Distributionally Robust Optimal Allocation with Costly Verification

Halil İbrahim Bayrak
Department of Industrial Engineering, Bilkent University, Turkey, halil.bayrak@bilkent.edu.tr,

Çağıl Koçyiğit
Luxembourg Centre for Logistics and Supply Chain Management, University of Luxembourg, Luxembourg, cagil.kocyigit@uni.lu,

Daniel Kuhn
Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland, daniel.kuhn@epfl.ch,

Mustafa Çelebi Pınar
Department of Industrial Engineering, Bilkent University, Turkey, mustafap@bilkent.edu.tr,

We consider the mechanism design problem of a principal allocating a single good to one of several agents without monetary transfers. Each agent desires the good and uses it to create value for the principal. We designate this value as the agent's private type. Even though the principal does not know the agents' types, she can verify them at a cost. The allocation of the good thus depends on the agents' self-declared types and the results of any verification performed, and the principal's payoff matches her value of the allocation minus the costs of verification. It is known that if the agents' types are independent, then a favored-agent mechanism maximizes her expected payoff. However, this result relies on the unrealistic assumptions that the agents' types follow known independent probability distributions. In contrast, we assume here that the agents' types are governed by an ambiguous joint probability distribution belonging to a commonly known ambiguity set and that the principal maximizes her worst-case expected payoff. We study support-only ambiguity sets, which contain all distributions supported on a rectangle, Markov ambiguity sets, which contain all distributions in a support-only ambiguity set satisfying some first-order moment bounds, and Markov ambiguity sets with independent types, which contain all distributions in a Markov ambiguity set under which the agents' types are mutually independent. In all cases, we construct explicit favored-agent mechanisms that are not only optimal but also Pareto robustly optimal.

Key words: mechanism design; costly verification; distributionally robust optimization; ambiguity aversion

---

1. Introduction

Consider a principal ("she") who allocates a good to one of several agents without using monetary transfers. Each agent ("he") derives strictly positive utility from owning the good and has a private type, which reflects the value he creates for the principal if receiving the good. The principal is unaware of the agents' types but can verify any of them at a cost. Any verification will perfectly reveal the corresponding agent's type to the principal. The good is allocated based on the agents' self-declared types as well as the results of any verification performed. The principal aims to design an allocation mechanism that maximizes her payoff, *i.e.*, the value of allocation minus any costs of verification.

This generic mechanism design problem arises in many different contexts. For example, the rector of a university may have funding for a new faculty position and needs to allocate it to one of the school's departments, the ministry of health may need to decide in which town to open up a new hospital, a venture capitalist may need to select a start-up business that should receive seed funding, the procurement manager of a manufacturing company may need to choose one of several suppliers, or a consulting company may need to identify a team that leads a new project. In all of these examples, the principal wishes to put the good into use where it best contributes to her organization or the society as a whole. Each agent desires the good and is likely to be well-informed about the value he will generate for the principal if he receives the good. In addition, monetary transfers may be inappropriate in all of the described situations, but the principal can collect information through costly investigation or audit.

Mechanism design problems of the above type are usually referred to as 'allocation with costly verification.' Ben-Porath et al. [5] describe the first formal model for their analysis and introduce the class of favored-agent mechanisms, which are attractive because of their simplicity and interpretability. As in most of the literature on mechanism design, Ben-Porath et al. [5] model the agents' types as independent random variables governed by a commonly known probability distribution, which allows them to prove that any mechanism that maximizes the principal's expected payoff is a randomization over favored-agent mechanisms. Any favored-agent mechanism is characterized by a favored agent and a threshold value, and it assigns the good to the favored agent without verification whenever the reported types of all other agents—adjusted for the costs of verification—fall below the given threshold. Otherwise, it allocates the good to any agent for which the reported type minus the cost of verification is maximal and verifies his reported type. This mechanism is incentive compatible, that is, no agent has an incentive to misreport his true type; see Section 2 for more details.

The vast majority of the literature on allocation with costly verification (see, *e.g.*, [17, 18] and the references therein) sustains the modeling assumptions of Ben-Porath et al. [5], thus assuming that the agents' types are independent random variables and that their distribution is common knowledge. In reality, however, it is often difficult to justify the precise knowledge of such a distribution. This prompts us to study allocation problems with costly verification under the more realistic assumption that the principal has only partial information about the distribution of the agents' types. Specifically, we assume that the distribution of the agents' types is unknown but belongs to a commonly known ambiguity set (*i.e.*, a family of multiple—perhaps infinitely many—distributions). In addition, we assume that the principal is ambiguity averse in the sense that she wishes to maximize her worst-case expected payoff in view of all distributions in the ambiguity set. Under these assumptions, the mechanism design problem at hand can be cast as a zero-sum game between the principal, who chooses a mechanism to allocate the good, and some fictitious adversary, who chooses the distribution of the agents' types from the ambiguity set in order to inflict maximum damage to the principal. Using techniques from distributionally robust optimization (see, *e.g.*, [11, 24]), we characterize optimal and Pareto robustly optimal mechanisms for well-known classes of ambiguity sets: (i) support-only ambiguity sets containing all distributions supported on a rectangle, (ii) Markov ambiguity sets containing all distributions in a support-only ambiguity set whose mean values fall within another (smaller) rectangle, and (iii) Markov ambiguity sets with independent types containing all distributions in a Markov ambiguity set under which the agents' types are mutually independent. We emphasize that support-only as well as Markov ambiguity sets contain distributions under which the agents' types are mutually dependent. Pareto robust optimality is an important solution concept in robust optimization [14]. In the distributionally robust context considered here, a mechanism is called Pareto robustly optimal if it is not Pareto robustly dominated, that is, if there is no other mechanism that generates a non-inferior expected payoff under every distribution and a strictly higher expected payoff under at least one distribution in the ambiguity set. Every Pareto robustly optimal mechanism is also robustly optimal, but the converse implication is not true in general. Mechanisms that fail to be Pareto robustly optimal would not be used by any rational agent.

The concept of Pareto robust optimality was invented because robustly optimal solutions are often highly degenerate, that is, typical (distributionally) robust optimization problems admit a multitude of robustly optimal solutions that all attain the same worst-case expected payoff. However, most of these solutions underperform when average (non-worst-case) conditions prevail [14].

This phenomenon is particularly pronounced in *adjustable* robust optimization [6]. We emphasize that the mechanism design problem studied in this paper can be viewed as an adjustable (distributionally) robust optimization problem because the allocation probabilities encoding different mechanisms represent functions of the agents' unknown types. We will show that this problem suffers indeed from massive solution degeneracy. We emphasize that this degeneracy cannot be avoided if the goal is to optimize worst-case performance in the face of distributional ambiguity.

The Pareto robustly optimal mechanisms form a small subset of the family of all robustly optimal mechanisms, and they are the only robustly optimal mechanisms of practical value. Indeed, any other robustly optimal mechanism unnecessarily sacrifices performance under at least one distribution in the ambiguity set. Seeking Pareto robustly optimal mechanisms is therefore a natural goal. By identifying Pareto robustly optimal *favored-agent* mechanisms, we also establish a bridge to the classical theory of allocation with costly verification [5]. On the methodological front, we develop a new *spatial induction* technique for proving Pareto robust optimality. This technique exploits the lack of locality of the incentive compatibility constraints in our robust mechanism design problem. In fact, while the constraints of standard adjustable robust optimization problems are *local* in the sense that they are separable with respect to different uncertainty realizations, the incentive compatibility constraints of our robust mechanism design problem are *non-local* in the sense that they couple the allocation probabilities of any feasible mechanism across different types. We exploit this non-locality to prove the Pareto robust optimality of a judiciously chosen favorite agent mechanism as follows. For the sake of contradiction, we first assume that there exists another feasible mechanism that Pareto robustly dominates the chosen favorite agent mechanism. Second, we use spatial induction to prove that both mechanisms must generate the *same* payoff in *every* scenario in the type space. To this end, we partition the type space into disjoint subsets that are tailored to the problem data and—in particular—to the ambiguity set at hand. We then use elementary arguments to prove that the two mechanisms generate the same payoff in every scenario in a first (particularly benign) subset; this is the *base step* of our spatial induction. Next, we use the non-locality of the incentive compatibility constraints to relate any scenario in the second subset to a scenario in the first subset of the type space. Exploiting our knowledge that the two mechanisms generate the same payoff throughout the first subset, we can then prove that they also generate the same payoff throughout the second subset. This is the first *induction step*. We then iterate through the remaining subsets of the type space one by one and apply each time a similar induction step that exploits the non-locality of the incentive compatibility constraints. This eventually proves that both mechanisms generate the same payoff throughout the *entire* type space, which contradicts the initial assumption that the prescribed favorite-agent mechanism is Pareto robustly dominated by some other mechanism. Hence, it is Pareto robustly optimal. Increasingly involved versions of this conceptual proof strategy based on spatial induction will be used to analyze support-only ambiguity sets as well as Markov ambiguity sets with and without independent types.

The main contributions of this paper can now be summarized as follows.

(i) For support-only ambiguity sets, we first show that not every robustly optimal mechanism represents a randomization over favored-agent mechanisms. This result is unexpected in view of the classical theory on stochastic mechanism design [5]. We then construct an explicit favored-agent mechanism that is not only robustly optimal but also Pareto robustly optimal. This mechanism selects the favored agent from among those whose types have the highest possible lower bound, and it sets the threshold to this lower bound.

(ii) For Markov ambiguity sets, we also construct an explicit favored-agent mechanism that is both robustly optimal as well as Pareto robustly optimal. This mechanism selects the favored agent from among those whose *expected* types have the highest possible lower bound, and it sets the threshold to the highest possible *actual* (not *expected*) type of the favored agent.

(iii) For Markov ambiguity sets with independent types, we identify again a favored-agent mechanism that is robustly optimal as well as Pareto robustly optimal. Here, the favored agent is chosen exactly as under an ordinary Markov ambiguity set, but the threshold is set to the lowest possible *expected* (not *actual*) type of the favored agent.

(iv) We develop a new spatial induction technique to prove the Pareto robust optimality of the above favored-agent mechanisms. This technique crucially exploits the non-locality of the incentive compatibility constraints of our mechanism design problem.

Our results show that favored-agent mechanisms continue to play an important role in allocation with costly verification even if the unrealistic assumption of a commonly known type distribution is abandoned. In addition, they suggest that robust optimality alone is not a sufficiently distinctive criterion to single out practically useful mechanisms under distributional ambiguity. However, our results also show that among possibly infinitely many robustly optimal mechanisms, one can always find a simple and interpretable Pareto robustly optimal favored-agent mechanism. Unlike in the classical theory that assumes the type distribution to be known [5], the favored agent as well as the threshold of our Pareto robustly optimal mechanisms are *in*dependent of the verification costs.

Literature review. The first treatise of allocation with costly verification is due to Townsend [21], who studies a principal-agent model with monetary transfers involving a single agent. Ben-Porath et al. [5] extend this model to multiple agents but rule out the possibility of monetary transfers. Their seminal work has inspired considerable follow-up research in economics. For example, Mylovanov and Zapechelnyuk [18] study a variant of the problem where verification is costless, but the principal can impose only limited penalties and only partially recover the good when agents misreport their types. Li [17] accounts both for costly verification and for limited penalties, thereby unifying the models in [5] and [18]. Chua et al. [10] further extend the model in [5] to multiple homogeneous goods, assuming that each agent can receive at most one good. Bayrak et al. [4] spearhead the study of allocation with costly verification under distributional ambiguity. However, for reasons of computational tractability, they focus on ambiguity sets that contain only two discrete distributions. In this paper, we investigate ambiguity sets that contain infinitely many (not necessarily discrete) type distributions characterized by support and moment constraints, and we derive robustly as well as Pareto robustly optimal mechanisms in closed form.

This paper also contributes to the growing literature on (distributionally) robust mechanism design. Note that any mechanism design problem is inherently affected by uncertainty due to the private information held by different agents. The vast majority of the extant mechanism design literature models uncertainty through random variables that are governed by a commonly known probability distribution. The robust mechanism design literature, on the other hand, explicitly accounts for (non-stochastic) distributional uncertainty and seeks mechanisms that maximize the worst-case payoff, minimize the worst-case regret or minimize the worst-case cost in view of all distributions consistent with the information available. Robust mechanism design problems have recently emerged in different contexts such as pricing (see, *e.g.*, [2, 7, 9, 16, 19, 23]), auction design (see, *e.g.*, [1, 3, 13, 15, 20]) or contracting (see, *e.g.*, [22]). This literature is too vast to be discussed in detail. To our best knowledge, however, we are the first to derive closed-form optimal and Pareto robustly optimal mechanisms for the allocation problem with costly verification under distributional ambiguity. Our paper is most closely related to the independent concurrent work by Chen et al. [8], who also study allocation problems with costly verification under distributional uncertainty. They assume that the agents have only access to a signal that correlates with their (unknown) types and that the principal has only access to the signal distribution, which is selected by a fictitious information designer. They identify the worst- and best-case signal distributions for the principal and the best-case signal distributions for the agents. They also study a distributionally robust mechanism design problem over a (what we call a) Markov ambiguity set, where the agents' types have known means. However, Chen et al. [8] do not address the multiplicity of robustly optimal mechanisms, and consequently they do not identify Pareto robustly optimal mechanisms.

The remainder of this paper is structured as follows. Section 2 introduces our model and establishes preliminary results. Sections 3, 4 and 5 solve the proposed mechanism design problem for support-only ambiguity sets, Markov ambiguity sets, and Markov ambiguity sets with independent types, respectively. Conclusions are drawn in Section 6, and all proofs are relegated to the appendix.

Notation. For any $\boldsymbol{t} \in \mathbb{R}^I$, we denote by t_i the i^{th} component and by $\boldsymbol{t}_{-i}$ the subvector of $\boldsymbol{t}$ without t_i. The indicator function of a logical expression E is defined as $\mathbb{1}_E = 1$ if E is true and as $\mathbb{1}_E = 0$ otherwise. For any Borel sets $\mathcal{S} \subseteq \mathbb{R}^n$ and $\mathcal{D} \subseteq \mathbb{R}^m$, we use $\mathcal{P}_0(\mathcal{S})$ and $\mathcal{L}(\mathcal{S}, \mathcal{D})$ to denote the family of all probability distributions on $\mathcal{S}$ and the set of all bounded Borel-measurable functions from $\mathcal{S}$ to $\mathcal{D}$, respectively. Random variables are designated by symbols with tildes (*e.g.*, $\tilde{\boldsymbol{t}}$), and their realizations are denoted by the same symbols without tildes (*e.g.*, $\boldsymbol{t}$).

2. Problem Statement and Preliminaries

A principal aims to allocate a single good to one of $I \geq 2$ agents. Each agent $i \in \mathcal{I} = \{1, 2, \ldots, I\}$ derives a strictly positive deterministic benefit from receiving the good and uses it to generate a value $t_i \in \mathcal{T}_i = [\underline{t}_i, \bar{t}_i]$ for the principal, where $0 \leq \underline{t}_i < \bar{t}_i < \infty$. We henceforth refer to t_i as agent i's type, and we assume that t_i is privately known to agent i but unknown to the principal and the other agents. Thus, the principal perceives the vector $\tilde{\boldsymbol{t}} = (\tilde{t}_1, \tilde{t}_2 \ldots, \tilde{t}_I)$ of all agents' types as a random vector governed by some probability distribution $\mathbb{P}_0$ on the type space $\mathcal{T} = \prod_{i \in \mathcal{I}} \mathcal{T}_i$. However, the principal can inspect agent i's type at cost $c_i > 0$, and the inspection perfectly reveals t_i. In contrast to much of the existing literature on mechanism design, we assume here that neither the principal nor the agents know $\mathbb{P}_0$. Instead, they are only aware that $\mathbb{P}_0$ belongs to some commonly known ambiguity set $\mathcal{P} \subseteq \mathcal{P}_0(\mathcal{T})$. On this basis, the principal aims to design a mechanism for allocating the good. A mechanism is an extensive-form game between the principal and the agents, where the principal commits in advance to her strategy (for a formal definition of extensive-form games, see, *e.g.*, [12]). Such a mechanism may involve multiple stages of cheap talk statements by the agents, while the principal's actions include the decisions on whether to inspect certain agents and how to allocate the good. Monetary transfers are not allowed, *i.e.*, the agents and the principal cannot exchange money at any time.

Given any mechanism represented as an extensive form game, we denote by $\mathcal{H}_i$ the family of all information sets of agent i and by $\mathcal{A}(h_i)$ the actions available to agent i at the nodes in the information set $h_i \in \mathcal{H}_i$. All agents select their actions strategically in view of their individual preferences and the available information. In particular, agent i's actions depend on his type t_i. Thus, we model any (mixed) strategy of agent i as a function $s_i \in \mathcal{L}(\mathcal{T}_i, \prod_{h_i \in \mathcal{H}_i} \mathcal{P}_0(\mathcal{A}(h_i)))$ that maps each of his possible types to a complete contingency plan $a_i \in \prod_{h_i \in \mathcal{H}_i} \mathcal{P}_0(\mathcal{A}(h_i))$, which represents a probability distribution over the actions available to agent i for all information sets $h_i \in \mathcal{H}_i$. In the following, we denote by $\text{prob}_i(a_i; \boldsymbol{t}, \boldsymbol{a}_{-i})$ the probability that agent $i \in \mathcal{I}$ receives the good under the principal's mechanism if the agents have types $\boldsymbol{t}$ and play the contingency plans $\boldsymbol{a} = (a_1, a_2, \ldots, a_I)$. We also restrict attention to mechanisms that admit an ex-post Nash equilibrium.

Definition 1 (Ex-Post Nash Equilibrium). An I-tuple $\boldsymbol{s} = (s_1, s_2, \ldots, s_I)$ of mixed strategies $s_i \in \mathcal{L}(\mathcal{T}_i, \prod_{h_i \in \mathcal{H}_i} \mathcal{P}_0(\mathcal{A}(h_i)))$, $i \in \mathcal{I}$, is called an *ex-post Nash equilibrium* if

$$\text{prob}_i(s_i(t_i); \boldsymbol{t}, \boldsymbol{s}_{-i}(\boldsymbol{t}_{-i})) \geq \text{prob}_i(a_i; \boldsymbol{t}, \boldsymbol{s}_{-i}(\boldsymbol{t}_{-i})) \quad \forall i \in \mathcal{I}, \forall \boldsymbol{t} \in \mathcal{T}, \forall a_i \in \textstyle\prod_{h_i \in \mathcal{H}_i} \mathcal{P}_0(\mathcal{A}(h_i)).$$

Recall that all agents assign a strictly positive deterministic value to the good, and therefore the expected utility of agent i conditional on $\tilde{\boldsymbol{t}} = \boldsymbol{t}$ is proportional to $\text{prob}_i(a_i; \boldsymbol{t}, \boldsymbol{a}_{-i})$. Under an ex-post Nash equilibrium, each agent i maximizes this probability simultaneously for all type scenarios $\boldsymbol{t} \in \mathcal{T}$. Hence, it is clear that insisting on the existence of an ex-post Nash equilibrium restricts the family of mechanisms to be considered. Note that Ben-Porath et al. [5] study the larger class of mechanisms that admit a Bayesian Nash equilibrium. However, these mechanisms generically depend on the type distribution $\mathbb{P}_0$ and can therefore not be implemented by a principle who lacks knowledge of $\mathbb{P}_0$. It is therefore natural to restrict attention to mechanisms that admit ex-post Nash equilibria, which remain well-defined in the face of distributional ambiguity. We further assume from now on that the principal is ambiguity averse in the sense that she wishes to maximize her worst-case expected payoff in view of all distributions in the ambiguity set $\mathcal{P}$.

The class of all mechanisms that admit an ex-post Nash equilibrium is vast. An important subclass is the family of all truthful direct mechanisms. A direct mechanism $(\boldsymbol{p}, \boldsymbol{q})$ consists of two I-tuples $\boldsymbol{p} = (p_1, p_2, \ldots, p_I)$ and $\boldsymbol{q} = (q_1, q_2, \ldots, q_I)$ of allocation functions $p_i, q_i \in \mathcal{L}(\mathcal{T}, [0,1])$, $i \in \mathcal{I}$. Any direct mechanism $(\boldsymbol{p}, \boldsymbol{q})$ is implemented as follows. First, the principal announces $\boldsymbol{p}$ and $\boldsymbol{q}$, and then she collects a bid $t'_i \in \mathcal{T}_i$ from each agent $i \in \mathcal{I}$. Next, the principal implements randomized allocation and inspection decisions. Specifically, $p_i(\boldsymbol{t}')$ represents the total probability that agent i receives the good, while $q_i(\boldsymbol{t}')$ represents the probability that agent i receives the good *and* is inspected. If the inspection reveals that agent i has misreported his type, the principal penalizes the agent by repossessing the good. Any direct mechanism $(\boldsymbol{p}, \boldsymbol{q})$ must satisfy the feasibility conditions

$$q_i(\boldsymbol{t}') \le p_i(\boldsymbol{t}') \;\; \forall i \in \mathcal{I} \quad \text{and} \quad \sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}') \le 1 \;\; \forall \boldsymbol{t}' \in \mathcal{T}. \tag{FC}$$

The first inequality in (FC) holds because only agents who receive the good may undergo an inspection. The second inequality in (FC) ensures that the principal allocates the good at most once.

A direct mechanism $(\boldsymbol{p}, \boldsymbol{q})$ is called truthful if it is optimal for each agent i to report his true type $t'_i = t_i$. Thus, $(\boldsymbol{p}, \boldsymbol{q})$ is truthful if and only if it satisfies the incentive compatibility constraints

$$p_i(\boldsymbol{t}) \ge p_i(t'_i, \boldsymbol{t}_{-i}) - q_i(t'_i, \boldsymbol{t}_{-i}) \quad \forall i \in \mathcal{I}, \, \forall t'_i \in \mathcal{T}_i, \, \forall \boldsymbol{t} \in \mathcal{T}, \tag{IC}$$

which ensure that if all other agents report their true types $\boldsymbol{t}_{-i}$, then the probability $p_i(\boldsymbol{t})$ of agent i receiving the good if he reports his true type t_i exceeds the probability $p_i(t'_i, \boldsymbol{t}_{-i}) - q_i(t'_i, \boldsymbol{t}_{-i})$ of agent i receiving the good if he misreports his type as $t'_i \neq t_i$. By leveraging a variant of the Revelation Principle detailed in [5], one can show that for any mechanism that admits an ex-post Nash equilibrium, there exists an equivalent truthful direct mechanism that duplicates or improves the principal's worst-case expected payoff; see the online appendix of [5] for details. Without loss of generality, the principal may thus focus on truthful direct mechanisms, which greatly simplifies the problem of finding an optimal mechanism. Consequently, the principal's mechanism design problem can be formalized as the following distributionally robust optimization problem.

$$\begin{array}{rll} z^\star = & \sup\limits_{\boldsymbol{p}, \boldsymbol{q}} & \inf\limits_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\sum\limits_{i \in \mathcal{I}} (p_i(\tilde{\boldsymbol{t}}) \tilde{t}_i - q_i(\tilde{\boldsymbol{t}}) c_i) \right] \\ & \text{s.t.} & p_i, q_i \in \mathcal{L}(\mathcal{T}, [0,1]) \;\; \forall i \in \mathcal{I} \\ & & \text{(IC), (FC)} \end{array} \tag{MDP}$$

From now on, we will use the shorthand $\mathcal{X}$ to denote the set of all $(\boldsymbol{p}, \boldsymbol{q})$ feasible in (MDP). Note that the feasible set $\mathcal{X}$ does not depend on the ambiguity set $\mathcal{P}$.

In the remainder, we will demonstrate that (MDP) often admits multiple optimal solutions. While different optimal mechanisms generate the same expected profit in the worst case, they may offer dramatically different expected profits under generic non-worst-case distributions. This observation prompts us to seek mechanisms that are not only worst-case optimal but perform also well under *all* type distributions in the ambiguity set $\mathcal{P}$. More precisely, we hope to identify a worst-case optimal mechanism for which there exists no other feasible mechanism that generates a non-inferior expected payoff under *every* distribution in $\mathcal{P}$ and a higher expected payoff under *at least one* distribution in $\mathcal{P}$. A mechanism with this property is called *Pareto robustly optimal*. This terminology is borrowed from the theory of Pareto efficiency in classical robust optimization [14].

Definition 2 (Pareto Robust Optimality). We say that a mechanism $(\boldsymbol{p}', \boldsymbol{q}')$ that is feasible in (MDP) *weakly Pareto robustly dominates* another feasible mechanism $(\boldsymbol{p}, \boldsymbol{q})$ if

$$\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i'(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i'(\tilde{\boldsymbol{t}})c_i)\right] \geq \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\right] \quad \forall \mathbb{P}\in\mathcal{P}. \tag{1}$$

If the inequality (1) holds for all $\mathbb{P}\in\mathcal{P}$ and is strict for at least one $\mathbb{P}\in\mathcal{P}$, we say that $(\boldsymbol{p}', \boldsymbol{q}')$ Pareto robustly dominates $(\boldsymbol{p}, \boldsymbol{q})$. A mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that is optimal in (MDP) is called Pareto robustly optimal if there exists no other feasible mechanism $(\boldsymbol{p}', \boldsymbol{q}')$ that Pareto robustly dominates $(\boldsymbol{p}, \boldsymbol{q})$.

Note that any mechanism that weakly Pareto robustly dominates an optimal mechanism is also optimal in (MDP). Moreover, a Pareto robustly optimal mechanism typically exists. However, there may not exist any mechanism that Pareto robustly dominates all other feasible mechanisms.

We now define the notion of a favored-agent mechanism, which was first introduced in [5].

Definition 3 (Favored-Agent Mechanism). A mechanism $(\boldsymbol{p}, \boldsymbol{q})$ is a *favored-agent mechanism* if there is a favored agent $i^\star\in\mathcal{I}$ and a threshold value $\nu^\star\in\mathbb{R}$ such that the following hold.

(i) If $\max_{i\neq i^\star} t_i - c_i < \nu^\star$, then $p_{i^\star}(\boldsymbol{t}) = 1$, $q_{i^\star}(\boldsymbol{t}) = 0$ and $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t}) = 0$ for all $i\neq i^\star$.

(ii) If $\max_{i\neq i^\star} t_i - c_i > \nu^\star$, then $p_{i'}(\boldsymbol{t}) = q_{i'}(\boldsymbol{t}) = 1$ for some $i'\in\arg\max_{i\in\mathcal{I}}(t_i - c_i)$ and $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t}) = 0$ for all $i\neq i'$.

If $\max_{i\neq i^\star} t_i - c_i = \nu^\star$, then we are free to define $(\boldsymbol{p}(\boldsymbol{t}), \boldsymbol{q}(\boldsymbol{t}))$ either as in (i) or as in (ii).

Intuitively, if t_i is smaller than the adjusted cost of inspection $c_i + \nu^\star$ for every agent $i\neq i^\star$, then we are in case (i), and the favored-agent mechanism allocates the good to the favored agent $i^\star$ without inspection. If there exists an agent $i\neq i^\star$ whose type t_i exceeds the adjusted cost of inspection $c_i+\nu^\star$, then we are in case (ii), and the favored-agent mechanism allocates the good to an agent i' with highest net payoff $t_{i'} - c_{i'}$, and this agent is inspected. Note that in case (ii) the good can also be allocated to the favored agent.

A favored-agent mechanism is uniquely determined by a favored agent $i^\star$, a threshold value $\nu^\star$, and two tie-breaking rules. The first tie-breaking rule determines the winning agent in case (ii) when $\arg\max_{i\in\mathcal{I}}(t_i - c_i)$ is not a singleton. From now on we will always use the lexicographic tie-breaking rule in this case, which sets $i' = \min\arg\max_{i\in\mathcal{I}}(t_i - c_i)$. The second tie-breaking rule determines whether $(\boldsymbol{p}(\boldsymbol{t}), \boldsymbol{q}(\boldsymbol{t}))$ should be constructed as in case (i) or as in case (ii) when $\max_{i\neq i^\star} t_i - c_i = \nu^\star$. From now on we say that a favored-agent mechanism is of type (i) if $(\boldsymbol{p}(\boldsymbol{t}), \boldsymbol{q}(\boldsymbol{t}))$ is always defined as in (i) and that it is of type (ii) if $(\boldsymbol{p}(\boldsymbol{t}), \boldsymbol{q}(\boldsymbol{t}))$ is always defined as in (ii) in case of a tie. Note that both tie-breaking rules are irrelevant in the Bayesian setting considered in [5], but they are relevant for us because the ambiguity sets $\mathcal{P}$ to be studied below contain discrete distributions, under which ties have a strictly positive probability.

