\bar{\mu}_s := 1/[2(n-1)]$ for $s < n$. The target trajectory is

Current state	$\bar{\mu}_{1,s}(1)$	$\bar{\mu}_{1,s}(2)$	$\bar{\mu}_{1,s}(3)$
$s < n$	0	$\bar{\mu}_s$	$\frac{3}{5}\bar{\mu}_s$
$s = n$	1	$1/2$	$7/10$

We now derive restrictions on continuation values and the first-date support. Consider any admissible implementation of this target, and abbreviate its actual before-offer values at date 2 by $\Psi_s := \Psi_{1,s}(2)$. At state n and date 2, every g_s yields payoff at most $-c_\star + 5 < 0$, whereas h yields the nonnegative payment $f(3, n)$. Thus h is the unique best response, $\Psi_n \geq 0$, and this group's contribution to total payment is $\Psi_n/2$. Since only state $s < n$ can send mass to itself at this date, its average transition probability satisfies

$$\mathbb{E}[x^{(j)}] = \frac{3}{5}, \qquad s < n, \tag{D.1}$$

where the expectation is over $\Gamma_{1,s}(2)$ and j indexes the selected basic action.

At date 1, all population mass is in state n. The target requires action g_s with probability $\bar{\mu}_s$ for each $s < n$, and action h with probability $1/2$. Whenever g_s is selected, its optimality relative to h implies

$$-c_\star + f(2,s) + \lambda_1 \Psi_s \geq f(2,n) + \lambda_1 \Psi_n.$$

Consequently, the payment bounds give

$$\Psi_s - \Psi_n \geq \frac{23}{10}, \qquad s < n. \tag{D.2}$$

Let $W(t)$ be the expected payment generated by decision date t, so that $W = W(1) + W(2)$. Dropping the nonnegative payments following h, we obtain

$$W(1) \geq \sum_{s<n} \bar{\mu}_s \big[c_\star - \lambda_1(\Psi_s - \Psi_n)\big]. \tag{D.3}$$

Moreover, the support of $\Gamma_{1,n}(1)$ must contain pairs involving at least n distinct actions. Each of the other $n-1$ groups must have a nonempty joint support, even though it has zero initial mass. This proves (3.2).

To bound second-date payments, fix $s < n$. For an offer $f(3)$ and a selected basic action j, write

$$\delta := f(3,s) - f(3,n), \qquad u := V_{1,s}(2, f(3); 0) = b^{(j)} + f(3,n) + x^{(j)}\delta.$$

Comparing the four affine action payoffs gives the following best-response intervals and smallest compatible optimized payoffs $\underline{u}^{(j)}$:

j	Best-response interval for δ	$\underline{u}^{(j)}$	Attaining $(f(3,s), f(3,n))$
1	$[-5, 1]$	1	$(0,0)$
2	$[1, 4]$	$6/5$	$(1,0)$
3	$[4, 41/10]$	$12/5$	$(4,0)$
4	$[41/10, 5]$	$123/50$	$(41/10, 0)$

For actions $2, 3, 4$, equality $u = \underline{u}^{(j)}$ forces the two payment coordinates listed in the last column. Direct substitution of the four pairs $(x^{(j)}, b^{(j)})$, together with $u \geq \underline{u}^{(j)}$, gives

$$\begin{aligned} b^{(j)} &\leq \frac{12}{5} - 4x^{(j)}, \\ b^{(j)} &\leq -\frac{78}{19}\left(x^{(j)} - \frac{3}{5}\right) + \frac{u - 12/5}{57}. \end{aligned} \tag{D.4}$$

The second inequality is strict for action 1; for each of actions $2, 3, 4$, it is an equality exactly when $u = \underline{u}^{(j)}$. Taking expectations under $\Gamma_{1,s}(2)$, using (D.1) and $\mathbb{E}[u] = \Psi_s$, yields

$$\mathbb{E}[b^{(j)}] \leq \min\left\{0, \frac{\Psi_s - 12/5}{57}\right\}.$$

The group's expected payment per agent is $\Psi_s - \mathbb{E}[b^{(j)}]$, so

$$W(2) \geq \sum_{s<n} \bar{\mu}_s \left[\Psi_s - \min\left\{0, \frac{\Psi_s - 12/5}{57}\right\}\right] + \frac{\Psi_n}{2}.$$

Combining this inequality with (D.3) gives

$$W \geq \frac{c_\star}{2} + \sum_{s<n} \bar{\mu}_s \varphi(\Psi_s) + \frac{99}{100}\Psi_n,$$

where $\varphi(u) := u/50 - \min\{0, (u - 12/5)/57\}$. This function is strictly increasing: its slopes below and above $12/5$ are $7/2850$ and $1/50$, respectively. By (D.2), $\Psi_s \geq 23/10 + \Psi_n$ for every $s < n$, and $\Psi_n \geq 0$. Hence

$$W \geq \frac{c_\star}{2} + \frac{1}{2}\varphi(23/10) = \frac{2081}{570}, \tag{D.5}$$

and equality requires

$$\Psi_s = \frac{23}{10} \quad (s < n), \qquad \Psi_n = 0. \tag{D.6}$$

We next construct an implementation attaining the payment lower bound. At date 2, assign the following three payment-action pairs to each state $s < n$:

Pair	$(\boldsymbol{e}_s, a_s^{(2)})$	$(4\boldsymbol{e}_s, a_s^{(3)})$	$((41/10)\boldsymbol{e}_s, a_s^{(4)})$
Probability	$5/57$	$47/57$	$5/57$

All selected actions are best responses. These probabilities give average transition weight $3/5$, before-offer payoff $23/10$, and average nontransfer payoff $-1/570$. At state n, assign the single pair $(0, h)$.

At date 1 and state n, offer the single vector $f^\star = 5\sum_{s<n} \boldsymbol{e}_s$. Select g_s with probability $\bar{\mu}_s$ for each $s < n$, and select h with probability $1/2$. In view of (D.6), all these actions yield payoff zero. At each zero-mass state $s < n$, assign the single pair $(0, a_s^{(2)})$. This action is optimal because the effective difference between the two destination payoffs is $\lambda_1(23/10) = 1127/500 \in (1, 4)$. Take payment marginals to obtain Φ, and use the indicated conditional action probabilities to obtain Λ. Complete the response rule at all unassigned offers by choosing the first maximizing basic action in a fixed ordering; this is a Borel rule.

The resulting contract implements the target and has payment

$$W = \frac{5}{2} + \frac{1}{2}\left(\frac{23}{10} + \frac{1}{570}\right) = \frac{2081}{570}.$$

It is therefore optimal. Its total joint support equals $2n - 1$ at date 1 and $3n - 2$ at date 2, and its total payment support at date 2 is also $3n - 2$.

To show that at least three offers are necessary in each state $s < n$ at date 2, consider any optimal implementation. Its payment must attain (D.5), so its continuation values satisfy (D.6). All the gaps used in the payment lower bound are nonnegative, and $\bar{\mu}_s > 0$ for each $s < n$. Consequently, each such group must satisfy

$$\mathbb{E}[b^{(j)}] = \frac{23/10 - 12/5}{57} = -\frac{1}{570},$$

and equality must hold almost surely in (D.4). Thus action 1 cannot be used; actions $2, 3, 4$ can only be used with respective optimized payoffs $6/5$, $12/5$, and $123/50$, and with the payment coordinates in the best-response table.

Let $\theta^{(2)}, \theta^{(3)}, \theta^{(4)}$ be their probabilities at a fixed state. Normalization, the required transition, and the before-offer payoff give

$$\begin{aligned} \theta^{(2)} + \theta^{(3)} + \theta^{(4)} &= 1, \\ \frac{2}{5}\theta^{(2)} + \frac{3}{5}\theta^{(3)} + \frac{4}{5}\theta^{(4)} &= \frac{3}{5}, \\ \frac{6}{5}\theta^{(2)} + \frac{12}{5}\theta^{(3)} + \frac{123}{50}\theta^{(4)} &= \frac{23}{10}. \end{aligned}$$

Their unique solution is $(\theta^{(2)}, \theta^{(3)}, \theta^{(4)}) = (5, 47, 5)/57$. Each action therefore has positive probability, and their required payment coordinates are distinct. It follows that both $\Gamma_{1,s}(2)$ and $\Phi_{1,s}(2)$ have support size at least three. The remaining state contributes at least one support point, proving (3.3) and the payment-support claim.

Finally, we extend the construction to a continuous scalar action space. Fix a state and let r be its number of basic actions, equal to 4 for $s < n$ and to n for $s = n$. Place these actions, in any fixed order, at the knots $z^{(j)} = (j-1)/(r-1)$ of $[0,1]$. Write $(P^{(j)}, b^{(j)})$ for their prescribed primitives at this state. Between two adjacent knots, set $\alpha = (a - z^{(j)})/(z^{(j+1)} - z^{(j)})$ and define

$$P_{1,s}(a) = (1-\alpha)P^{(j)} + \alpha P^{(j+1)},$$
$$b_{1,s}(a) = (1-\alpha)b^{(j)} + \alpha b^{(j+1)} - \alpha(1-\alpha).$$

These functions are continuous and time-independent, and the transition vectors remain probability vectors. For any offer f and next-date value vector ψ, let $u^{(j)}$ and $u^{(j+1)}$ be the total action payoffs at the two endpoints. At any interior point of the interval, the total payoff is

$$(1-\alpha)u^{(j)} + \alpha u^{(j+1)} - \alpha(1-\alpha) < \max\{u^{(j)}, u^{(j+1)}\}.$$

Thus only the basic actions can ever be best responses, including at unassigned offers and zero-mass groups. All preceding arguments and the constructed implementation therefore remain valid with $d = 1$ and $A_{1,s} = [0,1]$ at every state. □