All favored-agent mechanisms are feasible in (MDP), see Remark 1 in [5]. In particular, they are incentive compatible, that is, the agents have no incentive to misreport their types. To see this, recall that under a favored-agent mechanism the winning agent receives the good with probability one, and the losing agents receive the good with probability zero. Thus, if an agent wins by truthful bidding, he cannot increase his chances of receiving the good by lying about his type. If an agent loses by truthful bidding, on the other hand, he has certainly no incentive to lower his bid t_i because the chances of receiving the good are non-decreasing in t_i. Increasing his bid t_i may earn him the good provided that $t_i - c_i$ attains the maximum of $t_{i'} - c_{i'}$ over $i'\in\mathcal{I}$. However, in this case, the agent's type is inspected with probability one. Hence, the lie will be detected and the good will be repossessed. This shows that no agent benefits from lying under a favored-agent mechanism.

If $\mathcal{P} = \{\mathbb{P}_0\}$ is a singleton, the agents' types are independent under $\mathbb{P}_0$, and $\mathbb{P}_0$ has an everywhere positive density on $\mathcal{T}$, then problem (MDP) is solved by a favored-agent mechanism [5, Theorem 1]. The favored-agent mechanism with favored agent i and threshold ν_i generates an expected payoff of

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}_0}\left[\tilde{t}_i \mathbb{1}_{\tilde{y}_i \le \nu_i} + \max\left\{\tilde{t}_i - c_i, \tilde{y}_i\right\} \mathbb{1}_{\tilde{y}_i \ge \nu_i}\right] \\ &= \int_{-\infty}^{\nu_i} \mathbb{E}_{\mathbb{P}_0}\left[\tilde{t}_i\right] \rho_i(y_i) \mathrm{d}y_i + \int_{\nu_i}^{\infty} \mathbb{E}_{\mathbb{P}_0}\left[\max\left\{\tilde{t}_i - c_i, y_i\right\}\right] \rho_i(y_i) \mathrm{d}y_i,\end{aligned}$$

where the random variable $\tilde{y}_i = \max_{j\neq i} \tilde{t}_j - c_j$ with probability density function $\rho_i(y_i)$ is independent of $\tilde{t}_i$ under $\mathbb{P}_0$. The threshold value $\nu_i^\star$ that maximizes this expression thus solves the first-order optimality condition

$$\mathbb{E}_{\mathbb{P}_0}\left[\tilde{t}_i\right] = \mathbb{E}_{\mathbb{P}_0}\left[\max\left\{\tilde{t}_i - c_i, \nu_i\right\}\right]. \tag{2}$$

Note that $\nu_i^\star$ is unique because the right-hand side of (2) strictly increases in ν_i on the domain of interest; see [5, Theorem 2] for additional details. One can further prove that within the finite class of favored-agent mechanisms with optimal thresholds, the ones with the highest threshold are optimal. More specifically, any favored-agent mechanism with favored agent $i^\star \in \arg\max_{i\in\mathcal{I}} \nu_i^\star$ and threshold $\nu^\star = \max_{i\in\mathcal{I}} \nu_i^\star$ is optimal within the class of favored-agent mechanisms [5, Theorem 3]. Hence, any such mechanism must be optimal in (MDP). Finally, one can also show that for mutually distinct cost coefficients c_i, $i \in \mathcal{I}$, the optimal favored-agent mechanism is unique.

In the remainder of the paper, we will address instances of the mechanism design problem (MDP) where $\mathcal{P}$ is *not* a singleton, and we will prove that favored-agent mechanisms remain optimal. Under distributional ambiguity, however, the construction of $i^\star$ and $\nu^\star$ described above is no longer well-defined because it depends on a particular choice of the probability distribution of $\tilde{\boldsymbol{t}}$. We will show that if $\mathcal{P}$ is not a singleton, then there may be infinitely many optimal favored-agent mechanisms with different thresholds $\nu^\star$. In this situation, it is expedient to look for Pareto robustly optimal favored-agent mechanisms.

3. Support-Only Ambiguity Sets

We now investigate the mechanism design problem (MDP) under the assumption that $\mathcal{P} = \mathcal{P}_0(\mathcal{T})$ is the support-only ambiguity set that contains all distributions supported on the type space $\mathcal{T}$. As $\mathcal{P}$ contains all Dirac point distributions concentrating unit mass at any $\boldsymbol{t} \in \mathcal{T}$, the worst-case expected payoff over all distributions $\mathbb{P} \in \mathcal{P}$ simplifies to the worst-case payoff overall type profiles $\boldsymbol{t} \in \mathcal{T}$, and thus it is easy to verify that problem (MDP) simplifies to

$$\begin{aligned} z^\star = \ \sup_{\boldsymbol{p},\boldsymbol{q}} \quad & \inf_{\boldsymbol{t}\in\mathcal{T}} \sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \\ \text{s.t.} \quad & p_i, q_i \in \mathcal{L}(\mathcal{T}, [0,1]) \ \forall i \in \mathcal{I} \\ & (\text{IC}),\ (\text{FC}). \end{aligned} \tag{3}$$

Similarly, it is easy to verify that an optimal mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ for problem (3) is Pareto robustly optimal if there exists no other feasible mechanism $(\boldsymbol{p}, \boldsymbol{q})$ with

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \ge \sum_{i\in\mathcal{I}} (p_i^\star(\boldsymbol{t}) t_i - q_i^\star(\boldsymbol{t}) c_i) \quad \forall \boldsymbol{t} \in \mathcal{T},$$

where the inequality is strict for at least one type profile $\boldsymbol{t} \in \mathcal{T}$. If the principal knew the agents' types ex ante, she could simply allocate the good to the agent with the highest type and would not have to spend money on inspecting anyone. One can therefore show that the optimal value $z^\star$ of problem (3) is upper bounded by $\inf_{\boldsymbol{t}\in\mathcal{T}} \max_{i\in\mathcal{I}} t_i = \max_{i\in\mathcal{I}} \underline{t}_i$. The following proposition reveals that this upper bound is attained by an admissible mechanism.

Proposition 1. *Problem (3) is solvable, and its optimal value is given by* $z^\star = \max_{i\in\mathcal{I}} \underline{t}_i$.

The next theorem shows that there are infinitely many optimal favored-agent mechanisms that attain the optimal value $z^\star = \max_{i\in\mathcal{I}} \underline{t}_i$ of problem (3).

Theorem 1. *Any favored-agent mechanism with favored agent* $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{t}_i$ *and threshold value* $\nu^\star \geq \max_{i\in\mathcal{I}} \underline{t}_i$ *is optimal in problem (3).*

Remark 1. Theorem 1 is sharp in the sense that there are problem instances for which any favored-agent mechanism with favored agent $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{t}_i$ and threshold value $\nu < \max_{i\in\mathcal{I}} \underline{t}_i$ is strictly suboptimal in (3). To see this, consider an instance with $I = 2$ agents, where $\mathcal{T}_1 = [2, 8]$, $\mathcal{T}_2 = [0, 10]$ and $c_1 = c_2 = 1$. By Proposition 1, the supremum of (3) is given by $\max_{i\in\mathcal{I}} \underline{t}_i = 2$. Consider now any favored agent mechanism with favored agent $1 \in \arg\max_{i\in\mathcal{I}} \underline{t}_i$ and threshold value $\nu < \underline{t}_1 = 2$. This mechanism is strictly suboptimal. To see this, assume first that $\nu < 1$. If $\boldsymbol{t} = (2, 2)$, then the mechanism allocates the good to agent 1 or agent 2 with verification and earns $t_1 - c_1 = t_2 - c_2 = 1$. Thus, the worst-case payoff overall $\boldsymbol{t} \in \mathcal{T}$ cannot exceed 1, which is strictly smaller than the optimal worst-case payoff. Assume next that $\nu \in [1, 2)$. If $\boldsymbol{t} = (2, 2 + \nu/2) \in \mathcal{T}$, then the mechanism allocates the good to agent 2 with verification and earns $1 + \nu/2$. Thus, the worst-case payoff overall $\boldsymbol{t} \in \mathcal{T}$ cannot exceed $1 + \nu/2$, which is strictly smaller than the optimal worst-case payoff. In summary, the mechanism is strictly suboptimal for all $\nu < 2$. □

As the mechanism design problem (3) constitutes a convex program, any convex combination of optimal favored-agent mechanisms gives rise to yet another optimal mechanism. However, problem (3) also admits optimal mechanisms that can neither be interpreted as favored-agent mechanisms nor as convex combinations of favored-agent mechanisms. To see this, consider any favored-agent mechanism $(\boldsymbol{p}, \boldsymbol{q})$ with favored agent $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{t}_i$ and threshold value $\nu^\star \in \mathbb{R}$ satisfying $\nu^\star \geq \max_{i\in\mathcal{I}} \underline{t}_i$ and $\nu^\star > \max_{i\in\mathcal{I}} \bar{t}_i - c_i$. By Theorem 1, this mechanism is optimal in problem (3). The second condition on $\nu^\star$ implies that this mechanism allocates the good to the favored agent without inspection for every $\boldsymbol{t} \in \mathcal{T}$ (case (i) always prevails). Next, construct $\hat{\boldsymbol{t}} \in \mathcal{T}$ through $\hat{t}_i = \underline{t}_i$ for all $i \neq i^\star$ and $\hat{t}_{i^\star} = \bar{t}_{i^\star}$, and note that $\hat{\boldsymbol{t}} \neq \underline{\boldsymbol{t}}$ because $\underline{t}_{i^\star} < \bar{t}_{i^\star}$. Finally, introduce another mechanism $(\boldsymbol{p}, \boldsymbol{q}')$, where $\boldsymbol{q}'$ is defined through $q'_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{T}$ and $i \neq i^\star$ and

$$q'_{i^\star}(\boldsymbol{t}) = \begin{cases} \min\{1, (\bar{t}_{i^\star} - \underline{t}_{i^\star})/c_{i^\star}\} & \text{if } \boldsymbol{t} = \hat{\boldsymbol{t}}, \\ q_{i^\star}(\boldsymbol{t}) & \text{if } \boldsymbol{t} \in \mathcal{T} \setminus \{\hat{\boldsymbol{t}}\}. \end{cases}$$

One readily verifies that $(\boldsymbol{p}, \boldsymbol{q}')$ is feasible in (3). Indeed, as $(\boldsymbol{p}, \boldsymbol{q}')$ differs from $(\boldsymbol{p}, \boldsymbol{q})$ only in scenario $\hat{\boldsymbol{t}}$, and as $(\boldsymbol{p}, \boldsymbol{q})$ is feasible, it suffices to check the feasibility of $(\boldsymbol{p}, \boldsymbol{q}')$ in scenario $\hat{\boldsymbol{t}}$. Indeed, the modified allocation rule $\boldsymbol{q}'$ is valued in $[0, 1]^I$, and $(\boldsymbol{p}, \boldsymbol{q}')$ satisfies (FC) because $0 \leq q'_{i^\star}(\hat{\boldsymbol{t}}) \leq 1 = p_{i^\star}(\hat{\boldsymbol{t}})$, where the equality holds because the favored-agent mechanism $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent $i^\star$ with certainty. Similarly, the modified mechanism $(\boldsymbol{p}, \boldsymbol{q}')$ satisfies (IC) because

$$p_{i^\star}(t_{i^\star}, \hat{\boldsymbol{t}}_{-i^\star}) = 1 \geq p_{i^\star}(\hat{\boldsymbol{t}}) - q'_{i^\star}(\hat{\boldsymbol{t}}) \quad \forall t_{i^\star} \in \mathcal{T}_{i^\star}.$$

In summary, we have thus shown that the mechanism $(\boldsymbol{p}, \boldsymbol{q}')$ is feasible in (3). To show that it is also optimal, recall that $(\boldsymbol{p}, \boldsymbol{q})$ is optimal with worst-case payoff $\max_{i\in\mathcal{I}} \underline{t}_i$ and that $(\boldsymbol{p}, \boldsymbol{q}')$ differs from $(\boldsymbol{p}, \boldsymbol{q})$ only in scenario $\hat{\boldsymbol{t}}$. The principal's payoff in scenario $\hat{\boldsymbol{t}}$ amounts to

$$p_{i^\star}(\hat{\boldsymbol{t}})\hat{t}_{i^\star} - q'_{i^\star}(\hat{\boldsymbol{t}})c_{i^\star} = \hat{t}_{i^\star} - q'_{i^\star}(\hat{\boldsymbol{t}})c_{i^\star} \geq \hat{t}_{i^\star} - \frac{\hat{t}_{i^\star} - \underline{t}_{i^\star}}{c_{i^\star}} c_{i^\star} = \underline{t}_{i^\star} = \max_{i\in\mathcal{I}} \underline{t}_i,$$

where the inequality follows from the definition of $q'_{i^\star}(\hat{\boldsymbol{t}})$. Thus, the worst-case payoff of $(\boldsymbol{p}, \boldsymbol{q}')$ amounts to $\max_{i\in\mathcal{I}} \underline{t}_i$, and $(\boldsymbol{p}, \boldsymbol{q}')$ is indeed optimal in (3). However, $(\boldsymbol{p}, \boldsymbol{q}')$ is *not* a favored-agent mechanism for otherwise $q'_{i^\star}(\hat{\boldsymbol{t}})$ would have to vanish; see Definition 3. In addition, note that $p_{i^\star}(\hat{\boldsymbol{t}}) - q'_{i^\star}(\hat{\boldsymbol{t}}) < 1$ whereas $p_{i^\star}(t_{i^\star}, \hat{\boldsymbol{t}}_{-i^\star}) - q'_{i^\star}(t_{i^\star}, \hat{\boldsymbol{t}}_{-i^\star}) = 1$ for all $t_{i^\star} \neq \hat{t}_{i^\star}$. This implies via Lemma 1 below that $(\boldsymbol{p}, \boldsymbol{q}')$ is also *not* a convex combination of favored-agent mechanisms.

Lemma 1. *If a mechanism $(\boldsymbol{p},\boldsymbol{q})$ is a convex combination of favored-agent mechanisms, then the function $p_i(t_i,\boldsymbol{t}_{-i}) - q_i(t_i,\boldsymbol{t}_{-i})$ is constant in $t_i \in \mathcal{T}_i$ for any fixed $i \in \mathcal{I}$ and $\boldsymbol{t}_{-i} \in \mathcal{T}_{-i}$.*

In summary, we have shown that the robust mechanism design problem (3) admits infinitely many optimal solutions. Some of these solutions represent favored-agent mechanisms, while others represent convex combinations of favored-agent mechanisms. However, some optimal mechanisms are neither crisp favored-agent mechanisms nor convex combinations of favored-agent mechanisms. While all robustly optimal mechanisms generate the same payoff in the worst case, however, their payoffs may differ significantly in non-worst-case scenarios. This observation suggests that robust optimality alone is not a sufficient differentiator to distinguish desirable from undesirable mechanisms. Note also that the optimal mechanism constructed above by altering the inspection probabilities of an optimal favored-agent mechanism is Pareto robustly dominated by its underlying favored-agent mechanism. This observation prompts us to seek Pareto robustly optimal mechanisms for problem (3). The next theorem shows that among all robustly optimal favored-agent mechanisms identified in Theorem 1 there is always one that is also Pareto robustly optimal.

Theorem 2. *Any favored-agent mechanism of type (i) with favored agent $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{t}_i$ and threshold value $\nu^\star = \max_{i\in\mathcal{I}} \underline{t}_i$ is Pareto robustly optimal in problem (3).*

We sketch the proof idea in the special case when there are only two agents. To convey the key ideas without tedious case distinctions, we assume that $\underline{t}_1 > \underline{t}_2$ so that $\arg\max_{i\in\mathcal{I}} \underline{t}_i = \{1\}$ is a singleton, and we assume that $\bar{t}_2 > c_2 + \underline{t}_1$ and $\bar{t}_1 > c_2 + \underline{t}_1$. Throughout the subsequent discussion, we will use the following partition of the type space $\mathcal{T}$.

$$
\begin{aligned}
\mathcal{T}_I &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 \le \underline{t}_1 \text{ and } t_2 < t_1\} \\
\mathcal{T}_{II} &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 \le \underline{t}_1 \text{ and } t_2 \ge t_1\} \\
\mathcal{T}_{III} &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 > \underline{t}_1 \text{ and } t_2 - c_2 > t_1\} \\
\mathcal{T}_{IV} &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 > \underline{t}_1 \text{ and } t_2 - c_2 \le t_1\}
\end{aligned}
\tag{4}
$$

The sets $\mathcal{T}_I$, $\mathcal{T}_{II}$, $\mathcal{T}_{III}$ and $\mathcal{T}_{IV}$ are visualized in Figure 1. Note that all of them are nonempty thanks to the above assumptions about $\underline{t}_1$, $\underline{t}_2$ and c_2. We emphasize, however, that all simplifying assumptions as well as the restriction to two agents are relaxed in the formal proof of Theorem 2.

In the following we denote by $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ the favored-agent mechanism of type (i) with favored agent 1 and threshold value $\nu^\star = \underline{t}_1$, and we will prove that this mechanism is Pareto robustly optimal in problem (3). To this end, assume for the sake of contradiction that there exists another mechanism $(\boldsymbol{p},\boldsymbol{q})$ feasible in (3) that Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. Thus, we have

$$
\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \ge \sum_{i\in\mathcal{I}} (p_i^\star(\boldsymbol{t}) t_i - q_i^\star(\boldsymbol{t}) c_i) \quad \forall \boldsymbol{t} \in \mathcal{T}, \tag{5}
$$

where the inequality is strict for at least one $\boldsymbol{t} \in \mathcal{T}$. The right hand side of (5) represents the principal's payoff in scenario $\boldsymbol{t}$ under $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. By the definition of a type (i) favored-agent mechanism, this payoff amounts to t_1 when $t_2 - c_2 \le \underline{t}_1$ (*i.e.*, when $\boldsymbol{t} \in \mathcal{T}_I \cup \mathcal{T}_{II}$) and to $\max_{i\in\mathcal{I}} t_i - c_i$ when $t_2 - c_2 > \underline{t}_1$ (*i.e.*, when $\boldsymbol{t} \in \mathcal{T}_{III} \cup \mathcal{T}_{IV}$). We will now use what we call *spatial induction* to show that if (5) holds, then $(\boldsymbol{p},\boldsymbol{q})$ must generate the same payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ under every type profile $\boldsymbol{t} \in \mathcal{T}$. In other words, $(\boldsymbol{p},\boldsymbol{q})$ cannot generate a strictly higher payoff than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ under any type profile, which contradicts our assumption that $(\boldsymbol{p},\boldsymbol{q})$ Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$.

The idea of spatial induction can be explained as follows. In the base step, we first prove the equivalence of $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ on $\mathcal{T}_I$ by using elementary arguments. In the first induction step, we then prove the equivalence of these two mechanisms on $\mathcal{T}_{II}$ by exploiting the non-locality of the incentive compatibility constraint (IC), which couples the allocation probabilities of any feasible mechanism across different type profiles. This coupling allows us to prove the equivalence of $(\boldsymbol{p},\boldsymbol{q})$

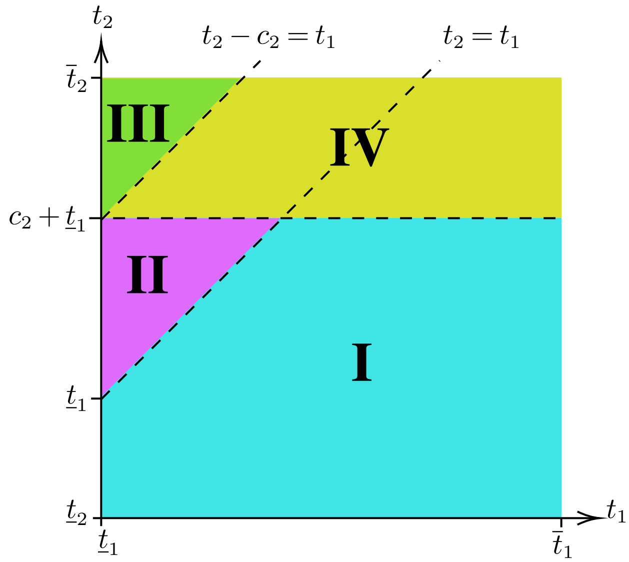

Figure 1 Partition (4) of the type space $\mathcal{T}$.

and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ on $\mathcal{T}_{II}$ from their equivalence on $\mathcal{T}_I$, which was established in the base step. In the remaining induction steps we iterate through the subsets $\mathcal{T}_{III}$ and $\mathcal{T}_{IV}$ one by one and use again the non-locality of the incentive compatibility constraint to prove the equivalence of $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ on the current subset from their equivalence on the subsets visited previously.

As for the base step, consider first a type profile $\boldsymbol{t} \in \mathcal{T}_I$. For inequality (5) to hold in this scenario, the principal must earn at least t_1 under the mechanism $(\boldsymbol{p}, \boldsymbol{q})$. As $t_2 < t_1$, $c_i > 0$ and $(\boldsymbol{p}, \boldsymbol{q})$ satisfies the (FC) constraints $\sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1$ and $q_i(\boldsymbol{t}) \geq 0$, this is only possible if $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$. Thus, the allocation probabilities of the mechanisms $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ coincide on $\mathcal{T}_I$.

As for the first induction step, consider now any $\boldsymbol{t} \in \mathcal{T}_{II}$. For inequality (5) to hold in scenario $\boldsymbol{t}$, the principal must earn at least t_1 under the mechanism $(\boldsymbol{p}, \boldsymbol{q})$. Incentive compatibility ensures that $p_1(\boldsymbol{t}) \geq p_1(\bar{t}_1, t_2) - q_1(\bar{t}_1, t_2) = 1$, where the equality holds because $(\bar{t}_1, t_2) \in \mathcal{T}_I$ thanks to the assumption $\bar{t}_1 > c_2 + \underline{t}_1$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 without inspection in $\mathcal{T}_I$. Thus, the mechanism $(\boldsymbol{p}, \boldsymbol{q})$ can only earn t_1 in scenario $\boldsymbol{t}$ if $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$. In summary, the allocation probabilities of $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ must coincide on $\mathcal{T}_{II}$.

Next, as for the second induction step, consider any $\boldsymbol{t} \in \mathcal{T}_{III}$. Incentive compatibility ensures that $p_2(\boldsymbol{t}) - q_2(\boldsymbol{t}) \leq p_2(t_1, \underline{t}_2) = 0$, where the equality holds because $(t_1, \underline{t}_2) \in \mathcal{T}_I \cup \mathcal{T}_{II}$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 without inspection throughout $\mathcal{T}_I \cup \mathcal{T}_{II}$. As the allocation probabilities are non-negative and satisfy the (FC) condition $p_2(\boldsymbol{t}) \geq q_2(\boldsymbol{t})$, we may conclude that $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$. Thus, the type of agent 2 is inspected if he wins the good in scenario $\boldsymbol{t}$. As $t_2 - c_2 > t_1 > t_1 - c_1$ for all $\boldsymbol{t} \in \mathcal{T}_{III}$, the inequality (5) implies that the principal must earn at least $t_2 - c_2$ under the mechanism $(\boldsymbol{p}, \boldsymbol{q})$ in scenario $\boldsymbol{t}$. This is only possible if $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t}) = 1$. In summary, the allocation probabilities of $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ must therefore also coincide in $\mathcal{T}_{III}$.

In the last induction step, finally, consider any $\boldsymbol{t} \in \mathcal{T}_{IV}$. Incentive compatibility ensures that $0 = p_1(\underline{t}_1, t_2) \geq p_1(\boldsymbol{t}) - q_1(\boldsymbol{t})$, where the equality holds because $(\underline{t}_1, t_2) \in \mathcal{T}_{III}$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 2 in $\mathcal{T}_{III}$. Incentive compatibility also ensures that $0 = p_2(t_1, \underline{t}_2) \geq p_2(\boldsymbol{t}) - q_2(\boldsymbol{t})$, where the equality holds because $(t_1, \underline{t}_2) \in \mathcal{T}_I \cup \mathcal{T}_{II}$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 in $\mathcal{T}_I \cup \mathcal{T}_{II}$. We may thus conclude that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I} = \{1, 2\}$. For the inequality (5) to hold in scenario $\boldsymbol{t}$, the principal must earn at least $\max_{i \in \mathcal{I}} t_i - c_i$ under the mechanism $(\boldsymbol{p}, \boldsymbol{q})$. As $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}$, this is only possible if $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to an agent $i' \in \arg\max_{i \in \mathcal{I}} t_i - c_i$ and inspects this agent. Thus, the principal's payoff under $(\boldsymbol{p}, \boldsymbol{q})$ matches her payoff under $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in region $\mathcal{T}_{IV}$.

The above spatial induction argument shows that the principal's earnings coincide under $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ throughout the entire type space $\mathcal{T}$. Therefore, $(\boldsymbol{p},\boldsymbol{q})$ cannot Pareto robustly dominate $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$, which in turn proves that $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ is Pareto robustly optimal in problem (3).

4. Markov Ambiguity Sets

Although simple and adequate for situations in which there is no distributional information at all, support-only ambiguity sets may be perceived as conservative in practice. In the following, we thus investigate the mechanism design problem (MDP) under the assumption that distributional uncertainty is captured by a Markov ambiguity set of the form

$$\mathcal{P}=\{\mathbb{P}\in\mathcal{P}_0(\mathcal{T}):\mathbb{E}_{\mathbb{P}}[\tilde{t}_i]\in[\underline{\mu}_i,\overline{\mu}_i]\ \forall i\in\mathcal{I}\},\tag{6}$$

where $\underline{\mu}_i$ and $\overline{\mu}_i$ denote lower and upper bounds on the expected type $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i]$ of agent $i\in\mathcal{I}$. We assume without much loss of generality that $\underline{t}_i<\underline{\mu}_i<\overline{\mu}_i<\overline{t}_i$ for all $i\in\mathcal{I}$. Under a Markov ambiguity set, the principal has information about the support as well as the mean of the agents' types.

Recall that if the principal knew the agents' types ex ante, then she could simply allocate the good to the agent with the highest type without inspection. Therefore, the optimal value $z^\star$ of problem (MDP) cannot exceed $\inf_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}[\max_{i\in\mathcal{I}}\tilde{t}_i]$. The next proposition shows that if $\mathcal{P}$ is a Markov ambiguity set, then this upper bound coincides with $\max_{i\in\mathcal{I}}\underline{\mu}_i$ and is attained by an admissible mechanism.

Proposition 2. *If $\mathcal{P}$ is a Markov ambiguity set of the form (6), then problem (MDP) is solvable, and $z^\star=\max_{i\in\mathcal{I}}\underline{\mu}_i$.*

Contrasting Proposition 2 with Proposition 1 shows that the principal can increase her optimal worst-case expected payoff from $\max_{i\in\mathcal{I}}\underline{t}_i$ to $\max_{i\in\mathcal{I}}\underline{\mu}_i$ by acquiring information about the mean values of the agents' types. The next theorem characterizes a class of favored-agent mechanisms that attain the optimal value $z^\star=\max_{i\in\mathcal{I}}\underline{\mu}_i$ of problem (MDP) under a Markov ambiguity set.

Theorem 3. *If $\mathcal{P}$ is a Markov ambiguity set of the form (6), then any favored-agent mechanism with favored agent $i^\star\in\arg\max_{i\in\mathcal{I}}\underline{\mu}_i$ and threshold value $\nu^\star\geq\overline{t}_{i^\star}$ is optimal in (MDP).*

Remark 2. Theorem 3 is sharp in the sense that there are problem instances for which any favored-agent mechanism with favored agent $i^\star\in\arg\max_{i\in\mathcal{I}}\underline{\mu}_i$ and threshold value $\nu<\overline{t}_{i^\star}$ is strictly suboptimal. To see this, consider an instance of problem (MDP) with $I=2$ agents, where $\mathcal{T}_1=[1,6]$, $\mathcal{T}_2=[0,10]$, $[\underline{\mu}_1,\overline{\mu}_1]=[4,5]$, $[\underline{\mu}_2,\overline{\mu}_2]=[3,7]$ and $c_1=c_2=2$. By Proposition 2, the optimal value of problem (MDP) is thus given by $\max_{i\in\mathcal{I}}\underline{\mu}_i=4$. Consider now any favored agent mechanism with favored agent $1\in\arg\max_{i\in\mathcal{I}}\underline{\mu}_i$ and threshold value $\nu<\overline{t}_1=6$. In the following, we prove that this mechanism is suboptimal. To this end, assume first that $\nu<1$. If $\boldsymbol{t}=\underline{\boldsymbol{\mu}}=(4,3)$, then the mechanism allocates the good to agent 1 with verification and earns $t_1-c_1=2$. As the discrete distribution that assigns unit probability to the scenario $\underline{\boldsymbol{\mu}}$ belongs to the Markov ambiguity set (6), the worst-case expected payoff across all admissible distributions cannot exceed 2, which is strictly smaller than the optimal worst-case expected payoff. Assume next that $\nu\in[1,6)$, and consider the discrete distribution $\mathbb{P}$ that assigns probability $\frac{1}{2}$ to the scenarios $(6,6.5+\nu/4)$ and $(2,0)$ each. One readily verifies that $\mathbb{P}$ belongs to the Markov ambiguity set (6). In scenario $(6,6.5+\nu/4)$, the mechanism allocates the good to agent 2 with verification because $t_2-c_2=(6.5+\nu/4)-2>\nu$ and $t_2-c_2>4=t_1-c_1$. In scenario $(2,0)$, on the other hand, the mechanism allocates the good to agent 1 without verification. Thus, the expected payoff of the mechanism under the distribution $\mathbb{P}$ amounts to $\frac{1}{2}(4.5+\nu/4)+\frac{1}{2}2=3.25+\nu/8$, and the worst-case expected payoff over all admissible distributions cannot exceed $3.25+\nu/8$, which is strictly smaller than the optimal worst-case expected payoff. In summary, the mechanism is strictly suboptimal for all $\nu<6$. □

In the remainder of this section, we will show that among all optimal favored-agent mechanism identified in Theorem 3 there is always one that is Pareto robustly optimal; see Theorem 4 below. The proof of this main result requires four preliminary lemmas. We stress that, even though all of these lemmas rely on the assumption that the set $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ is a singleton (meaning that the favored agent is uniquely determined), Theorem 4 will *not* depend on this restrictive assumption.

We first show that any type profile $\boldsymbol{t}\in\mathcal{T}$ has a strictly positive probability under some two-point distribution in the Markov ambiguity set. This result will help us to prove the subsequent lemmas.

LEMMA 2. *If $\mathcal{P}$ is a Markov ambiguity set of the form (6) and $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, then, for any type profile $\boldsymbol{t}\in\mathcal{T}$ there exist a scenario $\hat{\boldsymbol{t}}\in\mathcal{T}$ with $\max_{i\neq i^\star}\hat{t}_i < \hat{t}_{i^\star}$ and a discrete distribution $\mathbb{P}\in\mathcal{P}$ such that (i) $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i]=\underline{\mu}_i$ for all $i\in\mathcal{I}$, (ii) $\mathbb{P}(\tilde{\boldsymbol{t}}\in\{\boldsymbol{t},\hat{\boldsymbol{t}}\})=1$, and (iii) $\mathbb{P}(\tilde{\boldsymbol{t}}=\boldsymbol{t})>0$.*

The next proposition leverages Lemma 2 to establish a necessary and sufficient optimality condition for distributionally robust mechanism design problems with a Markov ambiguity set.

LEMMA 3. *If $\mathcal{P}$ is a Markov ambiguity set of the form (6) and $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, then a mechanism $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$ is optimal in (MDP) if and only if*

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \geq t_{i^\star} \quad \forall \boldsymbol{t}\in\mathcal{T}. \tag{7}$$

Lemma 3 reveals that the type $t_{i^\star}$ of agent $i^\star$ is an important reference point for the principal's payoff in any scenario $\boldsymbol{t}\in\mathcal{T}$ under an optimal mechanism. The next result reveals that the optimality condition (7) and the incentive compatibility constraints uniquely determine the allocation probabilities of any optimal mechanism throughout a subset of all scenarios. Specifically, any optimal mechanism in (MDP) must allocate the good to agent $i^\star$ without inspection if no other agent's type $t_i - c_i$, adjusted by the cost of verification, exceeds agent $i^\star$'s highest possible type $\bar{t}_{i^\star}$. This lemma will be a crucial ingredient for constructing Pareto robustly optimal mechanisms.

LEMMA 4. *If $\mathcal{P}$ is a Markov ambiguity set of the form (6), $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, and the mechanism $(\boldsymbol{p},\boldsymbol{q})$ is optimal in (MDP), then we have $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$ for all type profiles $\boldsymbol{t}\in\mathcal{T}$ with $\max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star}$.*

We sketch the proof of Lemma 4 in the special case when there are only two agents with $\underline{\mu}_2 < \underline{\mu}_1$ such that $\arg\max_{i\in\mathcal{I}}\underline{\mu}_i = \{1\}$ is a singleton. In the subsequent discussion, we assume that $\bar{t}_2 > c_2 + \bar{t}_1$, and we use the following partition of the type space $\mathcal{T}$, which is visualized in Figure 2.

$$\begin{aligned}
\mathcal{T}_I &= \{\boldsymbol{t}\in\mathcal{T} : t_2 < t_1\}\\
\mathcal{T}_{II} &= \{\boldsymbol{t}\in\mathcal{T} : t_2 \geq t_1 \text{ and } t_2 < \bar{t}_1\}\\
\mathcal{T}_{III} &= \{\boldsymbol{t}\in\mathcal{T} : t_2 \geq t_1,\ t_2 \geq \bar{t}_1 \text{ and } t_2 - c_2 < t_1\}\\
\mathcal{T}_{IV} &= \{\boldsymbol{t}\in\mathcal{T} : t_2 \geq t_1,\ t_2 \geq \bar{t}_1,\ t_2 - c_2 \geq t_1 \text{ and } t_2 - c_2 < \bar{t}_1\}\\
\mathcal{T}_{V} &= \{\boldsymbol{t}\in\mathcal{T} : t_2 \geq t_1,\ t_2 \geq \bar{t}_1,\ t_2 - c_2 \geq t_1 \text{ and } t_2 - c_2 \geq \bar{t}_1\}
\end{aligned} \tag{8}$$

Note that some inequalities in the definitions of these sets are redundant and only included for better readability. Using our simplifying assumptions on $\underline{\mu}_1$, $\underline{\mu}_2$, $\bar{t}_1$, $\bar{t}_2$ and c_2, one can show that all of the above sets are nonempty. We emphasize, however, that these simplifying assumptions as well as the restriction to two agents are relaxed in the formal proof of Lemma 4.

Select now any optimal mechanism $(\boldsymbol{p},\boldsymbol{q})$. Lemma 4 asserts that this mechanism allocates the good to agent 1 without inspection in every scenario $\boldsymbol{t}\in\mathcal{T}$ with $t_2 - c_2 < \bar{t}_1$, that is, in every scenario $\boldsymbol{t}\in\mathcal{T}\setminus\mathcal{T}_V$. We will prove this assertion via a spatial induction argument based on the subsets $\mathcal{T}_I$, $\mathcal{T}_{II}$, $\mathcal{T}_{III}$ and $\mathcal{T}_{IV}$ of the type space, where the induction step will critically exploit the non-locality of the incentive compatibility constraint (IC). In addition, we will exploit Lemma 3, which ensures that the mechanism $(\boldsymbol{p},\boldsymbol{q})$ generates a payoff of at least t_1 in each scenario $\boldsymbol{t}\in\mathcal{T}$.

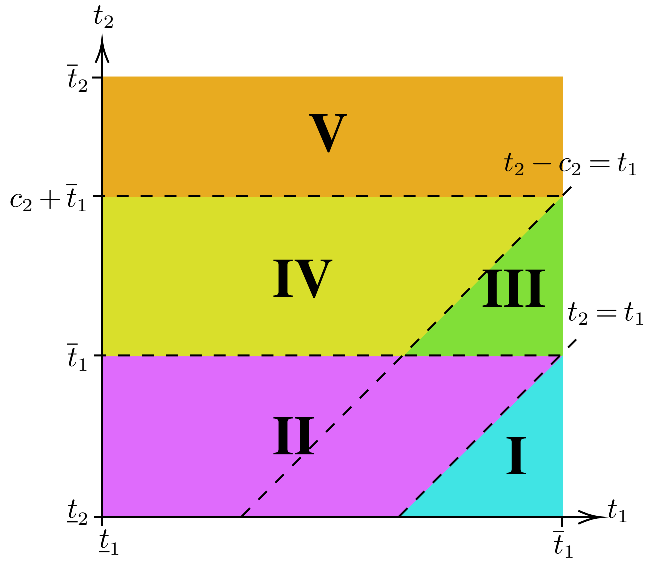

Figure 2 Partition (8) of the type space $\mathcal{T}$.

Consider first a type profile $\boldsymbol{t} \in \mathcal{T}_I$. As $t_2 < t_1$, $c_1 > 0$ and $(\boldsymbol{p}, \boldsymbol{q})$ satisfies the (FC) constraints $\sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \leq 1$ and $q_1(\boldsymbol{t}) \geq 0$, the principal cannot earn t_1 or more unless $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$.

Consider now any $\boldsymbol{t} \in \mathcal{T}_{II}$. Incentive compatibility ensures that $p_1(\boldsymbol{t}) \geq p_1(\bar{t}_1, t_2) - q_1(\bar{t}_1, t_2) = 1$, where the equality holds because $(\bar{t}_1, t_2) \in \mathcal{T}_I$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 without inspection in $\mathcal{T}_I$. Thus, the principal cannot earn t_1 or more in scenario $\boldsymbol{t}$ unless $q_1(\boldsymbol{t}) = 0$. In summary, we must again have $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$ on $\mathcal{T}_{II}$.

Next, consider any scenario $\boldsymbol{t} \in \mathcal{T}_{III}$. Incentive compatibility ensures that $p_2(\boldsymbol{t}) - q_2(\boldsymbol{t}) \leq p_2(t_1, \underline{t}_2) = 0$, where the equality holds because $(t_1, \underline{t}_2) \in \mathcal{T}_I \cup \mathcal{T}_{II}$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 without inspection throughout $\mathcal{T}_I \cup \mathcal{T}_{II}$. As the allocation probabilities are non-negative and satisfy the (FC) condition $p_2(\boldsymbol{t}) \geq q_2(\boldsymbol{t})$, we may conclude that $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$. Thus, agent 2 is inspected if he wins the good in scenario $\boldsymbol{t}$. As $t_2 - c_2 < t_1$ for all $\boldsymbol{t} \in \mathcal{T}_{III}$, the principal cannot earn t_1 or more unless $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$.

Finally, consider any $\boldsymbol{t} \in \mathcal{T}_{IV}$. Incentive compatibility ensures that $p_1(\boldsymbol{t}) \geq p_1(\bar{t}_1, t_2) - q_1(\bar{t}_1, t_2) = 1$, where the equality holds because $(\bar{t}_1, t_2) \in \mathcal{T}_{III}$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 without inspection in $\mathcal{T}_{III}$. Thus, the principal cannot earn t_1 or more in scenario $\boldsymbol{t}$ unless $q_1(\boldsymbol{t}) = 0$. Hence, we must again have $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$ on $\mathcal{T}_{IV}$.

Assume that $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton. Then, any favored-agent mechanism with favored agent $i^\star$ and threshold value $\bar{t}_{i^\star}$ allocates the good to the favored agent without inspection in all scenarios $\boldsymbol{t} \in \mathcal{T}$ with $\max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star}$, that is, it satisfies the implications of Lemma 4. Furthermore, both the type (i) and type (ii) versions of this favored-agent mechanism satisfy the optimality condition of Lemma 3. Thus, both versions of this mechanism are optimal. The next lemma implies that the type (ii) version of this mechanism is also Pareto robustly optimal.

Lemma 5. *Assume that $\mathcal{P}$ is a Markov ambiguity set of the form (6) and $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, and let $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ be the type (ii) favored-agent mechanism with favored agent $i^\star$ and threshold value $\nu^\star = \bar{t}_{i^\star}$. Then, any mechanism $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ must generate the same payoff as $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in every scenario $\boldsymbol{t} \in \mathcal{T}$, that is, we have*

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) = \sum_{i\in\mathcal{I}} (p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i) \quad \forall \boldsymbol{t} \in \mathcal{T}.$$

Lemma 5 asserts that if a feasible mechanism $(\boldsymbol{p},\boldsymbol{q})$ generates the same or a higher *expected* payoff than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ under every *distribution* $\mathbb{P}\in\mathcal{P}$, then it must generate the same payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in every *scenario* $\boldsymbol{t}\in\mathcal{T}$. This readily implies that $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ is Pareto robustly optimal in (MDP). To gain some intuition into this result, we sketch the proof of Lemma 5 when there are only two agents with $\underline{\mu}_2<\underline{\mu}_1$, such that $\arg\max_{i\in\mathcal{I}}\underline{\mu}_i=\{1\}$ is a singleton, and when $\bar{t}_2>c_2+\bar{t}_1$. To this end, we use again the partition (8) illustrated in Figure 2. In the following, we first show that any mechanism $(\boldsymbol{p},\boldsymbol{q})$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must be optimal. Hence, $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ generate the same payoff throughout $\mathcal{T}\setminus\mathcal{T}_V$ by virtue of Lemma 4. Next, we will prove that the two mechanisms earn the same payoff also on $\mathcal{T}_V$. This will establish the claim.

In accordance with the proof strategy outlined above, we fix any mechanism $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$, *i.e.*, we assume that

$$\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i-q_i(\tilde{\boldsymbol{t}})c_i)\right]\geq\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i^\star(\tilde{\boldsymbol{t}})\tilde{t}_i-q_i^\star(\tilde{\boldsymbol{t}})c_i)\right]\quad\forall\mathbb{P}\in\mathcal{P}.$$

Recall from Theorem 3 that $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ is optimal in (MDP). As the expected payoff of $(\boldsymbol{p},\boldsymbol{q})$ is non-inferior to that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ for any distribution $\mathbb{P}\in\mathcal{P}$, it is clear that $(\boldsymbol{p},\boldsymbol{q})$ is also optimal. As $\arg\max_{i\in\mathcal{I}}\underline{\mu}_i=\{1\}$ is a singleton and as $(\boldsymbol{p},\boldsymbol{q})$ is optimal, we further know from Lemma 4 that $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent 1 without inspection if $t_2-c_2<\bar{t}_1$, *i.e.*, if $\boldsymbol{t}\in\mathcal{T}\setminus\mathcal{T}_V$. Thus, the allocation probabilities of the mechanisms $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ coincide throughout $\mathcal{T}\setminus\mathcal{T}_V$.

Next, consider any $\boldsymbol{t}\in\mathcal{T}_V$. Incentive compatibility ensures that $p_2(\boldsymbol{t})-q_2(\boldsymbol{t})\leq p_2(t_1,\underline{t}_2)=0$, where the equality holds because $(t_1,\underline{t}_2)\in\mathcal{T}\setminus\mathcal{T}_V$ and because $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent 1 without inspection on $\mathcal{T}\setminus\mathcal{T}_V$. We may thus conclude that $p_2(\boldsymbol{t})=q_2(\boldsymbol{t})$. The principal's payoff under the mechanism $(\boldsymbol{p},\boldsymbol{q})$ in scenario $\boldsymbol{t}$ can thus be expressed as

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\leq p_2(\boldsymbol{t})(t_2-c_2)+p_1(\boldsymbol{t})t_1\leq t_2-c_2=\sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i-q_i^\star(\boldsymbol{t})c_i),\tag{9}$$

where the second inequality holds because $t_2-c_2\geq t_1$ on $\mathcal{T}_V$. This reasoning shows that the payoff of the mechanism $(\boldsymbol{p},\boldsymbol{q})$ cannot exceed that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ anywhere in $\mathcal{T}_V$. It remains to be shown that the mechanisms $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ actually generate the *same* payoff throughout $\mathcal{T}_V$, that is, it remains to be shown that (9) actually holds as an equality. To this end, assume for the sake of contradiction that $(\boldsymbol{p},\boldsymbol{q})$ generates a strictly lower payoff than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in scenario $\boldsymbol{t}\in\mathcal{T}_V$, that is,

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)<t_2-c_2=\sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i-q_i^\star(\boldsymbol{t})c_i).$$

By Lemma 2, there exists $\hat{\boldsymbol{t}}\in\mathcal{T}$ with $\max_{i\neq i^\star}\hat{t}_i<\hat{t}_{i^\star}$ and $\mathbb{P}\in\mathcal{P}$ such that (i) $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i]=\underline{\mu}_i$ for all $i\in\mathcal{I}$, (ii) $\mathbb{P}(\tilde{\boldsymbol{t}}\in\{\boldsymbol{t},\hat{\boldsymbol{t}}\})=1$, and (iii) $\mathbb{P}(\tilde{\boldsymbol{t}}=\boldsymbol{t})>0$. We already know that the payoff of $(\boldsymbol{p},\boldsymbol{q})$ cannot exceed that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in scenario $\hat{\boldsymbol{t}}$. Indeed, we have shown above that the two mechanisms generate the same payoff on $\mathcal{T}\setminus\mathcal{T}_V$ and that the payoff of $(\boldsymbol{p},\boldsymbol{q})$ does not exceed that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ on $\mathcal{T}_V$; see (9). By the properties (ii) and (iii) of $\mathbb{P}$, the expected payoff generated by $(\boldsymbol{p},\boldsymbol{q})$ is thus strictly lower than that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ under $\mathbb{P}$. As $\mathbb{P}\in\mathcal{P}$, however, this conclusion contradicts our initial assumption that $(\boldsymbol{p},\boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. Thus, the mechanisms $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must generate the same payoff also on $\mathcal{T}_V$. This observation implies that no feasible mechanism can Pareto robustly dominate $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$, and hence $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ is indeed Pareto robustly optimal.

The following main theorem shows that (MDP) admits a Pareto robustly optimal mechanism even if we abandon the simplifying assumption that $\arg\max_{i\in\mathcal{I}}\underline{\mu}_i$ is a singleton.

THEOREM 4. *If $\mathcal{P}$ is a Markov ambiguity set of the form (6), then any type (ii) favored-agent mechanism $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ with favored agent $i^\star\in\arg\max_{i\in\mathcal{I}}\underline{\mu}_i$ and threshold value $\nu^\star=\bar{t}_{i^\star}$ is Pareto robustly optimal in* (MDP).

The proof of Theorem 4 constructs a perturbed ambiguity set $\mathcal{P}_\varepsilon \subseteq \mathcal{P}$, which is obtained by replacing $\underline{\mu}_{i^\star}$ with $\underline{\mu}_{i^\star} + \epsilon$ in the original Markov ambiguity set (6). Here, we assume that $\varepsilon > 0$ is sufficiently small for $\mathcal{P}_\varepsilon$ to remain nonempty. By construction, the expected type of agent $i^\star$ has the largest lower bound in the perturbed ambiguity set $\mathcal{P}_\varepsilon$, whereas the expected types of all other agents have strictly smaller lower bounds. As $\mathcal{P}_\varepsilon \subseteq \mathcal{P}$, any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ with respect to $\mathcal{P}$ must also weakly Pareto robustly dominate $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ with respect to $\mathcal{P}_\varepsilon$. By Lemma 5 applied to $\mathcal{P}_\varepsilon$ instead of $\mathcal{P}$, we may thus conclude that the mechanisms $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generate the same payoff in every scenario $\boldsymbol{t} \in \mathcal{T}$. This in turn implies, however, that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ must be Pareto robustly optimal in (MDP) with respect to $\mathcal{P}$.

5. Markov Ambiguity Sets with Independent Types

The Markov ambiguity set studied in Section 4 contains distributions under which the agents' types are dependent. Sometimes, however, the principal may have good reasons to assume that the agents' types are in fact *in*dependent. In this section we thus study a subset of the Markov ambiguity set (6) studied in Section 4, which imposes the additional condition that the agents' types are mutually independent. Mathematically speaking, we thus investigate the Markov ambiguity set with independent types defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{T}) \ : \ \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{t}_i] \in [\underline{\mu}_i, \overline{\mu}_i] \ \forall i \in \mathcal{I}, \\ \tilde{t}_1, \ldots, \tilde{t}_I \text{ are mutually independent under } \mathbb{P} \end{array} \right\}. \tag{10}$$

By replacing the Markov ambiguity set (6) with its subset (10), which contains only distributions under which the agents' types are independent, we can only increase but not decrease the optimal value of problem (MDP). Proposition 2 thus implies that the optimal value of problem (MDP) with a Markov ambiguity set with independent types is bounded below by $\max_{i \in \mathcal{I}} \underline{\mu}_i$. The next proposition shows that this lower bound coincides in fact with the optimal value of problem (MDP).

Proposition 3. *If $\mathcal{P}$ is a Markov ambiguity set with independent types of the form (10), then problem (MDP) is solvable, and $z^\star = \max_{i \in \mathcal{I}} \underline{\mu}_i$.*

We do not provide a formal proof of Proposition 3 because it is similar to that of Proposition 2. However, the proof idea can be summarized as follows. Proposition 2 implies that $\max_{i \in \mathcal{I}} \underline{\mu}_i$ provides a lower bound on $z^\star$. Proposition 3 thus follows if we can show that $\max_{i \in \mathcal{I}} \underline{\mu}_i$ provides also an upper bound on $z^\star$. This is indeed the case because the agents' types are independent under the Dirac distribution $\delta_{\underline{\boldsymbol{\mu}}}$ that concentrates unit mass at $\underline{\boldsymbol{\mu}}$, which implies that $\delta_{\underline{\boldsymbol{\mu}}}$ belongs to the Markov ambiguity set with independent types, and because the expected payoff of *any* feasible mechanism under $\delta_{\underline{\boldsymbol{\mu}}}$ is bounded above by $\max_{i \in \mathcal{I}} \underline{\mu}_i$. Propositions 2 and 3 together suggest, perhaps surprisingly, that the principal does not benefit from knowing whether or not the agents' types are independent. At least, this information has no impact on the optimal worst-case expected payoff.

Theorem 5 below is reminiscent of Theorem 3 from Section 4 and shows that there is again a continuum of infinitely many optimal favored-agent mechanisms.

Theorem 5. *If $\mathcal{P}$ is a Markov ambiguity set of the form (10), then any favored-agent mechanism with favored agent $i^\star \in \arg\max_{i \in \mathcal{I}} \underline{\mu}_i$ and threshold value $\nu^\star \geq \max_{i \in \mathcal{I}} \underline{\mu}_i$ is optimal in (MDP).*

Comparing Theorems 3 and 5 reveals that information about the independence of the agents does not affect the choice of the optimal favored agent. However, it reduces the lowest optimal threshold value from $\bar{t}_{i^\star}$ to $\max_{i \in \mathcal{I}} \underline{\mu}_i$. Thus, the set of optimal favored-agent mechanisms *increases* if the principal learns that the agents are independent. This insight is unsurprising in view of the impossibility to monetize such independence information (at least in the worst case), which implies that all mechanisms that were optimal under a Markov ambiguity set of the form (6) remain optimal under a Markov ambiguity set of the form (10). However, the independence information allows the principal to choose an optimal threshold value that is independent of the favored agent.

Remark 3. Theorem 5 is sharp in the sense that there are problem instances for which any favored-agent mechanism with favored agent $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and threshold value $\nu < \max_{i\in\mathcal{I}} \underline{\mu}_i$ is strictly suboptimal. To see this, we revisit the instance of problem (MDP) described in Remark 2, which involves $I=2$ agents with $c_1=c_2=2$ and a Markov ambiguity set of the form (6) with parameters $\mathcal{T}_1=[1,6]$, $\mathcal{T}_2=[0,10]$, $[\underline{\mu}_1,\overline{\mu}_1]=[4,5]$ and $[\underline{\mu}_2,\overline{\mu}_2]=[3,7]$. Now, however, we additionally assume that the agents' types are independent such that the ambiguity becomes an instance of (10). Hence, by Proposition 3, the optimal value of problem (MDP) still amounts to $\max_{i\in\mathcal{I}} \underline{\mu}_i=4$. Consider now any favored agent mechanism with favored agent $1\in\arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and threshold value $\nu<\max_{i\in\mathcal{I}} \underline{\mu}_i=4$. In the following, we prove that this mechanism is suboptimal. To this end, assume first that $\nu<1$. If $\boldsymbol{t}=\underline{\boldsymbol{\mu}}=(4,3)$, then the mechanism allocates the good to agent 1 with verification and earns a payoff of $t_1-c_1=2$. As the discrete distribution that assigns unit probability to the scenario $\underline{\boldsymbol{\mu}}$ belongs to the Markov ambiguity set (10), the worst-case expected payoff over all admissible distributions cannot exceed 2, which is strictly smaller than the optimal worst-case expected payoff. Assume next that $\nu\in[1,4)$, and consider the discrete distribution $\mathbb{P}$ that assigns probability $\frac{1}{2}$ to the scenarios $(4,5+\nu/4)$ and $(4,2)$ each. One readily verifies that $\tilde{t}_1$ is deterministic under $\mathbb{P}$ and that $\mathbb{P}$ belongs to the Markov ambiguity set (10). In scenario $(4,5+\nu/4)$, the mechanism allocates the good to agent 2 with verification because $t_2-c_2=(5+\nu/4)-2>\nu$ and $t_2-c_2>2=t_1-c_1$. In scenario $(4,2)$, on the other hand, the mechanism allocates the good to agent 1 without verification. Consequently, the expected payoff of the mechanism under the distribution $\mathbb{P}$ amounts to $\frac{1}{2}(3+\nu/4)+\frac{1}{2}4=3.5+\nu/8$, and the worst-case expected payoff over all admissible distributions cannot exceed $3.5+\nu/8$, which is strictly smaller than the optimal worst-case expected payoff. In summary, the mechanism is strictly suboptimal for all $\nu<4$. □

As in Sections 3 and 4, we now show that among all optimal favored-agent mechanisms identified in Theorem 5, there is always one that is Pareto robustly optimal; see Theorem 6 below. We first prove this result under the simplifying assumption that $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ is a singleton, in which case the favored agent is uniquely determined. However, Theorem 6 will *not* depend on this assumption.

We first show that if $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i=\{i^\star\}$, then any type profile $\boldsymbol{t}\in\mathcal{T}$ has a strictly positive probability under some discrete distribution in the Markov ambiguity set (10) with a prescribed expected value of $\tilde{t}_{i^\star}$. It is further possible to require that, under this distribution, the type of each agent is supported on merely two points, the smaller of which falls strictly below $\underline{\mu}_{i^\star}$ for every $i\neq i^\star$.

Lemma 6. *If $\mathcal{P}$ is a Markov ambiguity set of the form (10) and $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i=\{i^\star\}$ is a singleton, then, for any type profile $\boldsymbol{t}\in\mathcal{T}$ and any expected value $\mu_{i^\star}\in[\underline{\mu}_{i^\star},\overline{\mu}_{i^\star}]$ there exist a scenario $\hat{\boldsymbol{t}}\in\mathcal{T}$ with $\max_{i\neq i^\star}\hat{t}_i<\underline{\mu}_{i^\star}$ and a discrete distribution $\mathbb{P}\in\mathcal{P}$ such that (i) $\mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}]=\mu_{i^\star}$, (ii) $\mathbb{P}(\tilde{t}_i\in\{t_i,\hat{t}_i\})=1$ for all $i\in\mathcal{I}$, and (iii) $\mathbb{P}(\tilde{\boldsymbol{t}}=\boldsymbol{t})>0$.*

Lemma 6 is instrumental to prove the following payoff equivalence result.

Lemma 7. *Assume that $\mathcal{P}$ is a Markov ambiguity set of the form (10) and $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i=\{i^\star\}$ is a singleton. If $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$ weakly Pareto robustly dominates $(\boldsymbol{p}',\boldsymbol{q}')\in\mathcal{X}$ and if*

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\leq\sum_{i\in\mathcal{I}}(p_i'(\boldsymbol{t})t_i-q_i'(\boldsymbol{t})c_i)\quad\forall\boldsymbol{t}\in\mathcal{T}' \tag{11}$$

for some rectangular set $\mathcal{T}'=\prod_{i\in\mathcal{I}}\mathcal{T}_i'\subseteq\mathcal{T}$ with (i) $[\underline{t}_i,\underline{\mu}_{i^\star})\subseteq\mathcal{T}_i'$ for all $i\neq i^\star$ and (ii) either $\mathcal{T}_{i^\star}'\subseteq[\underline{\mu}_{i^\star},\overline{\mu}_{i^\star}]$ or $\mathcal{T}_{i^\star}'=\mathcal{T}_{i^\star}$, then the inequality (11) must in fact hold as an equality.

Lemma 7 asserts that if $(\boldsymbol{p},\boldsymbol{q})$ generates the same or a higher *expected* payoff than $(\boldsymbol{p}',\boldsymbol{q}')$ under every distribution $\mathbb{P}\in\mathcal{P}$ and $(\boldsymbol{p}',\boldsymbol{q}')$ generates the same or a higher payoff than $(\boldsymbol{p},\boldsymbol{q})$ in every scenario in $\boldsymbol{t}\in\mathcal{T}'$, then the two mechanisms must generate the *same* payoff in every scenario $\boldsymbol{t}\in\mathcal{T}'$. In the special case when $(\boldsymbol{p}',\boldsymbol{q}')$ is the type (i) favored-agent mechanism with favored agent $i^\star$ and threshold value $\nu^\star=\underline{\mu}_{i^\star}$, where we continue to assume that $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i=\{i^\star\}$ is a singleton, we can prove a strengthened payoff equivalence result that holds across the entire type space $\mathcal{T}$.

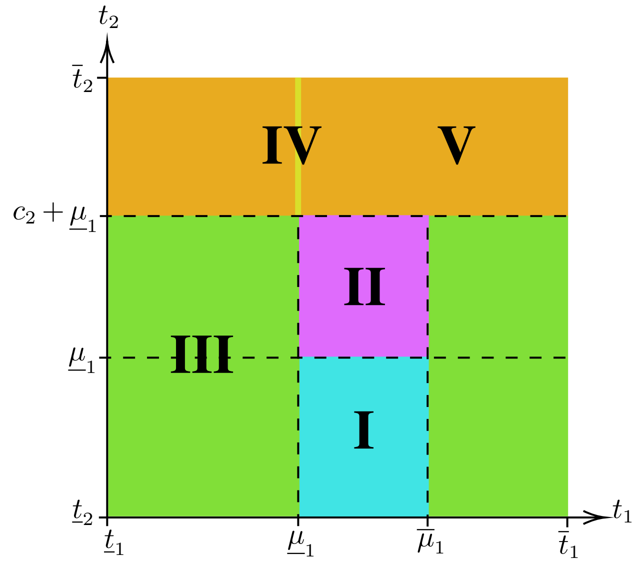

Figure 3 Partition (12) of the type space $\mathcal{T}$.

Lemma 8. *Assume that $\mathcal{P}$ is a Markov ambiguity set of the form (10) and $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, and let $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ be the type (i) favored-agent mechanism with favored agent $i^\star$ and threshold value $\nu^\star = \underline{\mu}_{i^\star}$. Then, any mechanism $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must generate the same payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in every scenario $\boldsymbol{t}\in\mathcal{T}$, that is, we have*

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) = \sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i) \quad \forall \boldsymbol{t}\in\mathcal{T}.$$

Lemma 8 is reminiscent of Lemma 5. It asserts that if a feasible mechanism $(\boldsymbol{p},\boldsymbol{q})$ generates the same or a higher *expected* payoff than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ under every *distribution* $\mathbb{P}\in\mathcal{P}$, then it must generate the same payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in every *scenario* $\boldsymbol{t}\in\mathcal{T}$. This readily implies that $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ is Pareto robustly optimal in (MDP). To gain some intuition into this result, we sketch the proof Lemma 8 when there are only two agents with $\underline{\mu}_2 < \underline{\mu}_1$, such that $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{1\}$ is a singleton, and where $\bar{t}_2 > c_2 + \underline{\mu}_1$. These simplifying assumptions prevent tedious case distinctions. Our arguments will rely on the following partition of the type space $\mathcal{T}$, which is illustrated in Figure 3.

$$\begin{aligned}
\mathcal{T}_I &= \{\boldsymbol{t}\in\mathcal{T} : t_1\in(\underline{\mu}_1,\overline{\mu}_1] \text{ and } t_2\le\underline{\mu}_1\}\\
\mathcal{T}_{II} &= \{\boldsymbol{t}\in\mathcal{T} : t_1\in(\underline{\mu}_1,\overline{\mu}_1],\ t_2>\underline{\mu}_1 \text{ and } t_2-c_2\le\underline{\mu}_1\}\\
\mathcal{T}_{III} &= \{\boldsymbol{t}\in\mathcal{T} : t_1\notin(\underline{\mu}_1,\overline{\mu}_1] \text{ and } t_2-c_2\le\underline{\mu}_1\}\\
\mathcal{T}_{IV} &= \{\boldsymbol{t}\in\mathcal{T} : t_1=\underline{\mu}_1 \text{ and } t_2-c_2>\underline{\mu}_1\}\\
\mathcal{T}_{V} &= \{\boldsymbol{t}\in\mathcal{T} : t_1\neq\underline{\mu}_1 \text{ and } t_2-c_2>\underline{\mu}_1\}
\end{aligned} \tag{12}$$

Lemma 9 in the appendix proves that the above rectangular sets form indeed a partition of $\mathcal{T}$. Under our simplifying assumptions about $\underline{\mu}_1$, $\underline{\mu}_2$, $\bar{t}_2$ and c_2, one can further show that all of these sets are nonempty. We emphasize, however, that the formal proof of Lemma 8 in the appendix does not rely on any of the simplifying assumptions imposed here. Choose now any mechanism $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. In the following, we will show that this mechanism generates the same payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in all scenarios $\boldsymbol{t}\in\mathcal{T}$. We will prove this assertion via a spatial induction argument centered around the subsets (12) of the type space, where the induction step critically relies on the non-locality of the incentive compatibility constraint (IC). In addition, we will repeatedly exploit Lemma 7.

Consider first any $\boldsymbol{t} \in \mathcal{T}_I$. The payoff generated by $(\boldsymbol{p}, \boldsymbol{q})$ in this scenario satisfies

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq \sum_{i\in\mathcal{I}} p_i(\boldsymbol{t})t_i \leq t_1 = \sum_{i\in\mathcal{I}} (p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the first inequality exploits the non-negativity of $q_i(\boldsymbol{t})$ and c_i, the second inequality holds because of (FC) and because $t_1 > t_2$ on $\mathcal{T}_I$, and the equality follows from the defining properties of the favored-agent mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. As $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ and as $\mathcal{T}_I = \mathcal{T}_{I1} \times \mathcal{T}_{I2}$ with $\mathcal{T}_{I1} = (\underline{\mu}_1, \overline{\mu}_1]$ and $\mathcal{T}_{I2} = [\underline{t}_2, \underline{\mu}_1]$, Lemma 7 ensures that the above inequality holds in fact as an equality. Thus, the payoffs of $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ coincide throughout $\mathcal{T}_I$. In addition, note that $t_2 < t_1$, $q_i(\boldsymbol{t}) \geq 0$, $c_i > 0$ and $\sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \leq 1$ for all $\boldsymbol{t} \in \mathcal{T}_I$. This implies that $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$ throughout $\mathcal{T}_I$ for otherwise the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ could not be equal to t_1.

Next, consider any $\boldsymbol{t} \in \mathcal{T}_{II}$. Incentive compatibility ensures that $p_2(\boldsymbol{t}) - q_2(\boldsymbol{t}) \leq p_2(t_1, \underline{t}_2) = 0$, where the equality holds because $(t_1, \underline{t}_2) \in \mathcal{T}_I$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 without inspection throughout $\mathcal{T}_I$. As the allocation probabilities are non-negative and satisfy the (FC) condition $p_2(\boldsymbol{t}) \geq q_2(\boldsymbol{t})$, we thus have $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$. This implies that

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq p_1(\boldsymbol{t})t_1 + p_2(\boldsymbol{t})(t_2 - c_2) \leq t_1 = \sum_{i\in\mathcal{I}} (p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the first inequality exploits the non-negativity of $q_1(\boldsymbol{t})$ and c_1, the second inequality holds because of (FC) and because of the defining properties $t_2 - c_2 \leq \underline{\mu}_1 < t_1$ of the set $\mathcal{T}_{II}$, and the equality follows from the definition of the favored-agent mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Next, observe that $\mathcal{T}' = \mathcal{T}_I \cup \mathcal{T}_{II}$ can be represented as $\mathcal{T}' = \mathcal{T}_1' \times \mathcal{T}_2'$, where $\mathcal{T}_1' = (\underline{\mu}_1, \overline{\mu}_1]$ and $\mathcal{T}_2' = [\underline{t}_2, c_2 + \underline{\mu}_1]$. The above reasoning implies that the payoff generated by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is non-inferior to that generated by $(\boldsymbol{p}, \boldsymbol{q})$ throughout $\mathcal{T}'$. As $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, Lemma 7 guarantees that the two mechanisms must in fact generate the *same* payoff throughout $\mathcal{T}' = \mathcal{T}_I \cup \mathcal{T}_{II}$. In addition, note that $c_1, c_2 > 0$ and that $t_2 - c_2 \leq \underline{\mu}_1 < t_1$, $q_i(\boldsymbol{t}) \geq 0$ and $\sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \leq 1$ for all $\boldsymbol{t} \in \mathcal{T}_{II}$. This implies that $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$ on $\mathcal{T}_{II}$ for otherwise the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ could not be equal to t_1.

Consider now any $\boldsymbol{t} \in \mathcal{T}_{III}$. Incentive compatibility ensures that $p_1(\boldsymbol{t}) \geq p_1(\overline{\mu}_1, t_2) - q_1(\overline{\mu}_1, t_2) = 1$, where the equality holds because $(\overline{\mu}_1, t_2) \in \mathcal{T}_I \cup \mathcal{T}_{II}$ and because we know from before that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 without inspection throughout $\mathcal{T}_I \cup \mathcal{T}_{II}$. Thus, we have again

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq p_1(\boldsymbol{t})t_1 - q_1(\boldsymbol{t})c_1 \leq t_1 = \sum_{i\in\mathcal{I}} (p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the first inequality follows from substituting $p_1(\boldsymbol{t}) = 1$ into (FC), which implies that $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t}) = 0$, the second inequality holds because $q_1(\boldsymbol{t})$ and c_1 are non-negative, and the equality follows from the definition of the favored-agent mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Next, observe that $\mathcal{T}'' = \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$ can be represented as $\mathcal{T}'' = \mathcal{T}_1'' \times \mathcal{T}_2''$, where $\mathcal{T}_1'' = \mathcal{T}_1 = [\underline{t}_1, \bar{t}_1]$ and $\mathcal{T}_2'' = [\underline{t}_2, c_2 + \underline{\mu}_1]$. The above reasoning implies that the payoff generated by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is non-inferior to that generated by $(\boldsymbol{p}, \boldsymbol{q})$ on the whole of $\mathcal{T}''$. As $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, Lemma 7 guarantees that the two mechanisms generate the *same* payoff on $\mathcal{T}'' = \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$. In addition, recall that $c_1 > 0$, that $p_1(\boldsymbol{t}) \geq q_1(\boldsymbol{t}) \geq 0$ and $\sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \leq 1$ by feasibility and that $p_1(\boldsymbol{t}) = 1$ by incentive compatibility. Hence, $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$ on $\mathcal{T}_{III}$ for otherwise the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ could not be equal to t_1.

It remains to be shown that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ and $(\boldsymbol{p}, \boldsymbol{q})$ generate the same payoff on the two remaining sets $\mathcal{T}_{IV}$ and $\mathcal{T}_V$. To this end, we first show that if agent 2 is allocated the good in any scenario $\boldsymbol{t} \in \mathcal{T}_{IV} \cup \mathcal{T}_V$, then he should be inspected. Indeed, incentive compatibility ensures that $p_2(\boldsymbol{t}) - q_2(\boldsymbol{t}) \leq p_2(t_1, \underline{t}_2) = 0$, where the equality holds because $(t_1, \underline{t}_2) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$ and because $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 1 without inspection throughout $\mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$. As the allocation probabilities are non-negative and satisfy the (FC) condition $p_2(\boldsymbol{t}) \geq q_2(\boldsymbol{t})$, we have $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$.

Consider now any $\boldsymbol{t} \in \mathcal{T}_{IV}$. As $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$, we have

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq p_1(\boldsymbol{t})t_1 + p_2(\boldsymbol{t})(t_2 - c_2) \leq t_2 - c_2 = \sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the first inequality exploits the non-negativity of $q_1(\boldsymbol{t})$ and c_1, the second inequality holds because of (FC) and because of the defining properties $t_1 = \underline{\mu}_1 < t_2 - c_2$ of the set $\mathcal{T}_{IV}$, and the equality follows from the definition of the favored-agent mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Next, observe that $\mathcal{T}''' = \{\boldsymbol{t} \in \mathcal{T}_{III} : t_1 = \underline{\mu}_1\} \cup \mathcal{T}_{IV}$ can be expressed as $\mathcal{T}''' = \{\underline{\mu}_1\} \times \mathcal{T}_2$. As $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, Lemma 7 guarantees that the two mechanisms generate in fact the *same* payoff on $\mathcal{T}'''$—and thus also on $\mathcal{T}_{IV}$. In addition, as $t_1 = \underline{\mu}_1 < t_2 - c_2$ and $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$, we have $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t}) = 1$ on $\mathcal{T}_{IV}$ for otherwise the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ could not be equal to $t_2 - c_2$.

Finally, consider any $\boldsymbol{t} \in \mathcal{T}_V$. Incentive compatibility ensures that $p_1(\boldsymbol{t}) - q_1(\boldsymbol{t}) \leq p_1(\underline{\mu}_1, t_2) = 0$, where the equality holds because $(\underline{\mu}_1, t_2) \in \mathcal{T}_{IV}$ and because $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent 2 throughout $\mathcal{T}_{IV}$. As the allocation probabilities further satisfy the (FC) condition $p_1(\boldsymbol{t}) \geq q_1(\boldsymbol{t})$, we may conclude that $p_1(\boldsymbol{t}) = q_1(\boldsymbol{t})$. Recalling our earlier insight that $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$, we thus obtain

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) = \sum_{i\in\mathcal{I}} p_i(\boldsymbol{t})(t_i - c_i) \leq \max_{i\in\mathcal{I}} t_i - c_i = \sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the inequality follows from the (FC) constraint $\sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \leq 1$, while the second equality follows from the definition of the favored-agent mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Together with our earlier results, this implies that the payoff generated by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is non-inferior to that generated by $(\boldsymbol{p}, \boldsymbol{q})$ throughout $\mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2$. As $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, Lemma 7 guarantees that the two mechanisms generate in fact the *same* payoff on $\mathcal{T}$. Thus, we have shown that any feasible mechanism that weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ must generate the same payoff as $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in every scenario $\boldsymbol{t} \in \mathcal{T}$, which implies that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is Pareto robustly optimal in (MDP).

The next theorem shows that the Pareto robust optimality result of Lemma 8 remains valid even when $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ is no longer guaranteed to be a singleton.

THEOREM 6. *If $\mathcal{P}$ is a Markov ambiguity set of the form (10), then any type (i) favored-agent mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ with favored agent $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and threshold value $\nu^\star = \max_{i\in\mathcal{I}} \underline{\mu}_i$ is Pareto robustly optimal in (MDP).*

The proof of Theorem 6 is omitted because it widely parallels that of Theorem 4. To close this section, we show that the favored-agent mechanism identified in Theorem 4 may cease to be Pareto robustly optimal when the Markov ambiguity set (6) is replaced with its subset (10) that imposes independence among the agents' types.

REMARK 4. Consider an instance of the robust mechanism design problem (MDP) with $I = 2$ agents, and assume that the input parameters satisfy $\underline{\mu}_1 > \underline{\mu}_2$ and $\bar{t}_2 - c_2 \geq \bar{t}_1 > \overline{\mu}_1$. In addition, let $(\boldsymbol{p}, \boldsymbol{q})$ be a type (ii) favored-agent mechanism with favored agent $1 \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and threshold value $\nu \geq \bar{t}_1$. In the special case where $\nu = \bar{t}_1$, Theorem 4 implies that $(\boldsymbol{p}, \boldsymbol{q})$ is Pareto robustly optimal in (MDP) provided that $\mathcal{P}$ is the Markov ambiguity set (6). In the following, we prove that, for any $\nu \geq \bar{t}_1$, $(\boldsymbol{p}, \boldsymbol{q})$ is Pareto robustly dominated by another feasible mechanism when $\mathcal{P}$ is the Markov ambiguity set (10) with independent types.[1] To this end, consider an arbitrary distribution $\mathbb{P}$ in the Markov ambiguity set (10). The expected payoff of $(\boldsymbol{p}, \boldsymbol{q})$ is then given by

$$\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\{1,2\}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i)\right] = \begin{cases} \quad\mathbb{P}\left(\tilde{t}_2 - c_2 \leq \overline{\mu}_1\right)\mathbb{E}_{\mathbb{P}}[\tilde{t}_1 \,|\, \tilde{t}_2 - c_2 \leq \overline{\mu}_1] \\ +\mathbb{P}\left(\tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)\right)\mathbb{E}_{\mathbb{P}}[\tilde{t}_1 \,|\, \tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)] \\ +\mathbb{P}\left(\tilde{t}_2 - c_2 \geq \nu\right)\mathbb{E}_{\mathbb{P}}[\max_{i\in\{1,2\}} \tilde{t}_i - c_i \,|\, \tilde{t}_2 - c_2 \geq \nu] \end{cases}$$

[1] One can use similar arguments to prove that the type (i) variant of the same favored-agent mechanism is also Pareto dominated. Details are omitted for the sake of brevity.

$$= \begin{cases} \quad \mathbb{P}\left(\tilde{t}_2 - c_2 \leq \overline{\mu}_1\right) \mathbb{E}_{\mathbb{P}}[\tilde{t}_1] \\ +\mathbb{P}\left(\tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)\right) \mathbb{E}_{\mathbb{P}}[\tilde{t}_1] \\ +\mathbb{P}\left(\tilde{t}_2 - c_2 \geq \nu\right) \mathbb{E}_{\mathbb{P}}[\max_{i\in\{1,2\}} \tilde{t}_i - c_i \,|\, \tilde{t}_2 - c_2 \geq \nu], \end{cases}$$

where the second equality follows from the independence of the agents' types under $\mathbb{P}$. Next, denote by $(\boldsymbol{p}', \boldsymbol{q}')$ the type (i) favored-agent mechanism with favored agent 1 and threshold value $\overline{\mu}_1$. By construction, the expected payoff of $(\boldsymbol{p}', \boldsymbol{q}')$ under $\mathbb{P}$ amounts to

$$\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\{1,2\}} (p_i'(\boldsymbol{t})t_i - q_i'(\boldsymbol{t})c_i)\right] = \begin{cases} \quad \mathbb{P}\left(\tilde{t}_2 - c_2 \leq \overline{\mu}_1\right) \mathbb{E}_{\mathbb{P}}[\tilde{t}_1] \\ +\mathbb{P}\left(\tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)\right) \mathbb{E}_{\mathbb{P}}[\max_{i\in\{1,2\}} \tilde{t}_i - c_i \,|\, \tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)] \\ +\mathbb{P}\left(\tilde{t}_2 - c_2 \geq \nu\right) \mathbb{E}_{\mathbb{P}}[\max_{i\in\{1,2\}} \tilde{t}_i - c_i \,|\, \tilde{t}_2 - c_2 \geq \nu]. \end{cases}$$

If $\mathbb{P}(\tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)) > 0$, then the expected payoff of $(\boldsymbol{p}', \boldsymbol{q}')$ exceeds that of $(\boldsymbol{p}, \boldsymbol{q})$ under $\mathbb{P}$ because

$$\max_{i\in\{1,2\}} t_i - c_i \geq t_2 - c_2 > \overline{\mu}_1 \geq \mathbb{E}_{\mathbb{P}}[\tilde{t}_1]$$

for all $\boldsymbol{t} \in \mathcal{T}$ with $t_2 - c_2 \in (\overline{\mu}_1, \nu)$. If $\mathbb{P}\left(\tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)\right) = 0$, on the other hand, then the expected payoffs of the two mechanisms coincide. In order to show that $(\boldsymbol{p}', \boldsymbol{q}')$ Pareto robustly dominates $(\boldsymbol{p}, \boldsymbol{q})$ it thus suffices to construct a distribution $\mathbb{P}^\star \in \mathcal{P}$ with $\mathbb{P}^\star(\tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)) > 0$. Such a distribution exists thanks to our assumption that $\overline{t}_2 - c_2 \geq \overline{t}_1 > \overline{\mu}_1$. Indeed, we can define $\mathbb{P}^\star$ as the two-point distribution that assigns probability $\alpha = (\underline{\mu}_2 - \underline{t}_2)/((\overline{t}_1 + \overline{\mu}_1)/2 + c_2 - \underline{t}_2)$ to scenario $(\underline{\mu}_1, (\overline{t}_1 + \overline{\mu}_1)/2 + c_2)$ and probability $1 - \alpha$ to scenario $(\underline{\mu}_1, \underline{t}_2)$. One readily verifies that this distribution belongs to the ambiguity set (10) and satisfies $\mathbb{P}^\star(\tilde{t}_2 - c_2 \in (\overline{\mu}_1, \nu)) \geq \alpha > 0$. Hence, the favored-agent mechanism $(\boldsymbol{p}, \boldsymbol{q})$ fails to be Pareto robustly optimal in problem (MDP) for any $\nu \geq \overline{t}_1$ if $\mathcal{P}$ is a Markov ambiguity set of the form (10) with independent types. □

In conjunction, Remarks 2 and 4 imply that for some instances of problem (MDP) there is *no* favored-agent mechanism that is Pareto robustly optimal simultaneously for *both* Markov ambiguity sets (6) and (10). To see this, consider the instance of problem (MDP) described in Remark 2, and note that this instance satisfies all assumptions of Remark 4. One readily verifies that every favored-agent mechanism with favored agent $i^\star \notin \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ is strictly suboptimal (and thus fails to be Pareto robustly optimal) for this problem instance under *both* ambiguity sets (6) and (10). By Remark 2, any favored-agent mechanism with favored agent $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and threshold value $\nu < \overline{t}_{i^\star}$ is strictly suboptimal (and thus fails to be Pareto robustly optimal) under the ambiguity set (6). By Remark 4, finally, any favored-agent mechanism with favored agent $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and threshold value $\nu \geq \overline{t}_{i^\star}$ fails to be Pareto robustly optimal under the ambiguity set (10). This implies that it is crucial for the principal to know whether or not the agents' types are independent.

6. Conclusions

This paper studies optimal allocation problems with costly verification. Many allocation problems of this kind recur infrequently or never, and therefore it is unreasonable to assume that the principal has full knowledge of the distribution of the agents' types. This prompts us to formulate these allocation problems as distributionally robust mechanism design problems that explicitly account for (and hedge against) distributional ambiguity. We show that—like in the classical stochastic setting [5]—simple and interpretable mechanisms are optimal despite the extra layer of complexity introduced by the distributional ambiguity. Specifically, for three natural but increasingly restrictive ambiguity sets for the type distribution, we identify a large family of robustly optimal favored-agent mechanisms that maximize the principal's worst-case expected payoff. Moreover, for each of the three ambiguity sets, we identify a Pareto robustly optimal mechanism from within the family of all robustly optimal favored-agent mechanisms. These Pareto robustly optimal mechanisms not only maximize the worst-case expected payoff of the principal but also perform well when non-worst-case conditions prevail. In fact, these mechanisms strike an optimal trade-off between the expected payoffs under all distributions in the ambiguity set.

The main results of this paper offer several insights of practical relevance. First, there is merit in acquiring information about the expected values of the agents' types. Indeed, at optimality, the principal's worst-case expected payoff is strictly higher under Markov ambiguity sets than under support-only ambiguity sets. For any given Markov ambiguity set, however, the principal does not benefit from knowing whether or not the agents' types are independent. At least, this information has no impact on the optimal worst-case expected payoff. Misrepresenting the agents' types as independent random variables may nevertheless have undesirable consequences, that is, it may mislead the principal into adopting a mechanism that fails to be Pareto robustly optimal and may even fail to be robustly optimal. We believe that the agents' types are unlikely to be independent in allocation problems with costly verification that arise naturally in reality. For example, a venture capitalist assigning seed funding to one of several start-up companies would be ill-advised to assume independence because innovations are often driven by societal trends, technical developments, or disruptive events (*e.g.*, the COVID-19 pandemic has led to a wave of supply chain start-ups), and therefore the potential gains from investing in these innovations cannot be independent.

References

[1] Anunrojwong J, Balseiro S, Besbes O (2022) On the robustness of second-price auctions in prior-independent mechanism design. *arXiv preprint arXiv:2204.10478* .

[2] Balseiro SR, Kim A, Russo D (2021) On the futility of dynamics in robust mechanism design. *Operations Research* 69(6):1767–1783.

[3] Bandi C, Bertsimas D (2014) Optimal design for multi-item auctions: Aa robust optimization approach. *Mathematics of Operations Research* 39(4):1012–1038.

[4] Bayrak HI, Güler K, Pınar MÇ (2017) Optimal allocation with costly inspection and discrete types under ambiguity. *Optimization Methods and Software* 32(4):699–718.

[5] Ben-Porath E, Dekel E, Lipman BL (2014) Optimal allocation with costly verification. *American Economic Review* 104(12):3779–3813.

[6] Bertsimas D, ten Eikelder S, den Hertog D, Trichakis N (2020) Pareto adaptive robust optimality via a Fourier-Motzkin elimination lens. *arXiv preprint arXiv:2012.04419* .

[7] Carrasco V, Luz VF, Kos N, Messner M, Monteiro P, Moreira H (2018) Optimal selling mechanisms under moment conditions. *Journal of Economic Theory* 177:245–279.

[8] Chen YC, Hu G, Yang X (2022) Information design in allocation with costly verification. *arXiv preprint arXiv:2210.16001* .

[9] Chen Z, Hu Z, Wang R (2021) Screening with limited information: The minimax theorem and a geometric approach. *Available at SSRN 3940212* .

[10] Chua GA, Hu G, Liu F (2019) Optimal multi-unit allocation with costly verification. *Available at SSRN 3407031* .

[11] Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58(3):595–612.

[12] Fudenberg D, Tirole J (1991) *Game Theory* (MIT Press).

[13] He W, Li J (2022) Correlation-robust auction design. *Journal of Economic Theory* 200:105403.

[14] Iancu DA, Trichakis N (2014) Pareto efficiency in robust optimization. *Management Science* 60(1):130–147.

[15] Koçyiğit Ç, Iyengar G, Kuhn D, Wiesemann W (2020) Distributionally robust mechanism design. *Management Science* 66(1):159–189.

[16] Koçyiğit Ç, Rujeerapaiboon N, Kuhn D (2021) Robust multidimensional pricing: Separation without regret. *Mathematical Programming* 1–34.

[17] Li Y (2020) Mechanism design with costly verification and limited punishments. *Journal of Economic Theory* 186:105000.

[18] Mylovanov T, Zapechelnyuk A (2017) Optimal allocation with ex post verification and limited penalties. *American Economic Review* 107(9):2666–94.

[19] Pınar MÇ, Kızılkale C (2017) Robust screening under ambiguity. *Mathematical Programming* 163(1):273–299.

[20] Suzdaltsev A (2020) An optimal distributionally robust auction. *arXiv preprint arXiv:2006.05192* .

[21] Townsend RM (1979) Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory* 21(2):265–293.

[22] Walton D, Carroll G (2022) A general framework for robust contracting models. *Econometrica* .

[23] Wang S, Liu S, Zhang J (2020) Minimax regret mechanism design with moments information. *Available at SSRN 3707021* .

[24] Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Operations Research* 62(6):1358–1376.

Appendix. Proofs

Proof of Proposition 1. Relaxing the incentive compatibility constraints and the first inequality in (FC) yields

$$\begin{aligned} z^\star \leq \quad & \sup_{\boldsymbol{p},\boldsymbol{q}} \quad \inf_{\boldsymbol{t}\in\mathcal{T}} \sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \\ & \text{s.t.} \quad p_i, q_i \in \mathcal{L}(\mathcal{T},[0,1]) \ \forall i\in\mathcal{I} \\ & \qquad \sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \leq 1 \ \forall \boldsymbol{t}\in\mathcal{T} \\ = \quad & \sup_{\boldsymbol{p}} \quad \inf_{\boldsymbol{t}\in\mathcal{T}} \sum_{i\in\mathcal{I}} p_i(\boldsymbol{t})t_i \\ & \text{s.t.} \quad p_i \in \mathcal{L}(\mathcal{T},[0,1]) \ \forall i\in\mathcal{I}, \ \sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \leq 1 \ \forall \boldsymbol{t}\in\mathcal{T}, \end{aligned}$$

where the equality holds because in the relaxed problem it is optimal to set $q_i(\boldsymbol{t}) = 0$ for all $i\in\mathcal{I}$ and $\boldsymbol{t}\in\mathcal{T}$. As the resulting maximization problem over $\boldsymbol{p}$ is separable with respect to $\boldsymbol{t}\in\mathcal{T}$, it is optimal to allocate the good in each scenario $\boldsymbol{t}\in\mathcal{T}$—with probability one—to an agent with maximal type. Therefore, $z^\star$ is bounded above by $\inf_{\boldsymbol{t}\in\mathcal{T}} \max_{i\in\mathcal{I}} t_i = \max_{i\in\mathcal{I}} \underline{t}_i$. However, this bound is attained by a mechanism that allocates the good to an agent $i' \in \arg\max_{i\in\mathcal{I}} \underline{t}_i$ irrespective of $\boldsymbol{t}\in\mathcal{T}$ and never inspects anyone's type. Since this mechanism is feasible, the claim follows. □

Proof of Theorem 1. Select an arbitrary favored-agent mechanism with $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{t}_i$ and $\nu^\star \geq \max_{i\in\mathcal{I}} \underline{t}_i$. Recall first that this mechanism is feasible in (3). Next, we will show that this mechanism attains a worst-case payoff that is at least as large as $\max_{i\in\mathcal{I}} \underline{t}_i$, which implies via Proposition 1 that it is in fact optimal in (3). To this end, fix an arbitrary type profile $\boldsymbol{t}\in\mathcal{T}$. If $\max_{i\neq i^\star} t_i - c_i < \nu^\star$, then condition (i) in Definition 3 implies that the principal's payoff amounts to $t_{i^\star} \geq \max_{i\in\mathcal{I}} \underline{t}_i$, where the inequality follows from the selection of $i^\star$. If $\max_{i\neq i^\star} t_i - c_i > \nu^\star$, then condition (ii) in Definition 3 implies that the principal's payoff amounts to $\max_{i\in\mathcal{I}} t_i - c_i > \nu^\star \geq \max_{i\in\mathcal{I}} \underline{t}_i$, where the second inequality follows from the selection of $\nu^\star$. If $\max_{i\neq i^\star} t_i - c_i = \nu^\star$, then the allocation functions are defined either as in condition (i) or as in condition (ii) of Definition 3. Thus, the principal's payoff amounts either to $t_{i^\star}$ or to $\max_{i\in\mathcal{I}} t_i - c_i \geq \nu^\star$, respectively, and is therefore again non-inferior to $\max_{i\in\mathcal{I}} \underline{t}_i$. In summary, we have shown that the principal's payoff is non-inferior to $z^\star = \max_{i\in\mathcal{I}} \underline{t}_i$ in all three cases. As scenario $\boldsymbol{t}\in\mathcal{T}$ was chosen arbitrarily, this reasoning implies that the principal's worst-case payoff is also non-inferior to $z^\star$. The favored-agent mechanism at hand is therefore optimal in (3) by virtue of Proposition 1. □

Proof of Lemma 1. Assume first that $(\boldsymbol{p},\boldsymbol{q})$ is a favored-agent mechanism with favored agent $i^\star\in\mathcal{I}$ and threshold value $\nu^\star\in\mathbb{R}$. Next, fix any agent $i\in\mathcal{I}$ and any type profile $\boldsymbol{t}_{-i}\in\mathcal{T}_{-i}$. If $i\neq i^\star$, then we have either $p_i(t_i,\boldsymbol{t}_{-i})=q_i(t_i,\boldsymbol{t}_{-i})=1$ or $p_i(t_i,\boldsymbol{t}_{-i})=q_i(t_i,\boldsymbol{t}_{-i})=0$ for all $t_i\in\mathcal{T}_i$. This implies that $p_i(t_i,\boldsymbol{t}_{-i})-q_i(t_i,\boldsymbol{t}_{-i})=0$ is constant in $t_i\in\mathcal{T}_i$. If $i=i^\star$, then the fixed type profile $\boldsymbol{t}_{-i^\star}$ uniquely determines whether the allocations are constructed as in case (i) or as in case (ii) of Definition 3. In case (i) we have $p_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})=1$ and $q_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})=0$ for all $t_{i^\star}\in\mathcal{T}_{i^\star}$, and thus $p_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})-q_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})=1$ is constant in $t_{i^\star}\in\mathcal{T}_{i^\star}$. In case (ii) we have either $p_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})=1$ and $q_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})=1$ or $p_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})=0$ and $q_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})=0$, and thus $p_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})-q_{i^\star}(t_{i^\star},\boldsymbol{t}_{-i^\star})=0$ is again constant in $t_{i^\star}\in\mathcal{T}_{i^\star}$. This establishes the claim for any favored-agent mechanism $(\boldsymbol{p},\boldsymbol{q})$. Assume now that $(\boldsymbol{p},\boldsymbol{q})=\sum_{k\in\mathcal{K}}\pi_k(\boldsymbol{p}^k,\boldsymbol{q}^k)$ is a convex combination of favored-agent mechanisms $(\boldsymbol{p}^k,\boldsymbol{q}^k)$, $k\in\mathcal{K}=\{1,\dots,K\}$. Next, fix any $i\in\mathcal{I}$ and $\boldsymbol{t}_{-i}\in\mathcal{T}_{-i}$. From the first part of the proof we know that $p_i^k(t_i,\boldsymbol{t}_{-i})-q_i^k(t_i,\boldsymbol{t}_{-i})$ is constant in $t_i\in\mathcal{T}_i$ for each $k\in\mathcal{K}$, and therefore $p_i(t_i,\boldsymbol{t}_{-i})-q_i(t_i,\boldsymbol{t}_{-i})$ is also constant in $t_i\in\mathcal{T}_i$. Similar arguments apply when $(\boldsymbol{p},\boldsymbol{q})$ represents a convex combination of infinitely many favored-agent mechanisms. □

Proof of Theorem 2. Throughout the proof, we use the following partition of the type space $\mathcal{T}$.

$$
\begin{aligned}
\mathcal{T}_I &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i \leq \underline{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i < t_{i^\star}\}\\
\mathcal{T}_{II} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i \leq \underline{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i \geq t_{i^\star}\}\\
\mathcal{T}_{III} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i > \underline{t}_{i^\star} \text{ and } t_i - c_i \notin (\underline{t}_{i^\star}, t_{i^\star}]\ \forall i\neq i^\star\}\\
\mathcal{T}_{IV} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i > \underline{t}_{i^\star} \text{ and } \exists i\neq i^\star \text{ such that } t_i - c_i \in (\underline{t}_{i^\star}, t_{i^\star}]\}
\end{aligned}
$$

Note that the set $\mathcal{T}_I$ is nonempty and contains at least $(\bar{t}_{i^\star},\underline{\boldsymbol{t}}_{-i^\star})$ since $\max_{i\in\mathcal{I}}\underline{t}_i=\underline{t}_{i^\star}<\bar{t}_{i^\star}$. However, the sets $\mathcal{T}_{II},\mathcal{T}_{III}$ and $\mathcal{T}_{IV}$ can be empty if $\underline{t}_{i^\star}$ or c_i, $i\neq i^\star$, are sufficiently large.

In the following, we denote by $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ the favored-agent mechanism of type (i) with favored agent $i^\star\in\arg\max_{i\in\mathcal{I}}\underline{t}_i$ and threshold value $\nu^\star=\max_{i\in\mathcal{I}}\underline{t}_i$. By construction, we thus have $\nu^\star=\underline{t}_{i^\star}$. Assume now for the sake of contradiction that there exists another mechanism $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$ that Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. Thus, the inequality (5) holds for all $\boldsymbol{t}\in\mathcal{T}$ and is strict for at least one $\boldsymbol{t}\in\mathcal{T}$. Note that the right-hand side of (5) represents the principal's payoff in scenario $\boldsymbol{t}$ under $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. By the definition of a type (i) favored-agent mechanism, this payoff amounts to $t_{i^\star}$ when $\max_{i\neq i^\star} t_i-c_i\leq\underline{t}_{i^\star}$ (*i.e.*, when $\boldsymbol{t}\in\mathcal{T}_I\cup\mathcal{T}_{II}$) and to $\max_{i\in\mathcal{I}} t_i-c_i$ when $\max_{i\neq i^\star} t_i-c_i>\underline{t}_{i^\star}$ (*i.e.*, when $\boldsymbol{t}\in\mathcal{T}_{III}\cup\mathcal{T}_{IV}$). We will show that if (5) holds, then $(\boldsymbol{p},\boldsymbol{q})$ must generate the same payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ under every type profile $\boldsymbol{t}\in\mathcal{T}$. In other words, $(\boldsymbol{p},\boldsymbol{q})$ cannot generate a strictly higher payoff than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ under any type profile, which contradicts our assumption that $(\boldsymbol{p},\boldsymbol{q})$ Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. The remainder of the proof is divided into four steps, each of which investigates one of the subsets $\mathcal{T}_I$, $\mathcal{T}_{II}$, $\mathcal{T}_{III}$ and $\mathcal{T}_{IV}$.

Step 1 ($\mathcal{T}_I$). Consider any type profile $\boldsymbol{t}\in\mathcal{T}_I$. For inequality (5) to hold in this scenario, the principal must earn at least $t_{i^\star}$ under mechanism $(\boldsymbol{p},\boldsymbol{q})$. We next show that this is only possible if $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$. To this end, assume for the sake of contradiction that either $p_{i^\star}(\boldsymbol{t})<1$ or $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})>0$. If $p_{i^\star}(\boldsymbol{t})<1$, then the principal's payoff under $(\boldsymbol{p},\boldsymbol{q})$ satisfies

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\leq\sum_{i\in\mathcal{I}}p_i(\boldsymbol{t})t_i<t_{i^\star}=\sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i-q_i^\star(\boldsymbol{t})c_i),$$

where the strict inequality holds because $\boldsymbol{t} \in \mathcal{T}_I$, which implies that $t_i < t_{i^\star}$ for all $i \neq i^\star$. Thus, inequality (5) is violated in scenario $\boldsymbol{t}$. If $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) > 0$, on the other hand, we have

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) = p_{i^\star}(\boldsymbol{t})t_{i^\star} - q_{i^\star}(\boldsymbol{t})c_{i^\star} < t_{i^\star} = \sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the strict inequality holds because $q_{i^\star}(\boldsymbol{t})$ and $c_{i^\star}$ are positive. Thus, inequality (5) is again violated in scenario $\boldsymbol{t}$. For inequality (5) to hold, we must therefore have $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$. Thus, the allocation probabilities of the mechanisms $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ coincide on $\mathcal{T}_I$.

Step 2 ($\mathcal{T}_{II}$). For inequality (5) to hold in any scenario $\boldsymbol{t} \in \mathcal{T}_{II}$, the principal must earn at least $t_{i^\star}$ under mechanism $(\boldsymbol{p},\boldsymbol{q})$. As in Step 1, we can show that this is only possible if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$. To this end, we partition $\mathcal{T}_{II}$ into the following subsets.

$$\begin{aligned}
\mathcal{T}_{II_1} &= \{\boldsymbol{t} \in \mathcal{T}_{II} : \max_{i\neq i^\star} t_i < \bar{t}_{i^\star}\}\\
\mathcal{T}_{II_2} &= \{\boldsymbol{t} \in \mathcal{T}_{II} : \max_{i\neq i^\star} t_i \geq \bar{t}_{i^\star} \text{ and } t_{i^\star} = \bar{t}_{i^\star}\}\\
\mathcal{T}_{II_3} &= \{\boldsymbol{t} \in \mathcal{T}_{II} : \max_{i\neq i^\star} t_i \geq \bar{t}_{i^\star} \text{ and } t_{i^\star} < \bar{t}_{i^\star}\}
\end{aligned}$$

Note that if $\max_{i\neq i^\star} \bar{t}_i < \underline{t}_{i^\star}$, then $\mathcal{T}_{II}$ as well as its subsets $\mathcal{T}_{II_1}$, $\mathcal{T}_{II_2}$ and $\mathcal{T}_{II_3}$ are all empty. If $\max_{i\neq i^\star} \bar{t}_i \geq \underline{t}_{i^\star}$, on the other hand, then $\mathcal{T}_{II}$ and its subset $\mathcal{T}_{II_1}$ are nonempty. Indeed, $\mathcal{T}_{II_1}$ contains the type profile $\boldsymbol{t}$ defined through $t_i = \min\{\underline{t}_{i^\star}, \bar{t}_i\}$ for all $i \in \mathcal{I}$. To see this, note that $\boldsymbol{t} \in \mathcal{T}$ by the construction of $i^\star$. In addition, we have $\boldsymbol{t} \in \mathcal{T}_{II_1}$ thanks to the assumption $\max_{i\neq i^\star} \bar{t}_i \geq \underline{t}_{i^\star}$, which implies that $\max_{i\neq i^\star} t_i = \underline{t}_{i^\star}$. We now investigate the sets $\mathcal{T}_{II_1}$, $\mathcal{T}_{II_2}$ and $\mathcal{T}_{II_3}$ one by one.

Fix first any type profile $\boldsymbol{t} \in \mathcal{T}_{II_1}$. Incentive compatibility ensures that $p_{i^\star}(\boldsymbol{t}) \geq p_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) - q_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 1$, where the equality holds because $(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_I$ and because we know from Step 1 that $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection in $\mathcal{T}_I$. Consequently, the mechanism $(\boldsymbol{p},\boldsymbol{q})$ can only earn $t_{i^\star}$ in scenario $\boldsymbol{t}$ if $q_{i^\star}(\boldsymbol{t}) = 0$. As $\boldsymbol{t} \in \mathcal{T}_{II_1}$ was chosen arbitrarily, the allocation probabilities $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must therefore coincide throughout $\mathcal{T}_{II_1}$.

Next, we study the subset $\mathcal{T}_{II_2}$. To this end, define the set-valued function $\mathcal{I}(\boldsymbol{t}) = \{i \in \mathcal{I} : t_i \geq \bar{t}_{i^\star}\}$ for $\boldsymbol{t} \in \mathcal{T}_{II_2}$. Note that $|\mathcal{I}(\boldsymbol{t})| \geq 2$ for all $\boldsymbol{t} \in \mathcal{T}_{II_2}$ thanks to the definition of $\mathcal{T}_{II_2}$, which implies that $i^\star \in \mathcal{I}(\boldsymbol{t})$ and $\arg\max_{i\neq i^\star} t_i \subseteq \mathcal{I}(\boldsymbol{t})$. We now prove by induction that the allocation probabilities $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must coincide on $\mathcal{T}_{II_2}^n = \{\boldsymbol{t} \in \mathcal{T}_{II_2} : |\mathcal{I}(\boldsymbol{t})| = n\}$ for all $n \geq 2$.

As for the base step, set $n = 2$ and fix any type profile $\boldsymbol{t} \in \mathcal{T}_{II_2}^2$. Thus, there exists exactly one agent $i^\circ \neq i^\star$ with type $t_{i^\circ} \geq \bar{t}_{i^\star}$. Incentive compatibility dictates that $p_{i^\circ}(\boldsymbol{t}) - q_{i^\circ}(\boldsymbol{t}) \leq p_{i^\circ}(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) = 0$, where the equality holds because $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I$ and because we know from Step 1 that $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection in $\mathcal{T}_I$. We thus have $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$. Inequality (5) further requires the mechanism $(\boldsymbol{p},\boldsymbol{q})$ to earn at least $\bar{t}_{i^\star}$ in scenario $\boldsymbol{t} \in \mathcal{T}_{II_2}^2$. All of this is only possible if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ because $t_{i^\circ} - c_{i^\circ} \leq \underline{t}_{i^\star} < \bar{t}_{i^\star}$ and $t_i < \bar{t}_{i^\star}$ for all $i \in \mathcal{I} \setminus \{i^\circ, i^\star\}$.

As for the induction step, assume that $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ for all $\boldsymbol{t} \in \mathcal{T}_{II_2}^n$ and for some $n \geq 2$, and fix an arbitrary type profile $\boldsymbol{t} \in \mathcal{T}_{II_2}^{n+1}$. Thus, there exist exactly n agents $i \neq i^\star$ with types $t_i \geq \bar{t}_{i^\star}$. For any such agent i, incentive compatibility dictates that $p_i(\boldsymbol{t}) - q_i(\boldsymbol{t}) \leq p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, where the equality follows from the induction hypothesis and the observation that $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_{II_2}^n$. We thus have $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t}) \setminus \{i^\star\}$. Inequality (5) further requires the mechanism $(\boldsymbol{p},\boldsymbol{q})$ to earn at least $\bar{t}_{i^\star}$ in scenario $\boldsymbol{t} \in \mathcal{T}_{II_2}^{n+1}$. In analogy to the base step, all of this is only possible if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ because $t_i - c_i \leq \underline{t}_{i^\star} < \bar{t}_{i^\star}$ for all $i \in \mathcal{I}(\boldsymbol{t}) \setminus \{i^\star\}$ and $t_i < \bar{t}_{i^\star}$ for all $i \in \mathcal{I} \setminus \mathcal{I}(\boldsymbol{t})$. This observation completes the induction step. In summary, the allocation probabilities $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must therefore coincide throughout $\cup_{n\geq 2}\mathcal{T}_{II_2}^n = \mathcal{T}_{II_2}$.

Finally, fix any type profile $\boldsymbol{t} \in \mathcal{T}_{II_3}$. Incentive compatibility ensures that $p_{i^\star}(\boldsymbol{t}) \geq p_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) - q_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 1$, where the equality holds because $(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_{II_2}$ and because we know from the above induction argument that $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection in $\mathcal{T}_{II_2}$.

Hence, the mechanism $(\boldsymbol{p},\boldsymbol{q})$ can only earn $t_{i^\star}$ in scenario $\boldsymbol{t}$ if $q_{i^\star}(\boldsymbol{t})=0$. As $\boldsymbol{t}\in\mathcal{T}_{II_3}$ was chosen arbitrarily, the allocation probabilities $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must therefore coincide throughout $\mathcal{T}_{II_3}$.

Step 3 ($\mathcal{T}_{III}$). In this part of the proof, we will demonstrate that

$$\sum_{i\in\arg\max_{i'\in\mathcal{I}} t_{i'}-c_{i'}} p_i(\boldsymbol{t})=1 \quad \text{and} \quad p_i(\boldsymbol{t})=q_i(\boldsymbol{t}) \quad \forall i\in\arg\max_{i'\in\mathcal{I}} t_{i'}-c_{i'}, \tag{13}$$

for every fixed $\boldsymbol{t}\in\mathcal{T}_{III}$. To prove (13), define the set-valued function $\mathcal{I}(\boldsymbol{t})=\{i\in\mathcal{I}: t_i>t_{i^\star}\}$ for $\boldsymbol{t}\in\mathcal{T}_{III}$. Note that $|\mathcal{I}(\boldsymbol{t})|\geq 1$ for all $\boldsymbol{t}\in\mathcal{T}_{III}$ thanks to the definition of $\mathcal{T}_{III}$, which ensures that there exists at least one agent $i\in\mathcal{I}$ with $t_i-c_i>t_{i^\star}$. We will now use induction to prove that (13) holds for all type profiles in $\mathcal{T}_{III}^n=\{\boldsymbol{t}\in\mathcal{T}_{III}:|\mathcal{I}(\boldsymbol{t})|=n\}$ for all $n\geq 1$.

As for the base step, set $n=1$ and fix any type profile $\boldsymbol{t}\in\mathcal{T}_{III}^1$. Thus, there exists exactly one agent $i^\circ\neq i^\star$ with $t_{i^\circ}>t_{i^\star}$. Incentive compatibility ensures that $p_{i^\circ}(\boldsymbol{t})-q_{i^\circ}(\boldsymbol{t})\leq p_{i^\circ}(\underline{t}_{i^\circ},\boldsymbol{t}_{-i^\circ})=0$, where the equality holds because $(\underline{t}_{i^\circ},\boldsymbol{t}_{-i^\circ})\in\mathcal{T}_I\cup\mathcal{T}_{II}$. We thus have $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$. If $p_{i^\circ}(\boldsymbol{t})<1$, then

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\leq p_{i^\circ}(\boldsymbol{t})(t_{i^\circ}-c_{i^\circ})+\sum_{i\neq i^\circ}p_i(\boldsymbol{t})t_i<\max_{i\in\mathcal{I}} t_i-c_i,$$

where the first inequality holds because $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$ and $c_i>0$ for all $i\neq i^\circ$. The second inequality follows from the assumption that $p_{i^\circ}(\boldsymbol{t})<1$ as well as the definition of $\mathcal{T}_{III}$ and the construction of i°, which imply that $t_{i^\circ}-c_{i^\circ}=\max_{i\in\mathcal{I}} t_i-c_i>t_{i^\star}$ and $t_{i^\star}\geq t_i$ for all $i\neq i^\circ$. This shows that $(\boldsymbol{p},\boldsymbol{q})$ earns strictly less than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in scenario $\boldsymbol{t}$, which contradicts inequality (5). Hence, our assumption must have been wrong, and $p_{i^\circ}(\boldsymbol{t})$ must equal 1. We have thus established (13) in scenario $\boldsymbol{t}$.

As for the induction step, assume that (13) holds throughout $\mathcal{T}_{III}^n$ for some $n\geq 1$, and fix any type profile $\boldsymbol{t}\in\mathcal{T}_{III}^{n+1}$. Thus, there exist exactly $n+1$ agents $i\neq i^\star$ with types $t_i>t_{i^\star}$. For any agent $i\in\mathcal{I}(\boldsymbol{t})$ incentive compatibility dictates that $p_i(\boldsymbol{t})-q_i(\boldsymbol{t})\leq p_i(\underline{t}_i,\boldsymbol{t}_{-i})=0$, where the equality holds because $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}^n$. Indeed, if $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_I\cup\mathcal{T}_{II}$, then the equality follows from the results of Steps 1 and 2, and if $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_{III}^n$, then the equality follows from the induction hypothesis. We thus have $p_i(\boldsymbol{t})=q_i(\boldsymbol{t})$ for all $i\in\mathcal{I}(\boldsymbol{t})$ and, by the definition of $\mathcal{T}_{III}$, in particular for all $i\in\arg\max_{j\in\mathcal{I}} t_j-c_j$. In addition, if the summation of $p_i(\boldsymbol{t})$ over all $i\in\arg\max_{i'\in\mathcal{I}} t_{i'}-c_{i'}$ is strictly smaller than 1, then

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\leq\sum_{i\in\mathcal{I}(\boldsymbol{t})}p_i(\boldsymbol{t})(t_i-c_i)+\sum_{i\notin\mathcal{I}(\boldsymbol{t})}p_i(\boldsymbol{t})t_i<\max_{i\in\mathcal{I}} t_i-c_i$$

where the first inequality holds because $p_i(\boldsymbol{t})=q_i(\boldsymbol{t})$ for all $i\in\mathcal{I}(\boldsymbol{t})$ and $c_i>0$ for all $i\notin\mathcal{I}(\boldsymbol{t})$. The strict inequality holds because $\sum_{i\in\mathcal{I}}p_i(\boldsymbol{t})\leq 1$ and $\max_{j\in\mathcal{I}} t_j-c_j>t_{i^\star}\geq t_i$ for all $i\notin\mathcal{I}(\boldsymbol{t})$ by the definition of $\mathcal{T}_{III}$ and because we assumed that the summation of $p_i(\boldsymbol{t})$ over $i\in\arg\max_{i'\in\mathcal{I}} t_{i'}-c_{i'}$ is strictly smaller than 1. This reasoning shows that $(\boldsymbol{p},\boldsymbol{q})$ earns strictly less than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in scenario $\boldsymbol{t}$, which contradicts inequality (5). Hence, our assumption must be false, and the summation of $p_i(\boldsymbol{t})$ over $i\in\arg\max_{i'\in\mathcal{I}} t_{i'}-c_{i'}$ equals 1. We have thus established (13) in scenario $\boldsymbol{t}$. As $\boldsymbol{t}\in\mathcal{T}_{III}^{n+1}$ was chosen arbitrarily, we may conclude that (13) holds throughout $\mathcal{T}_{III}^{n+1}$. This observation completes the induction step. In summary, the revenues generated by the mechanisms $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must therefore coincide throughout $\cup_{n\geq 1}\mathcal{T}_{III}^n=\mathcal{T}_{III}$.

Step 4 ($\mathcal{T}_{IV}$). In analogy to Step 3, we will show that (13) holds for every fixed $\boldsymbol{t}\in\mathcal{T}_{IV}$. This immediately implies that $(\boldsymbol{p},\boldsymbol{q})$ generates the same payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ throughout $\mathcal{T}_{IV}$. To prove (13), define the set-valued function $\mathcal{I}(\boldsymbol{t})=\{i\in\mathcal{I}: t_i>\underline{t}_{i^\star}\}$ for $\boldsymbol{t}\in\mathcal{T}_{IV}$. Note that $|\mathcal{I}(\boldsymbol{t})|\geq 2$ for all $\boldsymbol{t}\in\mathcal{T}_{IV}$ thanks to the definition of $\mathcal{T}_{IV}$, which implies that $i^\star\in\mathcal{I}(\boldsymbol{t})$ and $\arg\max_{i\neq i^\star} t_i-c_i\subseteq\mathcal{I}(\boldsymbol{t})$. To see that $i^\star\in\mathcal{I}(\boldsymbol{t})$, note that if $i^\star\notin\mathcal{I}(\boldsymbol{t})$ for some $\boldsymbol{t}\in\mathcal{T}_{IV}$, then $t_{i^\star}=\underline{t}_{i^\star}$, and there can be no $i\neq i^\star$ with

$t_i - c_i \in (\underline{t}_{i^\star}, t_{i^\star}] = \emptyset$, which contradicts the assumption that $\boldsymbol{t} \in \mathcal{T}_{IV}$. We will now use induction to prove that (13) holds for all type profiles in $\mathcal{T}_{IV}^n = \{\boldsymbol{t} \in \mathcal{T}_{IV} : |\mathcal{I}(\boldsymbol{t})| = n\}$ for all $n \geq 2$.

As for the base step, set $n = 2$ and fix an arbitrary type profile $\boldsymbol{t} \in \mathcal{T}_{IV}^2$. Thus, there exists exactly one agent $i^\circ \neq i^\star$ with $t_{i^\circ} > \underline{t}_{i^\star}$. Incentive compatibility for agent $i^\star$ ensures that $p_{i^\star}(\boldsymbol{t}) - q_{i^\star}(\boldsymbol{t}) \leq p_{i^\star}(\underline{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 0$, where the equality follows from (13) and the observation that $(\underline{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_{III}$. Thus, we have $p_{i^\star}(\boldsymbol{t}) = q_{i^\star}(\boldsymbol{t})$. Incentive compatibility for agent i° further dictates that $p_{i^\circ}(\boldsymbol{t}) - q_{i^\circ}(\boldsymbol{t}) \leq p_{i^\circ}(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) = 0$, where the equality holds because $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I \cup \mathcal{T}_{II}$. Indeed, recall that the allocation probabilities of $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ match and that the good is allocated to agent $i^\star$ on $\mathcal{T}_I \cup \mathcal{T}_{II}$. Thus, we have $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$. This reasoning shows that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})$. Assume now that the summation of $p_i(\boldsymbol{t})$ over all $i \in \arg\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}$ is strictly smaller than 1. Then, we have

$$\sum_{i \in \mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \leq \sum_{i \in \mathcal{I}(\boldsymbol{t})} p_i(\boldsymbol{t})(t_i - c_i) + \sum_{i \notin \mathcal{I}(\boldsymbol{t})} p_i(\boldsymbol{t}) t_i < \max_{i \in \mathcal{I}} t_i - c_i,$$

where the first inequality holds because $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})$ and $c_i > 0$ for all $i \notin \mathcal{I}(\boldsymbol{t})$. The strict inequality holds because $\sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1$, $\max_{j \in \mathcal{I}} t_j - c_j > \underline{t}_{i^\star} \geq t_i$ for all $i \notin \mathcal{I}(\boldsymbol{t})$ by the definition of $\mathcal{T}_{IV}$ and because we assumed that the summation of $p_i(\boldsymbol{t})$ over $i \in \arg\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}$ is strictly smaller than 1. Hence, $(\boldsymbol{p}, \boldsymbol{q})$ earns strictly less than $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in scenario $\boldsymbol{t}$, which contradicts inequality (5). This implies that our assumption was false and that the summation of $p_i(\boldsymbol{t})$ over all $i \in \arg\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}$ must be equal to 1. We have thus established (13) in scenario $\boldsymbol{t}$. As $\boldsymbol{t} \in \mathcal{T}_{IV}^2$ was chosen arbitrarily, (13) holds throughout $\mathcal{T}_{IV}^2$.

As for the induction step, assume that (13) holds throughout $\mathcal{T}_{IV}^n$ for some $n \geq 2$, and fix an arbitrary type profile $\boldsymbol{t} \in \mathcal{T}_{IV}^{n+1}$. Thus, there exist exactly n agents $i \neq i^\star$ with types $t_i > \underline{t}_{i^\star}$. Using the exact same reasoning as in the base step, we can prove that $p_{i^\star}(\boldsymbol{t}) = q_{i^\star}(\boldsymbol{t})$. In addition, for any agent $i \in \mathcal{I}(\boldsymbol{t}) \setminus \{i^\star\}$ incentive compatibility dictates that $p_i(\boldsymbol{t}) - q_i(\boldsymbol{t}) \leq p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, where the equality holds because $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}_{IV}^n$. Indeed, if $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$, then the equality follows from the results of Steps 1, 2 and 3, and if $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_{IV}^n$, then the equality follows from the induction hypothesis. In summary, we have thus shown that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})$. The first statement in (13) can be proved by repeating the corresponding arguments from the base step almost verbatim. Details are omitted for brevity. We have thus established (13) in an arbitrary scenario $\boldsymbol{t} \in \mathcal{T}_{IV}^{n+1}$. By induction, the revenues generated by the mechanisms $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ must therefore coincide throughout $\cup_{n \geq 2} \mathcal{T}_{IV}^n = \mathcal{T}_{IV}$. This observation completes the proof. □

Proof of Proposition 2. Relaxing the incentive compatibility constraints and the first inequality in (FC) yields

$$\begin{aligned} z^\star \leq \quad & \sup_{\boldsymbol{p}, \boldsymbol{q}} \quad \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\sum_{i \in \mathcal{I}} (p_i(\tilde{\boldsymbol{t}}) \tilde{t}_i - q_i(\tilde{\boldsymbol{t}}) c_i) \right] \\ & \text{s.t.} \quad p_i : \mathcal{T} \to [0,1] \text{ and } q_i : \mathcal{T} \to [0,1] \;\; \forall i \in \mathcal{I} \\ & \qquad\; \sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1 \;\; \forall \boldsymbol{t} \in \mathcal{T} \\ = \quad & \sup_{\boldsymbol{p}} \quad \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\sum_{i \in \mathcal{I}} p_i(\tilde{\boldsymbol{t}}) \tilde{t}_i \right] \\ & \text{s.t.} \quad p_i : \mathcal{T} \to [0,1] \;\; \forall i \in \mathcal{I}, \;\; \sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1 \;\; \forall \boldsymbol{t} \in \mathcal{T}, \end{aligned}$$

where the equality holds because it is optimal to set $q_i(\boldsymbol{t}) = 0$ for all $i \in \mathcal{I}$ and $\boldsymbol{t} \in \mathcal{T}$ in the relaxed problem. As $p_i \geq 0$ and $\sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1$ for all $\boldsymbol{t} \in \mathcal{T}$, we further have

$$\sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) t_i \leq \max_{i \in \mathcal{I}} t_i \;\; \forall \boldsymbol{t} \in \mathcal{T},$$

which imply that $z^\star$ is bounded above by $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\max_{i\in\mathcal{I}} \tilde{t}_i\right]$. Now, select an arbitrary $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and denote by $\delta_{\underline{\boldsymbol{\mu}}}$ the Dirac point mass at $\underline{\boldsymbol{\mu}}$. We have

$$\mathbb{E}_{\delta_{\underline{\boldsymbol{\mu}}}}\left[\max_{i\in\mathcal{I}} \tilde{t}_i\right] \geq \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\max_{i\in\mathcal{I}} \tilde{t}_i\right] \geq \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\tilde{t}_{i^\star}\right] = \max_{i\in\mathcal{I}} \underline{\mu}_i,$$

where the first inequality holds because $\delta_{\underline{\boldsymbol{\mu}}} \in \mathcal{P}$, the second inequality holds because $\max_{i\in\mathcal{I}} t_i \geq t_{i^\star}$ for any $\boldsymbol{t}\in\mathcal{T}$, and the equality follows from the selection of $i^\star$ and the definition of the Markov ambiguity set $\mathcal{P}$. As $\delta_{\underline{\boldsymbol{\mu}}}$ is the Dirac point mass at $\underline{\boldsymbol{\mu}}$, we also have $\mathbb{E}_{\delta_{\underline{\boldsymbol{\mu}}}}\left[\max_{i\in\mathcal{I}} \tilde{t}_i\right] = \max_{i\in\mathcal{I}} \underline{\mu}_i$ that implies $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\max_{i\in\mathcal{I}} \tilde{t}_i\right] = \max_{i\in\mathcal{I}} \underline{\mu}_i$. Therefore, the optimal value $z^\star$ is bounded above by $\max_{i\in\mathcal{I}} \underline{\mu}_i$. However, this bound is attained by a mechanism that allocates the good to an agent $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ irrespective of $\boldsymbol{t}\in\mathcal{T}$ and never inspects anyone's type. Since this mechanism is feasible, the claim follows. □

Proof of Theorem 3. Select an arbitrary favored-agent mechanism with $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and $\nu^\star \geq \bar{t}_{i^\star}$. Recall first that this mechanism is feasible in (MDP). Next, we will show that this mechanism attains a worst-case payoff that is at least as large as $\max_{i\in\mathcal{I}} \underline{\mu}_i$, which implies via Proposition 2 that this mechanism is optimal in (MDP). To this end, fix an arbitrary type profile $\boldsymbol{t}\in\mathcal{T}$. If $\max_{i\in\mathcal{I}} t_i - c_i < \nu^\star$, then condition (i) in Definition 3 implies that the principal's payoff amounts to $t_{i^\star}$. If $\max_{i\in\mathcal{I}} t_i - c_i > \nu^\star$, then condition (ii) in Definition 3 implies that the principal's payoff amounts to $\max_{i\in\mathcal{I}} t_i - c_i > \nu^\star \geq t_{i^\star}$, where the second inequality follows from the selection of $\nu^\star$. If $\max_{i\neq i^\star} t_i - c_i = \nu^\star$, then the allocation functions are defined either as in condition (i) or as in condition (ii) of Definition 3. Thus, the principal's payoff amounts either to $t_{i^\star}$ or to $\max_{i\in\mathcal{I}} t_i - c_i \geq \nu^\star \geq t_{i^\star}$, respectively. In summary, we have shown that the principal's payoff is bigger than or equal to $t_{i^\star}$ in all three cases. As the type profile $\boldsymbol{t}$ was chosen arbitrarily, this implies that the principal's expected payoff under any distribution $\mathbb{P}\in\mathcal{P}$ is bounded below by $\mathbb{E}_{\mathbb{P}}\left[\tilde{t}_{i^\star}\right]$. By the definition of the Markov ambiguity set $\mathcal{P}$, the expectation $\mathbb{E}_{\mathbb{P}}\left[\tilde{t}_{i^\star}\right]$ cannot be lower than $z^\star = \max_{i\in\mathcal{I}} \underline{\mu}_i$ for any $\mathbb{P}\in\mathcal{P}$. Therefore, the principal's worst-case expected payoff under the favored-agent mechanism is bounded below by $z^\star$. The favored-agent mechanism at hand is therefore optimal in (3) by virtue of Proposition 2. □

Proof of Lemma 2. For any $\boldsymbol{t}\in\mathcal{T}$, we will show that there exists a scenario $\hat{\boldsymbol{t}}\in\mathcal{T}$ that satisfies $\max_{i\neq i^\star} \hat{t}_i < \hat{t}_{i^\star}$ and $\alpha\boldsymbol{t} + (1-\alpha)\hat{\boldsymbol{t}} = \underline{\boldsymbol{\mu}}$ for some $\alpha\in(0,1]$. This implies that the discrete distribution $\mathbb{P} = \alpha\delta_{\boldsymbol{t}} + (1-\alpha)\delta_{\hat{\boldsymbol{t}}}$ belongs to the Markov ambiguity set $\mathcal{P}$ and moreover satisfies the properties (i)–(iii).

To this end, consider any $\boldsymbol{t}\in\mathcal{T}$. If $\boldsymbol{t} = \underline{\boldsymbol{\mu}}$, set $\hat{\boldsymbol{t}} = \boldsymbol{t} = \underline{\boldsymbol{\mu}}$. As $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, scenario $\hat{\boldsymbol{t}}$ satisfies $\max_{i\neq i^\star} \hat{t}_i < \hat{t}_{i^\star}$. Moreover, note that $\alpha\boldsymbol{t} + (1-\alpha)\hat{\boldsymbol{t}} = \underline{\boldsymbol{\mu}}$ for any $\alpha\in(0,1]$. Similarly, for any $\alpha\in(0,1]$, $\mathbb{P} = \alpha\delta_{\boldsymbol{t}} + (1-\alpha)\delta_{\hat{\boldsymbol{t}}} = \delta_{\underline{\boldsymbol{\mu}}}$ is the Dirac point mass at $\underline{\boldsymbol{\mu}}$ and trivially satisfies the desired properties (i)–(iii).

If $\boldsymbol{t} \neq \underline{\boldsymbol{\mu}}$, define function $\hat{\boldsymbol{t}}(\alpha)$ through

$$\hat{\boldsymbol{t}}(\alpha) = \frac{1}{1-\alpha}(\underline{\boldsymbol{\mu}} - \boldsymbol{t}) + \boldsymbol{t}.$$

Note that, for any $\alpha\in[0,1)$, $\hat{\boldsymbol{t}}(\alpha)$ satisfies

$$\alpha\boldsymbol{t} + (1-\alpha)\hat{\boldsymbol{t}}(\alpha) = \alpha\boldsymbol{t} + (1-\alpha)\left(\frac{1}{1-\alpha}(\underline{\boldsymbol{\mu}} - \boldsymbol{t}) + \boldsymbol{t}\right) = \underline{\boldsymbol{\mu}}.$$

Thus, for any $\alpha\in[0,1)$, $\hat{\boldsymbol{t}} = \hat{\boldsymbol{t}}(\alpha)$ satisfies $\alpha\boldsymbol{t} + (1-\alpha)\hat{\boldsymbol{t}} = \underline{\boldsymbol{\mu}}$. We will next show that there exists an $\alpha\in(0,1)$ for which $\hat{\boldsymbol{t}} = \hat{\boldsymbol{t}}(\alpha)$ also satisfies $\max_{i\neq i^\star} \hat{t}_i < \hat{t}_{i^\star}$. To this end, first note that $\hat{\boldsymbol{t}}(\alpha)$ is a continuous function of $\alpha \in [0,1)$ and $\hat{\boldsymbol{t}}(0) = \underline{\boldsymbol{\mu}}$. Thus, for any $\varepsilon > 0$, there exists $\alpha \in (0,1)$ such that $\hat{\boldsymbol{t}}(\alpha) \in \prod_{i\in\mathcal{I}}[\underline{\mu}_i - \varepsilon, \underline{\mu}_i + \varepsilon]$. We next show that any $\varepsilon > 0$ that belongs to the set

$$L = (0, \min_{i\in\mathcal{I}} \underline{\mu}_i - \underline{t}_i) \cap (0, \min_{i\in\mathcal{I}} \bar{t}_i - \underline{\mu}_i) \cap \big(0, (\underline{\mu}_{i^\star} - \max_{i\neq i^\star} \underline{\mu}_i)/2\big)$$

ensures that $\prod_{i\in\mathcal{I}}[\underline{\mu}_i - \varepsilon, \underline{\mu}_i + \varepsilon] \subseteq \{\boldsymbol{t} \in \mathcal{T} : \max_{i\neq i^\star} t_i < t_{i^\star}\}$. Note that set L is non-empty because $\underline{t}_i < \underline{\mu}_i < \overline{\mu}_i < \bar{t}_i$ for all $i \in \mathcal{I}$ and $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton. Consider any $\varepsilon \in L$. As $\varepsilon < \min_{i\in\mathcal{I}} \underline{\mu}_i - \underline{t}_i$, any $\boldsymbol{t} \in \prod_{i\in\mathcal{I}}[\underline{\mu}_i - \varepsilon, \underline{\mu}_i + \varepsilon]$ satisfies

$$t_i \geq \underline{\mu}_i - \varepsilon > \underline{\mu}_i - (\min_{j\in\mathcal{I}} \underline{\mu}_j - \underline{t}_j) \geq \underline{\mu}_i - (\underline{\mu}_i - \underline{t}_i) = \underline{t}_i \quad \forall i \in \mathcal{I}.$$

Similarly, as $\varepsilon < \min_{i\in\mathcal{I}} \bar{t}_i - \underline{\mu}_i$, any $\boldsymbol{t} \in \prod_{i\in\mathcal{I}}[\underline{\mu}_i - \varepsilon, \underline{\mu}_i + \varepsilon]$ satisfies

$$t_i \leq \underline{\mu}_i + \varepsilon < \underline{\mu}_i + (\min_{j\in\mathcal{I}} \bar{t}_j - \underline{\mu}_j) \leq \underline{\mu}_i + \bar{t}_i - \underline{\mu}_i = \bar{t}_i \quad \forall i \in \mathcal{I}.$$

Therefore, we have shown that $\prod_{i\in\mathcal{I}}[\underline{\mu}_i - \varepsilon, \underline{\mu}_i + \varepsilon] \subseteq \mathcal{T}$. Finally, any $\boldsymbol{t} \in \prod_{i\in\mathcal{I}}[\underline{\mu}_i - \varepsilon, \underline{\mu}_i + \varepsilon]$ satisfies

$$\begin{aligned} t_{i^\star} &\geq \underline{\mu}_{i^\star} - \varepsilon > \underline{\mu}_{i^\star} - (\underline{\mu}_{i^\star} - \max_{j\neq i^\star} \underline{\mu}_j)/2 = (\underline{\mu}_{i^\star} + \max_{j\neq i^\star} \underline{\mu}_j)/2 \\ &= \max_{j\neq i^\star} \underline{\mu}_j + (\underline{\mu}_{i^\star} - \max_{j\neq i^\star} \underline{\mu}_j)/2 > \max_{j\neq i^\star} \underline{\mu}_j + \varepsilon \geq \underline{\mu}_i + \varepsilon \geq t_i \quad \forall i \neq i^\star, \end{aligned}$$

where the second and third inequalities follow from $\varepsilon < (\underline{\mu}_{i^\star} - \max_{i\neq i^\star} \underline{\mu}_i)/2$. Thus, we have shown that $\prod_{i\in\mathcal{I}}[\underline{\mu}_i - \varepsilon, \underline{\mu}_i + \varepsilon] \subseteq \{\boldsymbol{t} \in \mathcal{T} : \max_{i\neq i^\star} t_i < t_{i^\star}\}$ for any $\varepsilon \in L$. As for any $\varepsilon \in L$ there exists $\alpha \in (0,1)$ such that $\hat{\boldsymbol{t}}(\alpha) \in \prod_{i\in\mathcal{I}}[\underline{\mu}_i - \varepsilon, \underline{\mu}_i + \varepsilon]$, the claim follows. □

Proof of Lemma 3. Consider an arbitrary mechanism $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$. If $(\boldsymbol{p}, \boldsymbol{q})$ satisfies (7), then the principal's expected payoff $\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\right]$ under any distribution $\mathbb{P} \in \mathcal{P}$ is at least $\mathbb{E}_{\mathbb{P}}\left[\tilde{t}_{i^\star}\right] \geq \max_{i\in\mathcal{I}} \underline{\mu}_i$, where the inequality follows from the definition of the Markov ambiguity set $\mathcal{P}$. By virtue of Proposition 2, this mechanism is therefore optimal (MDP). We thus have shown that if $(\boldsymbol{p}, \boldsymbol{q})$ satisfies (7), then it is optimal in (MDP).

We next show that if $(\boldsymbol{p}, \boldsymbol{q})$ is optimal in (MDP), then it must satisfy (7). To this end, assume for the sake of contradiction that $(\boldsymbol{p}, \boldsymbol{q})$ is optimal and $\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) < t_{i^\star}$ for some $\boldsymbol{t} \in \mathcal{T}$. Consider an arbitrary $\boldsymbol{t} \in \mathcal{T}$ for which inequality (7) fails. By Lemma 2, we know that there exist a scenario $\hat{\boldsymbol{t}} \in \mathcal{T}$, where $\max_{i\neq i^\star} \hat{t}_i < \hat{t}_{i^\star}$, and a discrete distribution $\mathbb{P} \in \mathcal{P}$ that satisfy the following properties: (i) $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] = \underline{\mu}_i \;\; \forall i \in \mathcal{I}$, (ii) $\mathbb{P}(\tilde{\boldsymbol{t}} \in \{\boldsymbol{t}, \hat{\boldsymbol{t}}\}) = 1$, (iii) $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. The principal's payoff $\sum_{i\in\mathcal{I}}(p_i(\hat{\boldsymbol{t}})\hat{t}_i - q_i(\hat{\boldsymbol{t}})c_i)$ in scenario $\hat{\boldsymbol{t}}$ is bounded above by $\sum_{i\in\mathcal{I}} p_i(\hat{\boldsymbol{t}})\hat{t}_i \leq \hat{t}_{i^\star}$, where the inequality holds because $\sum_{i\in\mathcal{I}} p_i(\hat{\boldsymbol{t}}) \leq 1$ and $\hat{t}_i \leq \hat{t}_{i^\star}$ for all $i \in \mathcal{I}$. The principal's expected payoff under $\mathbb{P}$ therefore satisfies

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\right] &= \mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) \sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) + \mathbb{P}(\tilde{\boldsymbol{t}} = \hat{\boldsymbol{t}}) \sum_{i\in\mathcal{I}}(p_i(\hat{\boldsymbol{t}})\hat{t}_i - q_i(\hat{\boldsymbol{t}})c_i) \\ &< \mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t})t_{i^\star} + \mathbb{P}(\tilde{\boldsymbol{t}} = \hat{\boldsymbol{t}})\hat{t}_{i^\star} = \underline{\mu}_{i^\star}, \end{aligned}$$

where the first equality follows from property (ii), the inequality holds because of property (iii) and because we have assumed that $\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) < t_{i^\star}$ and we have shown that $\sum_{i\in\mathcal{I}}(p_i(\hat{\boldsymbol{t}})\hat{t}_i - q_i(\hat{\boldsymbol{t}})c_i) \leq \hat{t}_{i^\star}$, and the last equality follows from properties (i) and (ii). As the principal's expected payoff under $\mathbb{P}$ is strictly smaller than $z^\star = \underline{\mu}_{i^\star}$, mechanism $(\boldsymbol{p}, \boldsymbol{q})$ cannot be optimal. The claim thus follows. □

Proof of Lemma 4. Throughout the proof, we use the following partition of the type space $\mathcal{T}$.

$$\begin{aligned}
\mathcal{T}_I &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i < \underline{t}_{i^\star}\}\\
\mathcal{T}_{II} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i \geq \underline{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i < \bar{t}_{i^\star}\}\\
\mathcal{T}_{III} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i \geq \underline{t}_{i^\star},\ \max_{i\neq i^\star} t_i \geq \bar{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i - c_i < \underline{t}_{i^\star}\}\\
\mathcal{T}_{IV} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i \geq \underline{t}_{i^\star},\ \max_{i\neq i^\star} t_i \geq \bar{t}_{i^\star},\ \max_{i\neq i^\star} t_i - c_i \geq \underline{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star}\}\\
\mathcal{T}_V &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i \geq \underline{t}_{i^\star},\ \max_{i\neq i^\star} t_i \geq \bar{t}_{i^\star},\ \max_{i\neq i^\star} t_i - c_i \geq \underline{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i - c_i \geq \bar{t}_{i^\star}\}
\end{aligned} \tag{14}$$

Note again that some of the conditions in the definitions above are redundant and introduced for ease of readability. Note also that the set $\mathcal{T}_I$ is nonempty and contains at least $\underline{\boldsymbol{\mu}} = (\underline{\mu}_1, \ldots, \underline{\mu}_I)$ because $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton. However, the sets $\mathcal{T}_{II}, \mathcal{T}_{III}, \mathcal{T}_{IV}$ and $\mathcal{T}_V$ can be empty if $\underline{t}_{i^\star}$ or c_i, $i\neq i^\star$, are sufficiently large.

In the following, we will use Lemma 3 that shows that any optimal mechanism should satisfy (7). In other words, any optimal mechanism should earn a payoff that is at least $\underline{t}_{i^\star}$ in any scenario $\boldsymbol{t}\in\mathcal{T}$. To prove the claim, we will show that if a feasible mechanism $(\boldsymbol{p},\boldsymbol{q})$ violates $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ for some $\boldsymbol{t}\in\mathcal{T}$ such that $\max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star}$, then it cannot satisfy (7). Consequently, mechanism $(\boldsymbol{p},\boldsymbol{q})$ cannot be optimal. The remainder of the proof is divided into four steps, each of which investigates one of the subsets $\mathcal{T}_I$, $\mathcal{T}_{II}$, $\mathcal{T}_{III}$ and $\mathcal{T}_{IV}$. We have $\max_{i\neq i^\star} t_i - c_i \geq \bar{t}_{i^\star}$ for any $\boldsymbol{t}\in\mathcal{T}_V$, and for this reason we do not need to investigate this set.

Step 1 ($\mathcal{T}_I$). Assume for the sake of contradiction that a mechanism $(\boldsymbol{p},\boldsymbol{q})$ is optimal in (MDP) and satisfy $p_{i^\star}(\boldsymbol{t}) < 1$ or $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) > 0$ in some scenario $\boldsymbol{t}\in\mathcal{T}_I$. If $p_{i^\star}(\boldsymbol{t}) < 1$, then the principal's payoff can be written as

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq \sum_{i\in\mathcal{I}} p_i(\boldsymbol{t})t_i < \underline{t}_{i^\star},$$

where the strict inequality holds because $\sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \leq 1$ and $\boldsymbol{t}\in\mathcal{T}_I$, which implies that $t_i < \underline{t}_{i^\star}$ for all $i\neq i^\star$. Thus, inequality (7) is violated in scenario $\boldsymbol{t}$. If $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) > 0$, on the other hand, we have

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) = p_{i^\star}(\boldsymbol{t})t_{i^\star} - q_{i^\star}(\boldsymbol{t})c_{i^\star} < t_{i^\star},$$

where the strict inequality holds because $q_{i^\star}(\boldsymbol{t})$ and $c_{i^\star}$ are positive. Thus, inequality (7) is again violated in scenario $\boldsymbol{t}$. For inequality (7) to hold, we must therefore have $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ for any $\boldsymbol{t}\in\mathcal{T}_I$.

Step 2 ($\mathcal{T}_{II}$). Consider any type profile $\boldsymbol{t}\in\mathcal{T}_{II}$. Incentive compatibility ensures that $p_{i^\star}(\boldsymbol{t}) \geq p_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) - q_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 1$, where the equality holds because $(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_I$ and because we know from Step 1 that $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection in $\mathcal{T}_I$. Consequently, a feasible mechanism $(\boldsymbol{p},\boldsymbol{q})$ can earn at least $\underline{t}_{i^\star}$ in scenario $\boldsymbol{t}$ only if $q_{i^\star}(\boldsymbol{t}) = 0$. As $\boldsymbol{t}\in\mathcal{T}_{II}$ was chosen arbitrarily, any optimal mechanism $(\boldsymbol{p},\boldsymbol{q})$ should satisfy $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ throughout $\mathcal{T}_{II}$.

Step 3 ($\mathcal{T}_{III}$). Define the set-valued function $\mathcal{I}(\boldsymbol{t}) = \{i\in\mathcal{I} : t_i \geq \underline{t}_{i^\star}\}$ for $\boldsymbol{t}\in\mathcal{T}_{III}$. Note that $|\mathcal{I}(\boldsymbol{t})| \geq 2$ for all $\boldsymbol{t}\in\mathcal{T}_{III}$ because $i^\star \in \mathcal{I}(\boldsymbol{t})$ and because the definition of $\mathcal{T}_{III}$ ensures that $\max_{i\neq i^\star} t_i \geq \underline{t}_{i^\star}$. We now prove by induction that $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ for all type profiles in $\mathcal{T}_{III}^n = \{\boldsymbol{t}\in\mathcal{T}_{III} : |\mathcal{I}(\boldsymbol{t})| = n\}$ for all $n\geq 2$.

As for the base step, set $n = 2$ and fix any $\boldsymbol{t}\in\mathcal{T}_{III}^2$. Thus, there exists exactly one agent $i^\circ \neq i^\star$ with $t_{i^\circ} \geq \underline{t}_{i^\star}$. Incentive compatibility ensures that $p_{i^\circ}(\boldsymbol{t}) - q_{i^\circ}(\boldsymbol{t}) \leq p_{i^\circ}(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) = 0$, where the equality holds because $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I \cup \mathcal{T}_{II}$ and because we know from Step 1 and 2 that $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection in $\mathcal{T}_I \cup \mathcal{T}_{II}$. We thus have $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$. As $t_{i^\circ} - c_{i^\circ} < \underline{t}_{i^\star}$ and $t_j < \underline{t}_{i^\star}$ for all $j\in\mathcal{I}\setminus\mathcal{I}(\boldsymbol{t})$, the mechanism $(\boldsymbol{p},\boldsymbol{q})$ can satisfy the inequality (7) for $\boldsymbol{t}\in\mathcal{T}_{III}^2$ only if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$.

As for the induction step, assume that $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$ for all $\boldsymbol{t}\in\mathcal{T}_{III}^n$ and for some $n\geq 2$, and fix an arbitrary type profile $\boldsymbol{t}\in\mathcal{T}_{III}^{n+1}$. Thus, there exist exactly n agents $i\neq i^\star$ with types $t_i\geq t_{i^\star}$. For any such agent i, incentive compatibility dictates that $p_i(\boldsymbol{t})-q_i(\boldsymbol{t})\leq p_i(\underline{t}_i,\boldsymbol{t}_{-i})=0$, where the equality holds because $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}^n$. Indeed, if $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_I\cup\mathcal{T}_{II}$, then the equality follows from the results of Steps 1 and 2, and if $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_{III}^n$, then the equality follows from the induction hypothesis. We thus have $p_i(\boldsymbol{t})=q_i(\boldsymbol{t})$ for all $i\in\mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}$. In analogy to the base step, a feasible mechanism $(\boldsymbol{p},\boldsymbol{q})$ can satisfy the inequality (7) for $\boldsymbol{t}\in\mathcal{T}_{III}^{n+1}$ only if $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$ because $t_i-c_i<t_{i^\star}$ for all $i\in\mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}$, and $t_j<t_{i^\star}$ for $j\in\mathcal{I}\setminus\mathcal{I}(\boldsymbol{t})$. This observation completes the induction step. In summary, the allocation probabilities of any optimal mechanism $(\boldsymbol{p},\boldsymbol{q})$ should satisfy $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$ throughout $\cup_{n\geq 2}\mathcal{T}_{III}^n=\mathcal{T}_{III}$.

Step 4 ($\mathcal{T}_{IV}$). Fix now any arbitrary type profile $\boldsymbol{t}\in\mathcal{T}_{IV}$. Incentive compatibility ensures that $p_{i^\star}(\boldsymbol{t})\geq p_{i^\star}(\bar{t}_{i^\star},\boldsymbol{t}_{-i^\star})-q_{i^\star}(\bar{t}_{i^\star},\boldsymbol{t}_{-i^\star})=1$, where the equality holds because $(\bar{t}_{i^\star},\boldsymbol{t}_{-i^\star})\in\mathcal{T}_{III}$ and because we know from Step 3 that any optimal mechanism $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection in $\mathcal{T}_{III}$. Consequently, a feasible mechanism $(\boldsymbol{p},\boldsymbol{q})$ can earn at least $t_{i^\star}$ in scenario $\boldsymbol{t}$ only if $q_{i^\star}(\boldsymbol{t})=0$. As $\boldsymbol{t}\in\mathcal{T}_{IV}$ was chosen arbitrarily, any optimal mechanism $(\boldsymbol{p},\boldsymbol{q})$ should satisfy $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$ throughout $\mathcal{T}_{IV}$. This observation completes the proof. □

Proof of Lemma 5. We will again use the partition $\mathcal{T}_I-\mathcal{T}_V$ given in (14). Similarly to the sketch of the proof idea, we first show that $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ generate the same payoff throughout $\mathcal{T}\setminus\mathcal{T}_V$. Then, we will prove that the two mechanisms generate the same payoff also in $\mathcal{T}_V$.

To this end, fix a mechanism $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$ and assume that $(\boldsymbol{p},\boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. Mechanism $(\boldsymbol{p},\boldsymbol{q})$ thus earns at least as high expected payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ under every $\mathbb{P}\in\mathcal{P}$, *i.e.*, condition (1) holds. As $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ is optimal by Theorem 3, this implies that $(\boldsymbol{p},\boldsymbol{q})$ is also optimal in (MDP). As $\arg\max_{i\in\mathcal{I}}\underline{\mu}_i=\{i^\star\}$ is a singleton, we thus know from Lemma 4 that $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to the favored agent $i^\star$ without inspection if $\max_{i\neq i^\star}t_i-c_i<\bar{t}_{i^\star}$, *i.e.*, if $\boldsymbol{t}\in\mathcal{T}\setminus\mathcal{T}_V$. Thus, the allocation probabilities of the mechanisms $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ coincide on $\mathcal{T}\setminus\mathcal{T}_V$, and they earn the same payoff throughout $\mathcal{T}\setminus\mathcal{T}_V$.

In the following we show that $(\boldsymbol{p},\boldsymbol{q})$ can weakly Pareto robustly dominate $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ only if

$$\sum_{i\in\arg\max_{i'\in\mathcal{I}}t_{i'}-c_{i'}}p_i(\boldsymbol{t})=1\quad\text{and}\quad p_i(\boldsymbol{t})=q_i(\boldsymbol{t})\quad\forall i\in\arg\max_{i'\in\mathcal{I}}t_{i'}-c_{i'}\tag{15}$$

for all $\boldsymbol{t}\in\mathcal{T}_V$. Note that (15) immediately implies that $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ generate the same payoff $\max_{i\in\mathcal{I}}t_i-c_i$ throughout $\mathcal{T}_V$.

Define now the set-valued function $\mathcal{I}(\boldsymbol{t})=\{i\in\mathcal{I}:t_i\geq\bar{t}_{i^\star}\}$ for $\boldsymbol{t}\in\mathcal{T}_V$. Note that $|\mathcal{I}(\boldsymbol{t})|\geq 1$ for all $\boldsymbol{t}\in\mathcal{T}_V$ thanks to the definition of $\mathcal{T}_V$, which ensures that there exists at least one agent $i\neq i^\star$ with $t_i-c_i\geq\bar{t}_{i^\star}$ and $\arg\max_{i\neq i^\star}t_i-c_i\subseteq\mathcal{I}(\boldsymbol{t})$. We now prove by induction that (15) holds for all type profiles in $\mathcal{T}_V^n=\{\boldsymbol{t}\in\mathcal{T}_V:|\mathcal{I}(\boldsymbol{t})|=n\}$ for all $n\geq 1$.

As for the base step, set $n=1$ and fix any $\boldsymbol{t}\in\mathcal{T}_V^1$. Thus, there exists exactly one agent $i^\circ\neq i^\star$ such that $t_{i^\circ}\geq\bar{t}_{i^\star}$. Incentive compatibility ensures that $p_{i^\circ}(\boldsymbol{t})-q_{i^\circ}(\boldsymbol{t})\leq p_{i^\circ}(\underline{t}_{i^\circ},\boldsymbol{t}_{-i^\circ})=0$, where the equality holds because $(\underline{t}_{i^\circ},\boldsymbol{t}_{-i^\circ})\in\mathcal{T}\setminus\mathcal{T}_V$ and because $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection on $\mathcal{T}\setminus\mathcal{T}_V$. We thus have $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$. If $p_{i^\circ}(\boldsymbol{t})<1$, then

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\leq p_{i^\circ}(\boldsymbol{t})(t_{i^\circ}-c_{i^\circ})+\sum_{i\neq i^\circ}p_i(\boldsymbol{t})t_i<\max_{i\in\mathcal{I}}t_i-c_i,\tag{16}$$

where the first inequality holds because $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$ and $c_i>0$ for all $i\neq i^\circ$. The second inequality follows from the assumption that $p_{i^\circ}(\boldsymbol{t})<1$ as well as the definition of $\mathcal{T}_V^1$ and the construction of i°, which imply that $t_{i^\circ}-c_{i^\circ}=\max_{i\in\mathcal{I}}t_i-c_i\geq\bar{t}_{i^\star}$ and $\bar{t}_{i^\star}>t_i$ for all $i\neq i^\circ$. This shows that $(\boldsymbol{p},\boldsymbol{q})$ earns strictly less than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in scenario $\boldsymbol{t}$. We next show that this fact contradicts inequality (1).

Due to Lemma 2, there exists $\hat{\boldsymbol{t}} \in \mathcal{T}$, where $\max_{i \neq i^\star} \hat{t}_i < \hat{t}_{i^\star}$, and $\mathbb{P} \in \mathcal{P}$ that satisfy: (i) $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] = \underline{\mu}_i$ $\forall i \in \mathcal{I}$, (ii) $\mathbb{P}(\tilde{\boldsymbol{t}} \in \{\boldsymbol{t}, \hat{\boldsymbol{t}}\}) = 1$, (iii) $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. As $\hat{\boldsymbol{t}} \in \mathcal{T} \setminus \mathcal{T}_V$ by definition, we have

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[\sum_{i \in \mathcal{I}} (p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\right] &= \alpha \sum_{i \in \mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) + (1-\alpha)\sum_{i \in \mathcal{I}} (p_i(\hat{\boldsymbol{t}})\hat{t}_i - q_i(\hat{\boldsymbol{t}})c_i) \\ &< \alpha(t_{i^\circ} - c_{i^\circ}) + (1-\alpha)\hat{t}_{i^\star} = \mathbb{E}_{\mathbb{P}}\left[\sum_{i \in \mathcal{I}} (p_i^\star(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i^\star(\tilde{\boldsymbol{t}})c_i)\right],\end{aligned}$$

where $\alpha \in (0,1]$ indicates the probability of $\tilde{\boldsymbol{t}} = \boldsymbol{t}$, and the inequality follows from (16) and the fact that the payoff at scenario $\hat{\boldsymbol{t}}$ is smaller than $\hat{t}_{i^\star}$ because $\max_{i \neq i^\star} \hat{t}_i < \hat{t}_{i^\star}$ and because $(\boldsymbol{p}, \boldsymbol{q})$ satisfies (FC) and $c_i > 0$ for all $i \in \mathcal{I}$. The strict inequality above implies that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ earns a strictly higher expected payoff than $(\boldsymbol{p}, \boldsymbol{q})$ under $\mathbb{P} \in \mathcal{P}$. It thus contradicts inequality (1) and our assumption that $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Hence, we have established (15) in scenario $\boldsymbol{t}$.

As for the induction step, assume that (15) holds throughout $\mathcal{T}_V^n$ for some $n \geq 1$, and fix an arbitrary type profile $\boldsymbol{t} \in \mathcal{T}_V^{n+1}$. Thus, there exist exactly $n+1$ agents i with types $t_i \geq \bar{t}_{i^\star}$. For any agent $i \in \mathcal{I}(\boldsymbol{t}) \setminus \{i^\star\}$ incentive compatibility dictates that $p_i(\boldsymbol{t}) - q_i(\boldsymbol{t}) \leq p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, where the equality follows from Lemma 4 and the induction hypothesis because $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_V^n \cup (\mathcal{T} \setminus \mathcal{T}_V)$. If $i^\star \in \mathcal{I}(\boldsymbol{t})$, then we can make a similar argument for $i^\star$. In fact, incentive compatibility dictates that $p_{i^\star}(\boldsymbol{t}) - q_{i^\star}(\boldsymbol{t}) \leq p_{i^\star}(\underline{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 0$, where the equality follows from the induction hypothesis because $(\underline{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_V^n$. In summary, we have thus shown that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})$. The first condition in (15) can be proved by repeating the corresponding arguments from the base step almost verbatim. Details are omitted for brevity. We have thus established (15) in an arbitrary scenario $\boldsymbol{t} \in \mathcal{T}_V^{n+1}$. By induction, the revenues generated by the mechanisms $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ must therefore coincide throughout $\cup_{n \geq 1} \mathcal{T}_V^n = \mathcal{T}_V$. This observation completes the proof. □

Proof of Theorem 4. Let $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ denote the allocation probabilities of the favored-agent mechanism described in Theorem 4. We know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is optimal from Theorem 3. To show that it is also Pareto robustly optimal, fix a mechanism $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$ and suppose that $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, *i.e.*, condition (1) holds. We will show that $(\boldsymbol{p}, \boldsymbol{q})$ cannot (strictly) Pareto robustly dominate $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$.

If $\arg\max_{i \in \mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, we know from Lemma 5 that $(\boldsymbol{p}, \boldsymbol{q})$ cannot generate strictly higher expected payoff under any $\mathbb{P} \in \mathcal{P}$, and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is thus Pareto robustly optimal. Suppose now that $\arg\max_{i \in \mathcal{I}} \underline{\mu}_i$ is not a singleton. Select any $\varepsilon \in (0, \overline{\mu}_{i^\star} - \underline{\mu}_{i^\star})$ that exists because $\underline{\mu}_{i^\star} < \overline{\mu}_{i^\star}$, and define

$$\mathcal{P}_\varepsilon = \{\mathbb{P} \in \mathcal{P} : \mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}] \in [\underline{\mu}_{i^\star} + \varepsilon, \overline{\mu}_{i^\star}]\}.$$

Set $\mathcal{P}_\varepsilon$ represents another Markov ambiguity set where the lowest mean value $\underline{\mu}_{i^\star}$ of bidder $i^\star$ is shifted to $\underline{\mu}_{i^\star} + \varepsilon$. Note that agent $i^\star$ becomes the unique agent with the maximum lowest mean value under $\mathcal{P}_\varepsilon$. As $\mathcal{P}_\varepsilon \subset \mathcal{P}$ by construction, we have

$$\mathbb{E}_{\mathbb{P}}\left[\sum_{i \in \mathcal{I}} (p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\right] \geq \mathbb{E}_{\mathbb{P}}\left[\sum_{i \in \mathcal{I}} (p_i^\star(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i^\star(\tilde{\boldsymbol{t}})c_i)\right] \quad \forall \mathbb{P} \in \mathcal{P}_\varepsilon.$$

Thus, $(\boldsymbol{p}, \boldsymbol{q})$ also weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ under the Markov ambiguity set $\mathcal{P}_\varepsilon$. By Lemma 5, we can now conclude that $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generate the same payoff for the principal in any scenario $\boldsymbol{t} \in \mathcal{T}$. This implies that the expected payoff of $(\boldsymbol{p}, \boldsymbol{q})$ cannot exceed the one of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ under any distribution $\mathbb{P}$ supported on $\mathcal{T}$. Thus, none of the inequalities in (1) can be strict, and $(\boldsymbol{p}, \boldsymbol{q})$ cannot Pareto robustly dominate $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. The claim thus follows. □

Proof of Theorem 5. Select any favored-agent mechanism with $i^\star \in \arg\max_{i\in\mathcal{I}} \underline{\mu}_i$ and $\nu^\star \geq \max_{i\in\mathcal{I}} \underline{\mu}_i$, denote by $(\boldsymbol{p},\boldsymbol{q})$ its allocation probabilities. Recall first that this mechanism is feasible in (MDP). We will prove that $(\boldsymbol{p},\boldsymbol{q})$ attains a worst-case expected payoff that is at least as large as $\max_{i\in\mathcal{I}} \underline{\mu}_i$, which implies via Proposition 3 that it is optimal in (MDP).

To this end, fix an arbitrary distribution $\mathbb{P}\in\mathcal{P}$ and suppose for ease of exposition that $\mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i < \nu^\star\right)$, $\mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i = \nu^\star\right)$ and $\mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i > \nu^\star\right)$ are all strictly positive. We can write the principal's expected payoff from $(\boldsymbol{p},\boldsymbol{q})$ under $\mathbb{P}$ as

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\right] =& \,\mathbb{P}\left(\max_{i\neq i^\star}\tilde{t}_i - c_i < \nu^\star\right)\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i < \nu^\star\right] \\ &+\mathbb{P}\left(\max_{i\neq i^\star}\tilde{t}_i - c_i = \nu^\star\right)\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i = \nu^\star\right] \\ &+\mathbb{P}\left(\max_{i\neq i^\star}\tilde{t}_i - c_i > \nu^\star\right)\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i > \nu^\star\right].\end{aligned} \tag{17}$$

If one or more of $\mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i < \nu^\star\right)$, $\mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i = \nu^\star\right)$ and $\mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i > \nu^\star\right)$ are zero, the right-hand side of equation (17) can be adjusted by removing the respective terms, and the proof proceeds similarly.

In the following, we will show that all of the conditional expectations above, and therefore the principal's expected payoff under $\mathbb{P}$, are greater than or equal to $z^\star = \max_{i\in\mathcal{I}} \underline{\mu}_i$. If $\max_{i\neq i^\star} t_i - c_i < \nu^\star$, condition (i) in Definition 3 implies that the principal's payoff amounts to $t_{i^\star}$. This implies that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i < \nu^\star\right] &= \mathbb{E}_{\mathbb{P}}\left[\tilde{t}_{i^\star}\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i < \nu^\star\right] \\ &= \mathbb{E}_{\mathbb{P}}\left[\tilde{t}_{i^\star}\right] = \mu_{i^\star} = \max_{i\in\mathcal{I}}\underline{\mu}_i,\end{aligned}$$

where the second equality holds because the agents' types are independent. If $\max_{i\neq i^\star} t_i - c_i > \nu^\star$, then condition (ii) in Definition 3 implies that the principal's payoff amounts to $\max_{i\in\mathcal{I}} t_i - c_i$. We thus have

$$\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i > \nu^\star\right] = \mathbb{E}_{\mathbb{P}}\left[\max_{i\in\mathcal{I}}\tilde{t}_i - c_i\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i > \nu^\star\right] > \nu^\star \geq \max_{i\in\mathcal{I}}\underline{\mu}_i.$$

If $\max_{i\neq i^\star} t_i - c_i = \nu^\star$, then the allocation functions are defined either as in condition (i) or as in condition (ii) of Definition 3. If the allocation functions are defined as in condition (i), we have

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i = \nu^\star\right] &= \mathbb{E}_{\mathbb{P}}\left[\tilde{t}_{i^\star}\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i = \nu^\star\right] \\ &= \mathbb{E}_{\mathbb{P}}\left[\tilde{t}_{i^\star}\right] = \mu_{i^\star} = \max_{i\in\mathcal{I}}\underline{\mu}_i,\end{aligned}$$

where the second equality again holds because the agents' types are independent. If the allocation functions are defined as in condition (ii), on the other hand, then

$$\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i = \nu^\star\right] = \mathbb{E}_{\mathbb{P}}\left[\max_{i\in\mathcal{I}}\tilde{t}_i - c_i\,\middle|\, \max_{i\neq i^\star}\tilde{t}_i - c_i = \nu^\star\right] \geq \nu^\star \geq \max_{i\in\mathcal{I}}\underline{\mu}_i.$$

In summary, we have shown that all of the conditional expectations in (17), and therefore also the principal's expected payoff under $\mathbb{P}$, are non-inferior to $\max_{i\in\mathcal{I}} \underline{\mu}_i$. As the distribution $\mathbb{P}\in\mathcal{P}$ was chosen arbitrarily, this reasoning implies that the principal's worst-case expected payoff is also non-inferior to $\max_{i\in\mathcal{I}} \underline{\mu}_i$. The favored-agent mechanism at hand is therefore optimal in (MDP) by virtue of Proposition 3. □

Proof of Lemma 6. Consider arbitrary $\boldsymbol{t}\in\mathcal{T}$ and $\mu_{i^\star}\in[\underline{\mu}_{i^\star},\overline{\mu}_{i^\star}]$. We will construct a scenario $\hat{\boldsymbol{t}}\in\mathcal{T}$, where $\max_{i\neq i^\star}\hat{t}_i<\underline{\mu}_{i^\star}$, and a discrete distribution $\mathbb{P}\in\mathcal{P}$ that satisfies (i)–(iii). To this end, we define $\hat{t}_i$ through

$$\hat{t}_i=\begin{cases}t_i & \text{if } t_i=\underline{\mu}_i\\ \underline{t}_i & \text{if } t_i>\underline{\mu}_i\\ \underline{\mu}_i+\varepsilon & \text{if } t_i<\underline{\mu}_i\end{cases}\quad\forall i\in\mathcal{I}\setminus\{i^\star\}\quad\text{and}\quad \hat{t}_{i^\star}=\begin{cases}t_{i^\star} & \text{if } t_{i^\star}=\mu_{i^\star}\\ \underline{t}_{i^\star} & \text{if } t_{i^\star}>\mu_{i^\star}\\ \mu_{i^\star}+\varepsilon & \text{if } t_{i^\star}<\mu_{i^\star},\end{cases}$$

where $\varepsilon\in(0,\min_{i\in\mathcal{I}}\overline{t}_i-\overline{\mu}_i)\cap\big(0,(\underline{\mu}_{i^\star}-\max_{i\neq i^\star}\underline{\mu}_i)/2\big)$ is a fixed positive number. Note that there exists such $\varepsilon>0$ because $\underline{\mu}_i<\overline{\mu}_i<\overline{t}_i$ for all $i\in\mathcal{I}$ and $\arg\max_{i\in\mathcal{I}}\underline{\mu}_i=\{i^\star\}$ is a singleton. We next show that $\hat{t}_i\in\mathcal{T}_i$ for all $i\in\mathcal{I}$ (*i.e.*, $\hat{\boldsymbol{t}}\in\mathcal{T}$) and $\max_{i\neq i^\star}\hat{t}_i<\underline{\mu}_{i^\star}$. For any $i\in\mathcal{I}$, we have

$$\hat{t}_i\le\overline{\mu}_i+\varepsilon\le\overline{\mu}_i+\min_{j\in\mathcal{I}}(\overline{t}_j-\overline{\mu}_j)\le\overline{\mu}_i+\overline{t}_i-\overline{\mu}_i=\overline{t}_i,$$

where the first inequality follows from the definition of $\hat{t}_i$, and the second inequality holds because $\varepsilon<\min_{j\in\mathcal{I}}\overline{t}_j-\overline{\mu}_j$. The definition of $\hat{t}_i$ implies that we also have $\hat{t}_i\ge\underline{t}_i$. We thus showed that $\hat{\boldsymbol{t}}\in\mathcal{T}$.

For all $i\neq i^\star$, we moreover have

$$\hat{t}_i\le\underline{\mu}_i+\varepsilon\le\underline{\mu}_i+(\underline{\mu}_{i^\star}-\max_{j\neq i^\star}\underline{\mu}_j)/2\le\underline{\mu}_i+(\underline{\mu}_{i^\star}-\underline{\mu}_i)/2<\underline{\mu}_{i^\star},$$

where the first inequality again follows from the definition of $\hat{t}_i$, the second inequality holds because $\varepsilon<(\underline{\mu}_{i^\star}-\max_{i\neq i^\star}\underline{\mu}_i)/2$, and the fourth inequality holds because $\arg\max_{i\in\mathcal{I}}\underline{\mu}_i=\{i^\star\}$ is a singleton. We thus showed that $\max_{i\neq i^\star}\hat{t}_i<\underline{\mu}_{i^\star}$.

Next, we will construct a discrete distribution $\mathbb{P}$ through the marginal distributions $\mathbb{P}_i=\alpha_i\delta_{t_i}+(1-\alpha_i)\delta_{\hat{t}_i}$ of $\tilde{t}_i$'s, where $\alpha_i\in(0,1]$ for all $i\in\mathcal{I}$. We will then show that $\mathbb{P}$ belongs to the Markov ambiguity set $\mathcal{P}$ and moreover satisfies the properties (i)–(iii). To this end, we define α_i through

$$\alpha_i=\begin{cases}1 & \text{if } t_i=\hat{t}_i,\\ (\underline{\mu}_i-\hat{t}_i)/(t_i-\hat{t}_i) & \text{if } t_i\neq\hat{t}_i,\end{cases}\quad\forall i\in\mathcal{I}\setminus\{i^\star\}$$

and

$$\alpha_{i^\star}=\begin{cases}1 & \text{if } t_{i^\star}=\hat{t}_{i^\star},\\ (\mu_{i^\star}-\hat{t}_{i^\star})/(t_{i^\star}-\hat{t}_{i^\star}) & \text{if } t_{i^\star}\neq\hat{t}_{i^\star}.\end{cases}$$

We first show that $\alpha_i\in(0,1]$ for all $i\in\mathcal{I}$. For any $i\in\mathcal{I}$, it is sufficient to show that the claim holds if $t_i\neq\hat{t}_i$. For any $i\neq i^\star$, if $t_i\neq\hat{t}_i$ and $t_i>\underline{\mu}_i$, we have

$$\alpha_i=(\underline{\mu}_i-\hat{t}_i)/(t_i-\hat{t}_i)=(\underline{\mu}_i-\underline{t}_i)/(t_i-\underline{t}_i)\in(0,1),$$

where the second equality follows from the definition of $\hat{t}_i$, and the inclusion holds because $t_i>\underline{\mu}_i>\underline{t}_i$. If $t_i\neq\hat{t}_i$ and $t_i<\underline{\mu}_i$, on the other hand, we have $\alpha_i=-\varepsilon/(t_i-\underline{\mu}_i-\varepsilon)\in(0,1)$, where the equality again follows from the definition of $\hat{t}_i$, and the inclusion holds because $t_i<\underline{\mu}_i<\underline{\mu}_i+\varepsilon$. Note that if $t_i=\underline{\mu}_i$, then $\hat{t}_i=t_i$ by definition, and $\alpha_i=1$. One can similarly show that $\alpha_{i^\star}\in(0,1]$ by replacing $\underline{\mu}_{i^\star}$ with $\mu_{i^\star}$ in the above arguments. Thus, $\alpha_i\in(0,1]$ for all $i\in\mathcal{I}$. We now define $\mathbb{P}$ through the marginal distributions $\mathbb{P}_i=\alpha_i\delta_{\boldsymbol{t}_i}+(1-\alpha_i)\delta_{\hat{\boldsymbol{t}}_i}$, $i\in\mathcal{I}$, as follows.

$$\mathbb{P}(\tilde{\boldsymbol{t}}=\boldsymbol{t})=\prod_{i\in\mathcal{I}}\mathbb{P}_i(\tilde{t}_i=t_i)\quad\forall\boldsymbol{t}\in\mathcal{T}$$

By construction, $\tilde{t}_i$'s are mutually independent under $\mathbb{P}$. Hence, the expected type of each $i \in \mathcal{I}$ amounts to $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] = \alpha_i t_i + (1-\alpha_i)\hat{t}_i$.

We next show that $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] \in [\underline{\mu}_i, \overline{\mu}_i]$ for all $i \in \mathcal{I}$, which implies that $\mathbb{P} \in \mathcal{P}$. For any $i \neq i^\star$, if $t_i = \hat{t}_i$, then we have $t_i = \hat{t}_i = \underline{\mu}_i$ by definition of $\hat{t}_i$. The expected type therefore amounts to $\underline{\mu}_i$. If $t_i \neq \hat{t}_i$, on the other hand, we have

$$\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] = \alpha_i t_i + (1-\alpha_i)\hat{t}_i = \alpha_i (t_i - \hat{t}_i) + \hat{t}_i = \frac{\underline{\mu}_i - \hat{t}_i}{t_i - \hat{t}_i}(t_i - \hat{t}_i) + \hat{t}_i = \underline{\mu}_i,$$

where the third equality follows from the definition of α_i. One can verify that $\mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}] = \mu_{i^\star}$ using similar arguments. We thus showed that $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] \in [\underline{\mu}_i, \overline{\mu}_i]$ for all $i \in \mathcal{I}$, and therefore $\mathbb{P} \in \mathcal{P}$.

It remains to show that $\mathbb{P}$ satisfies (i)–(iii). As we have $\mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}] = \mu_{i^\star}$, property (i) holds. The definition of $\mathbb{P}$ implies that (ii) and (iii) also hold. □

Proof of Lemma 7. Consider any subset $\mathcal{T}' = \prod_{i\in\mathcal{I}} \mathcal{T}'_i$ of $\mathcal{T}$ such that (i) and (ii) holds. Also, consider any $(\boldsymbol{p}, \boldsymbol{q}), (\boldsymbol{p}', \boldsymbol{q}') \in \mathcal{X}$ such that $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}', \boldsymbol{q}')$ and (11) holds. Suppose for the sake of contradiction that (11) is strict for some $\boldsymbol{t} \in \mathcal{T}'$.

We will characterize a discrete distribution $\mathbb{P} \in \mathcal{P}$ under which the expected payoff of mechanism $(\boldsymbol{p}, \boldsymbol{q})$ is strictly lower than that of $(\boldsymbol{p}', \boldsymbol{q}')$, which contradicts that $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}', \boldsymbol{q}')$. By Lemma 6, for scenario $\boldsymbol{t}$ and for any $\mu_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$, there exist a scenario $\hat{\boldsymbol{t}} \in \mathcal{T}$, where $\max_{i\neq i^\star} \hat{t}_i < \underline{\mu}_{i^\star}$, and a discrete distribution $\mathbb{P} \in \mathcal{P}$ that satisfy the following properties: (i) $\mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}] = \mu_{i^\star}$, (ii) $\mathbb{P}(\tilde{t}_i \in \{t_i, \hat{t}_i\}) = 1$ for all $i \in \mathcal{I}$, (iii) $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. We next show that there is always a $\mu_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$ such that distribution $\mathbb{P}$ also satisfies $\mathbb{P}(\tilde{\boldsymbol{t}} \in \mathcal{T}') = 1$. Note that if $\mathbb{P}$ satisfies $\mathbb{P}(\tilde{t}_i \in \mathcal{T}'_i) = 1$ for all $i \in \mathcal{I}$, then it also satisfies $\mathbb{P}(\tilde{\boldsymbol{t}} \in \mathcal{T}') = 1$ as $\mathcal{T}' = \prod_{i\in\mathcal{I}} \mathcal{T}'_i$. First, suppose that $\mathcal{T}'_{i^\star} \subseteq [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$. For $\mu_{i^\star} = t_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$, properties (i)–(iii) on $\mathbb{P}$ imply that if $\hat{t}_{i^\star} \neq t_{i^\star}$, then $\mathbb{P}(\tilde{t}_{i^\star} = \hat{t}_{i^\star}) = 0$. We thus have $\mathbb{P}(\tilde{t}_{i^\star} = t_{i^\star}) = 1$, which implies that $\mathbb{P}(\tilde{t}_{i^\star} \in \mathcal{T}'_{i^\star}) = 1$ as $t_{i^\star} \in \mathcal{T}'_{i^\star}$. For any $i \in \mathcal{I} \setminus \{i^\star\}$, as $\hat{t}_i < \underline{\mu}_{i^\star}$, we have $\hat{t}_i \in \mathcal{T}'_i \supseteq [\underline{t}_i, \underline{\mu}_{i^\star})$ irrespective of the value of $\mu_{i^\star}$. Condition (ii) on $\mathbb{P}$ thus implies that $\mathbb{P}(\tilde{t}_i \in \mathcal{T}'_i) = 1$. Suppose now that $\mathcal{T}'_{i^\star} = \mathcal{T}_{i^\star}$. Condition (ii) on $\mathbb{P}$ implies that $\mathbb{P}(\tilde{t}_{i^\star} \in \mathcal{T}'_{i^\star}) = 1$ as $t_{i^\star}, \hat{t}_{i^\star} \in \mathcal{T}_{i^\star} = \mathcal{T}'_{i^\star}$. We already showed that $\mathbb{P}(\tilde{t}_i \in \mathcal{T}'_i) = 1$ for every other $i \in \mathcal{I} \setminus \{i^\star\}$ irrespective of the value of $\mu_{i^\star}$. We can thus conclude that there always exists a $\mu_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$ such that distribution $\mathbb{P}$ from Lemma 6 also satisfies $\mathbb{P}(\tilde{\boldsymbol{t}} \in \mathcal{T}') = 1$.

Now, keeping in mind that $\mathbb{P}$ is a discrete distribution with properties $\mathbb{P}(\tilde{\boldsymbol{t}} \in \mathcal{T}') = 1$ and $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$, we can bound the principal's expected payoff from $(\boldsymbol{p}, \boldsymbol{q})$ under $\mathbb{P}$ as follows:

$$\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}} (p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\right] < \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}} (p'_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q'_i(\tilde{\boldsymbol{t}})c_i)\right],$$

where the strict inequality follows from (11) and the assumption that (11) is strict for $\boldsymbol{t} \in \mathcal{T}'$. Therefore, we conclude that $(\boldsymbol{p}, \boldsymbol{q})$ cannot weakly Pareto robustly dominate $(\boldsymbol{p}', \boldsymbol{q}')$ unless the inequalities in (11) hold with equality. □

LEMMA 9. *The sets $\mathcal{T}_I, \mathcal{T}_{II}, \mathcal{T}_{III}, \mathcal{T}_{IV}, \mathcal{T}_V$ defined in (12) is a partition of the type space $\mathcal{T}$.*

Proof. We give another partition of the set $\mathcal{T}$ in the following.

$$\begin{aligned}
\mathcal{T}_I &= \{\boldsymbol{t} \in \mathcal{T} : t_1 \in (\underline{\mu}_1, \overline{\mu}_1] \text{ and } t_2 \leq \underline{\mu}_1\} \\
\mathcal{T}_{II} &= \{\boldsymbol{t} \in \mathcal{T} : t_1 \in (\underline{\mu}_1, \overline{\mu}_1],\ t_2 > \underline{\mu}_1 \text{ and } t_2 - c_2 \leq \underline{\mu}_1\} \\
\mathcal{T}_{II'} &= \{\boldsymbol{t} \in \mathcal{T} : t_1 \in (\underline{\mu}_1, \overline{\mu}_1],\ t_2 > \underline{\mu}_1 \text{ and } t_2 - c_2 > \underline{\mu}_1\} \\
\mathcal{T}_{III} &= \{\boldsymbol{t} \in \mathcal{T} : t_1 \notin (\underline{\mu}_1, \overline{\mu}_1] \text{ and } t_2 - c_2 \leq \underline{\mu}_1\} \\
\mathcal{T}_{III'} &= \{\boldsymbol{t} \in \mathcal{T} : t_1 \notin (\underline{\mu}_1, \overline{\mu}_1] \text{ and } t_2 - c_2 > \underline{\mu}_1\}
\end{aligned}$$

Note that the condition $t_2 > \underline{\mu}_1$ in $\mathcal{T}_{II'}$ is redundant but makes it easy to see that $\mathcal{T}_I, \mathcal{T}_{II}, \mathcal{T}_{II'}, \mathcal{T}_{III}, \mathcal{T}_{III'}$ is a partition of the type space $\mathcal{T}$. We next show that we can replace the sets $\mathcal{T}_{II'}$ and $\mathcal{T}_{III'}$ with

$$\mathcal{T}_{IV} = \{\boldsymbol{t} \in \mathcal{T} : t_1 = \underline{\mu}_1 \text{ and } t_2 - c_2 > \underline{\mu}_1\}$$
$$\mathcal{T}_V = \{\boldsymbol{t} \in \mathcal{T} : t_1 \neq \underline{\mu}_1 \text{ and } t_2 - c_2 > \underline{\mu}_1\}$$

and obtain a partition of $\mathcal{T}$. To this end, first note that the intersection of $\mathcal{T}_{IV}$ and $\mathcal{T}_V$ is empty. Moreover, their union is given by $\{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 > \underline{\mu}_1\}$ that is the same as the union of $\mathcal{T}_{II'}$ and $\mathcal{T}_{III'}$. Thus, $\mathcal{T}_I, \mathcal{T}_{II}, \mathcal{T}_{III}, \mathcal{T}_{IV}, \mathcal{T}_V$ is a partition of the type space $\mathcal{T}$. □

Proof of Lemma 8. Consider the following partition of the set $\mathcal{T}$.

$$\mathcal{T}_I = \{\boldsymbol{t} \in \mathcal{T} : t_{i^\star} \in (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}] \text{ and } \max_{i \neq i^\star} t_i \leq \underline{\mu}_{i^\star}\}$$
$$\mathcal{T}_{II} = \{\boldsymbol{t} \in \mathcal{T} : t_{i^\star} \in (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}] \text{ and } \max_{i \neq i^\star} t_i > \underline{\mu}_{i^\star} \text{ and } \max_{i \neq i^\star} t_i - c_i \leq \underline{\mu}_{i^\star}\}$$
$$\mathcal{T}_{II'} = \{\boldsymbol{t} \in \mathcal{T} : t_{i^\star} \in (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}] \text{ and } \max_{i \neq i^\star} t_i > \underline{\mu}_{i^\star} \text{ and } \max_{i \neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}\}$$
$$\mathcal{T}_{III} = \{\boldsymbol{t} \in \mathcal{T} : t_{i^\star} \notin (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}] \text{ and } \max_{i \neq i^\star} t_i - c_i \leq \underline{\mu}_{i^\star}\}$$
$$\mathcal{T}_{III'} = \{\boldsymbol{t} \in \mathcal{T} : t_{i^\star} \notin (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}] \text{ and } \max_{i \neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}\}$$

We can replace $\mathcal{T}_{II'}$ and $\mathcal{T}_{III'}$ with the following two sets to obtain a different partition of $\mathcal{T}$.

$$\mathcal{T}_{IV} = \{\boldsymbol{t} \in \mathcal{T} : t_{i^\star} = \underline{\mu}_{i^\star} \text{ and } \max_{i \neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}\}$$
$$\mathcal{T}_V = \{\boldsymbol{t} \in \mathcal{T} : t_{i^\star} \neq \underline{\mu}_{i^\star} \text{ and } \max_{i \neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}\}$$

This is because $\mathcal{T}_{IV}$ and $\mathcal{T}_V$ are disjoint sets that have the same union as the union of $\mathcal{T}_{II'}$ and $\mathcal{T}_{III'}$. Throughout the proof we consider the partition $\mathcal{T}_I, \mathcal{T}_{II}, \mathcal{T}_{III}, \mathcal{T}_{IV}, \mathcal{T}_V$. Note that $\mathcal{T}_I$ and $\mathcal{T}_{III}$ are nonempty as $\arg\max_{i \in \mathcal{I}} \underline{\mu}_i = \{i^\star\}$ and $[\underline{\mu}_i, \overline{\mu}_i] \in (\underline{t}_i, \bar{t}_i)$ for all $i \in \mathcal{I}$, but sets $\mathcal{T}_{II}$, $\mathcal{T}_{IV}$ and $\mathcal{T}_V$ can be empty if $\underline{\mu}_{i^\star}$ or c_i for all $i \neq i^\star$ are sufficiently large.

The remainder of the proof is divided into four steps, each of which proves the claim for one of the subsets $\mathcal{T}_I$, $\mathcal{T}_{II}$, $\mathcal{T}_{III}$, $\mathcal{T}_{IV}$ and $\mathcal{T}_V$.

Step 1 ($\mathcal{T}_I$). For any $\boldsymbol{t} \in \mathcal{T}_I$, the principal's payoff under $(\boldsymbol{p}, \boldsymbol{q})$ satisfies

$$\sum_{i \in \mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \leq \sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) t_i \leq t_{i^\star} = \sum_{i \in \mathcal{I}} (p_i^\star(\boldsymbol{t}) t_i - q_i^\star(\boldsymbol{t}) c_i),$$

where the first inequality holds because $q_i(\boldsymbol{t})$ and c_i are non-negative, the second inequality follows from (FC) and that $\max_{i \neq i^\star} t_i \leq \underline{\mu}_{i^\star} < t_{i^\star}$, and the equality follows from the definition of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. The payoff of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is thus larger than or equal to the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ in every $\boldsymbol{t} \in \mathcal{T}_I$. Moreover, note that $\mathcal{T}_I$ can be written as $\prod_{i \in \mathcal{I}} \mathcal{T}_{Ii}$ where $\mathcal{T}_{Ii^\star} = (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$ and $\mathcal{T}_{Ii} = [\underline{t}_i, \underline{\mu}_{i^\star}] \cap \mathcal{T}_i$ for all $i \neq i^\star$. The set $\mathcal{T}_I$ thus satisfies the assumptions (i) and (ii) in Lemma 7. By Lemma 7, we can thus conclude that the payoffs of $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ coincide in $\mathcal{T}_I$. In addition, note that, for any $\boldsymbol{t} \in \mathcal{T}_I$, we have $\max_{i \neq i^\star} t_i < t_{i^\star}$, $q_i(\boldsymbol{t}) \geq 0$, $c_i > 0$ and $\sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1$. This implies that the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ can match the payoff $t_{i^\star}$ of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ only if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$.

Step 2 ($\mathcal{T}_{II}$). We will prove that if mechanism $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ then it must satisfy $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ for any $\boldsymbol{t} \in \mathcal{T}_{II}$, which implies that the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ matches that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ throughout $\mathcal{T}_{II}$. To this end, define the set-valued function $\mathcal{I}(\boldsymbol{t}) = \{i \in \mathcal{I} : t_i > \underline{\mu}_{i^\star}\}$ for $\boldsymbol{t} \in \mathcal{T}_{II}$. Note that $|\mathcal{I}(\boldsymbol{t})| \geq 2$ for all $\boldsymbol{t} \in \mathcal{T}_{II}$ by the definition of $\mathcal{T}_{II}$, which ensures that $\max_{i \neq i^\star} t_i > \underline{\mu}_{i^\star}$ and $t_{i^\star} \in (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$. We now prove by induction that the claim holds in $\mathcal{T}_{II}^n = \{t \in \mathcal{T}_{II} : |\mathcal{I}(\boldsymbol{t})| = n\}$ for all $n \geq 2$.

As for the base step, set $n = 2$ and fix any $\boldsymbol{t} \in \mathcal{T}_{II}^2$. Thus, there exists exactly one agent $i^\circ \neq i^\star$ that satisfies $t_{i^\circ} > \underline{\mu}_{i^\star}$. Incentive compatibility ensures that $p_{i^\circ}(\boldsymbol{t}) - q_{i^\circ}(\boldsymbol{t}) \leq p_{i^\circ}(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) = 0$, where the equality holds because $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I$ and because we know from Step 1 that $(\boldsymbol{p}, \boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection in $\mathcal{T}_I$. We thus have $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$. Then, we have

$$\begin{aligned}\sum_{j\in\mathcal{I}}(p_j(\boldsymbol{t})t_j - q_j(\boldsymbol{t})c_j) &\le \sum_{j\neq i^\circ} p_j(\boldsymbol{t})t_j + p_{i^\circ}(\boldsymbol{t})(t_{i^\circ} - c_{i^\circ}) \\ &\le t_{i^\star} = \sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),\end{aligned}$$

where the first inequality holds because $q_j(\boldsymbol{t})$ and c_j are non-negative and $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$, the second inequality from (FC) and that $t_{i^\circ} - c_{i^\circ} \le \underline{\mu}_{i^\star} < t_{i^\star}$ and $t_j \le \underline{\mu}_{i^\star}$ for all $j \in \mathcal{I}\setminus\{i^\circ, i^\star\}$, and the equality follows from the definition of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. As scenario $\boldsymbol{t}$ is chosen arbitrarily, the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ thus cannot exceed that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ throughout $\mathcal{T}_{II}^2$. Recalling the conclusion from Step 1, we now know that this relation between the payoffs is true for the set $\mathcal{T}_I \cup \mathcal{T}_{II}^2$.

For any $i^\circ \in \mathcal{I}\setminus\{i^\star\}$, define $\mathcal{T}_{II}^2(i^\circ)$ as the subset of $\mathcal{T}_{II}^2$ where i° is the only agent with type $t_{i^\circ} > \underline{\mu}_{i^\star}$ and note that $\mathcal{T}_{II}^2 = \cup_{i^\circ\in\mathcal{I}\setminus\{i^\star\}}\mathcal{T}_{II}^2(i^\circ)$. Consider an arbitrary $i^\circ \in \mathcal{I}\setminus\{i^\star\}$ and the set $\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ)$, which can be written as $\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ) = \prod_{i\in\mathcal{I}}(\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ))_i$, where $(\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ))_{i^\star} = (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$, $(\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ))_{i^\circ} = [\underline{t}_{i^\circ}, c_{i^\circ} + \underline{\mu}_{i^\star}] \cap \mathcal{T}_{i^\circ}$ and $(\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ))_i = [\underline{t}_i, \underline{\mu}_{i^\star}] \cap \mathcal{T}_i$ for all $i \notin \{i^\star, i^\circ\}$. The set $\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ)$ satisfies the assumptions (i) and (ii) in Lemma 7. Mechanisms $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ thus generate the same payoff throughout $\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ)$ by Lemma 7. By definition, the payoff of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ amounts to $t_{i^\star}$ in $\mathcal{T}_I \cup \mathcal{T}_{II}^2(i^\circ)$. For any $\boldsymbol{t} \in \mathcal{T}_{II}^2(i^\circ)$, as $t_{i^\circ} - c_{i^\circ} < t_{i^\star}$, $t_i < t_{i^\star}$ for all $i \notin \{i^\star, i^\circ\}$ and $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$, $(\boldsymbol{p}, \boldsymbol{q})$ can generate a payoff of $t_{i^\star}$ only if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$. As i° is chosen arbitrarily, we have $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ throughout $\mathcal{T}_{II}^2$.

As for the induction step, assume that $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ for all $\boldsymbol{t} \in \mathcal{T}_{II}^n$ and for some $n \ge 2$, and fix a scenario $\boldsymbol{t} \in \mathcal{T}_{II}^{n+1}$. Thus, there exists exactly $n+1$ agents i that satisfy $t_i > \underline{\mu}_{i^\star}$. For any agent $i \in \mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}$, incentive compatibility dictates that $p_i(\boldsymbol{t}) - q_i(\boldsymbol{t}) \le p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, where the equality follows from $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_{II}^n$ and the induction hypothesis. We thus have $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}$. Then,

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \le \sum_{i\notin\mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}} p_i(\boldsymbol{t})t_i + \sum_{i\in\mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}} p_i(\boldsymbol{t})(t_i - c_i) \le t_{i^\star} = \sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the first inequality holds because $q_i(\boldsymbol{t})$ and c_i are non-negative and $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}$, the second inequality follows from (FC) and that $t_i - c_i \le \underline{\mu}_{i^\star}$ for all $i \in \mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}$ and $t_i \le \underline{\mu}_{i^\star} < t_{i^\star}$ for $i \in \mathcal{I}\setminus\mathcal{I}(\boldsymbol{t})$, and the equality follows from the definition of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. As scenario $\boldsymbol{t}$ is chosen arbitrarily, the payoff from $(\boldsymbol{p}, \boldsymbol{q})$ is thus less than or equal to that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ throughout $\mathcal{T}_{II}^{n+1}$. By Step 1, this relationship between the payoffs holds true for the set $\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}$.

For any subset $\mathcal{I}' \ni i^\star$ of agents with $|\mathcal{I}'| = n+1$, define $\mathcal{T}_{II}^{n+1}(\mathcal{I}')$ as the subset of $\mathcal{T}_{II}$ where $t_i \le \underline{\mu}_{i^\star}$ for all $i \notin \mathcal{I}'$. Note that $\mathcal{T}_{II}^{n+1}(\mathcal{I}') \subseteq \cup_{k=2}^{n+1}\mathcal{T}_{II}^k$, and the union of $\mathcal{T}_{II}^{n+1}(\mathcal{I}')$ over all $\mathcal{I}' \subseteq \mathcal{I}$ with $|\mathcal{I}'| = n+1$ and $i^\star \in \mathcal{I}'$ gives us the set $\cup_{k=2}^{n+1}\mathcal{T}_{II}^k$. Consider now an arbitrary $\mathcal{I}' \ni i^\star$ with $|\mathcal{I}'| = n+1$ and the set $\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}')$, which can be written as $\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}') = \prod_{i\in\mathcal{I}}(\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}'))_i$, where $(\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}'))_{i^\star} = (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$, $(\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}'))_i = [\underline{t}_i, c_i + \underline{\mu}_{i^\star}] \cap \mathcal{T}_i$ for $i \in \mathcal{I}'\setminus\{i^\star\}$ and $(\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}'))_i = [\underline{t}_i, \underline{\mu}_{i^\star}] \cap \mathcal{T}_i$ for all $i \in \mathcal{I}\setminus\mathcal{I}'$. The set $\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}')$ satisfies the assumptions (i) and (ii) in Lemma 7. The payoffs of $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ thus coincide in $\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}')$ by Lemma 7. By definition, the payoff of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ amounts to $t_{i^\star}$ throughout $\mathcal{T}_I \cup \mathcal{T}_{II}^{n+1}(\mathcal{I}')$. For any $\boldsymbol{t} \in \mathcal{T}_{II}^{n+1}(\mathcal{I}')$, as $t_i - c_i < t_{i^\star}$ and $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}'\setminus\{i^\star\}$ and $t_i < t_{i^\star}$ for all $i \in \mathcal{I}\setminus\mathcal{I}'$, mechanism $(\boldsymbol{p}, \boldsymbol{q})$ can generate a payoff of $t_{i^\star}$ only if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$. As $\mathcal{I}'$ is chosen arbitrarily, we have $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ throughout $\mathcal{T}_{II}^{n+1}$. This thus completes the induction step.

In summary, the allocation probabilities of any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ should satisfy $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ throughout $\mathcal{T}_{II}$.

Step 3 ($\mathcal{T}_{III}$). Next, fix any type profile $\boldsymbol{t} \in \mathcal{T}_{III}$. Incentive compatibility ensures that $p_{i^\star}(\boldsymbol{t}) \ge p_{i^\star}(\overline{\mu}_{i^\star}, \boldsymbol{t}_{-i^\star}) - q_{i^\star}(\overline{\mu}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 1$, where the equality holds because $(\overline{\mu}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_I \cup \mathcal{T}_{II}$ and because we know from Step 1 and 2 that $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent $i^\star$ without inspection in $\mathcal{T}_I \cup \mathcal{T}_{II}$. We thus have $p_{i^\star}(\boldsymbol{t})=1$ and

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \le p_{i^\star}(\boldsymbol{t})t_{i^\star} + q_{i^\star}(\boldsymbol{t})c_{i^\star} \le t_{i^\star} = \sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the first inequality follows from (FC) and non-negativity of $q_i(\boldsymbol{t})$, the second inequality holds because $q_{i^\star}(\boldsymbol{t})\ge 0$ and $c_{i^\star}>0$, and the equality follows from the definition of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. As scenario $\boldsymbol{t}$ is chosen arbitrarily, $(\boldsymbol{p},\boldsymbol{q})$ cannot generate a payoff higher than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ throughout $\mathcal{T}_{III}$. By Steps 1 and 2, this relation between the payoffs holds for the set $\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}$. Note that the set $\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}$ can be written as $\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}=\prod_{i\in\mathcal{I}}(\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III})_i$, where $(\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III})_{i^\star}=\mathcal{T}_{i^\star}$ and $(\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III})_i=[\underline{t}_i, c_i+\underline{\mu}_{i^\star}]\cap\mathcal{T}_i$ for all $i\in\mathcal{I}\setminus\{i^\star\}$. The set $\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}$ thus satisfies the assumptions (i) and (ii) in Lemma 7. The payoffs of $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ thus coincide throughout $\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}$ by Lemma 7. As $q_{i^\star}(\boldsymbol{t})\ge 0$, $c_{i^\star}>0$ and $(\boldsymbol{p},\boldsymbol{q})$ satisfies the (FC), mechanism $(\boldsymbol{p},\boldsymbol{q})$ can generate a payoff of $t_{i^\star}$ in a scenario $\boldsymbol{t}\in\mathcal{T}_{III}$ only if $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$.

Step 4 ($\mathcal{T}_{IV}$). In this step, we will show that any mechanism $(\boldsymbol{p},\boldsymbol{q})$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must satisfy

$$\sum_{i\in\arg\max_{i'\in\mathcal{I}} t_{i'}-c_{i'}} p_i(\boldsymbol{t})=1 \quad\text{and}\quad p_i(\boldsymbol{t})=q_i(\boldsymbol{t}) \quad \forall i\in\arg\max_{i'\in\mathcal{I}} t_{i'}-c_{i'} \tag{18}$$

for all $\boldsymbol{t}\in\mathcal{T}_{IV}$. This immediately implies that $(\boldsymbol{p},\boldsymbol{q})$ generates the same payoff as $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ throughout $\mathcal{T}_{IV}$. To this end, define the set-valued function $\mathcal{I}(\boldsymbol{t})=\{i\in\mathcal{I}: t_i>\underline{\mu}_{i^\star}\}$ for $\boldsymbol{t}\in\mathcal{T}_{IV}$. Note that $|\mathcal{I}(\boldsymbol{t})|\ge 1$ and $i^\star\notin\mathcal{I}(\boldsymbol{t})$ for all $\boldsymbol{t}\in\mathcal{T}_{IV}$ thanks to the definition of $\mathcal{T}_{IV}$, which ensures that $\max_{i\ne i^\star} t_i - c_i>\underline{\mu}_{i^\star}$ and $t_{i^\star}=\underline{\mu}_{i^\star}$. We will prove by induction that (18) holds in $\mathcal{T}_{IV}^n=\{\boldsymbol{t}\in\mathcal{T}_{IV}: |\mathcal{I}(\boldsymbol{t})|=n\}$ for all $I-1\ge n\ge 1$.

As for the base step, set $n=1$ and fix a scenario $\boldsymbol{t}\in\mathcal{T}_{IV}^1$. Thus, exactly one agent i° satisfies $t_{i^\circ}>\underline{\mu}_{i^\star}$. Incentive compatibility ensures that $p_{i^\circ}(\boldsymbol{t})-q_{i^\circ}(\boldsymbol{t})\le p_{i^\circ}(\underline{t}_{i^\circ},\boldsymbol{t}_{-i^\circ})=0$, where the equality follows from that $(\underline{t}_{i^\circ},\boldsymbol{t}_{-i^\circ})\in\mathcal{T}_{III}$ and Step 3. We thus have $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$. Then,

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\le\sum_{i\ne i^\circ}p_i(\boldsymbol{t})t_i+p_{i^\circ}(\boldsymbol{t})(t_{i^\circ}-c_{i^\circ})\le t_{i^\circ}-c_{i^\circ}=\sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i-q_i^\star(\boldsymbol{t})c_i),$$

where the first inequality holds because $q_i(\boldsymbol{t})$ and c_i are non-negative and $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$, the second inequality follows from (FC) and that $t_{i^\circ}-c_{i^\circ}>\underline{\mu}_{i^\star}$ and $t_i\le\underline{\mu}_{i^\star}$ for all $i\in\mathcal{I}\setminus\{i^\circ\}$, and the equality follows from the definition of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. As scenario $\boldsymbol{t}$ is chosen arbitrarily, $(\boldsymbol{p},\boldsymbol{q})$ generates a payoff less than or equal to that of from $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ throughout $\mathcal{T}_{IV}^1$.

For an arbitrary $i^\circ\in\mathcal{I}\setminus\{i^\star\}$, we now define the set $\mathcal{T}'(i^\circ)=\prod_{i\in\mathcal{I}}\mathcal{T}_i'$, where $\mathcal{T}_{i^\star}'=\{\underline{\mu}_{i^\star}\}$, $\mathcal{T}_{i^\circ}'=\mathcal{T}_{i^\circ}$ and $\mathcal{T}_i'=[\underline{t}_i,\underline{\mu}_{i^\star}]$ for all $i\in\mathcal{I}\setminus\{i^\star,i^\circ\}$. Note that $\mathcal{T}'(i^\circ)\subseteq\mathcal{T}_{III}\cup\mathcal{T}_{IV}^1$. By Step 3 and the findings of Step 4 thus far, the payoff of $(\boldsymbol{p},\boldsymbol{q})$ cannot be higher than that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ throughout $\mathcal{T}'(i^\circ)$. Denote by $\mathcal{T}_{IV}^1(i^\circ)$ the subset of $\mathcal{T}_{IV}^1$ where i° is the only agent whose type $t_{i^\circ}>\underline{\mu}_{i^\star}$, and note that $\cup_{i^\circ\in\mathcal{I}\setminus\{i^\star\}}\mathcal{T}_{IV}^1(i^\circ)=\mathcal{T}_{IV}^1$. We have $\mathcal{T}_{IV}^1(i^\circ)\subseteq\mathcal{T}'(i^\circ)$, and $\mathcal{T}'(i^\circ)$ satisfies the assumptions (i) and (ii) in Lemma 7. The payoffs of $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ thus coincide throughout $\mathcal{T}_{IV}^1(i^\circ)$ by Lemma 7. As we have $t_{i^\circ}-c_{i^\circ}>t_{i^\star}=\underline{\mu}_{i^\star}\ge t_i$ for all $i\in\mathcal{I}\setminus\{i^\circ,i^\star\}$ and $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$ for any $\boldsymbol{t}\in\mathcal{T}_{IV}^1(i^\circ)$, the payoff of $(\boldsymbol{p},\boldsymbol{q})$ can match the payoff $\max_{i\in\mathcal{I}} t_i-c_i=t_{i^\circ}-c_{i^\circ}$ of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ only if $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})=1$. We thus established (18) in $\mathcal{T}_{IV}^1(i^\circ)$. As agent i° is chosen arbitrarily, the claim holds throughout $\mathcal{T}_{IV}^1$.

As for the induction step, assume that (18) holds throughout $\mathcal{T}_{IV}^n$ for some $n\ge 1$, and fix a scenario $\boldsymbol{t}\in\mathcal{T}_{IV}^{n+1}$. Thus, there exists exactly $n+1$ agents $i\ne i^\star$ that satisfy $t_i>\underline{\mu}_{i^\star}$. For any agent $i\in\mathcal{I}(\boldsymbol{t})$, incentive compatibility dictates that $p_i(\boldsymbol{t})-q_i(\boldsymbol{t})\le p_i(\underline{t}_i,\boldsymbol{t}_{-i})=0$, where the equality follows from $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_{III}\cup\mathcal{T}_{IV}^n$. Indeed, if $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_{III}$, then agent $i^\star\notin\mathcal{I}(\boldsymbol{t})$ receives the good so that and $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, and if $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}^n_{IV}$, the equality follows from the induction hypothesis. We thus have $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})$ and, by the definition of $\mathcal{T}_{IV}$, in particular for all $i \in \arg\max_{j\in\mathcal{I}} t_j - c_j$. Then,

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \le \sum_{i\notin\mathcal{I}(\boldsymbol{t})} p_i(\boldsymbol{t})t_i + \sum_{i\in\mathcal{I}(\boldsymbol{t})} p_i(\boldsymbol{t})(t_i - c_i) \le \max_{i\in\mathcal{I}} t_i - c_i = \sum_{i\in\mathcal{I}}(p^\star_i(\boldsymbol{t})t_i - q^\star_i(\boldsymbol{t})c_i),$$

where the first inequality holds because $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})$ and $q_i(\boldsymbol{t})$ and c_i are non-negative, the second inequality follows from (FC) and that $\max_{j\in\mathcal{I}} t_j - c_j > \underline{\mu}_{i^\star} = t_{i^\star} \ge t_i$ for all $i \notin \mathcal{I}(\boldsymbol{t})$, and the equality follows from the definition of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Thus, the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ is less than or equal to the payoff of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in $\mathcal{T}^{n+1}_{IV}$.

For an arbitrary $\mathcal{I}' \subseteq \mathcal{I} \setminus \{i^\star\}$ with $|\mathcal{I}'| = n+1$, we now define $\mathcal{T}'(\mathcal{I}') = \prod_{i\in\mathcal{I}} \mathcal{T}'_i$, where $\mathcal{T}'_{i^\star} = \{\underline{\mu}_{i^\star}\}$, $\mathcal{T}'_i = \mathcal{T}_i$ for all $i \in \mathcal{I}'$, and $\mathcal{T}'_i = [\underline{t}_i, \underline{\mu}_{i^\star}]$ for all $i \in \mathcal{I} \setminus (\mathcal{I}' \cup \{i^\star\})$. Note that $\mathcal{T}'(\mathcal{I}') \subseteq \mathcal{T}_{III} \cup \bigcup_{k=1}^{n+1} \mathcal{T}^k_{IV}$. By Step 3 and the findings of this step thus far, the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ cannot be higher than that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ throughout $\mathcal{T}'(\mathcal{I}')$. Denote by $\mathcal{T}^{n+1}_{IV}(\mathcal{I}')$ the subset of $\mathcal{T}_{IV}$ where $t_i \le \underline{\mu}_{i^\star}$ for all $i \notin \mathcal{I}'$. Note that the union of $\mathcal{T}^{n+1}_{IV}(\mathcal{I}')$ over all $\mathcal{I}' \subseteq \mathcal{I} \setminus \{i^\star\}$ with $|\mathcal{I}'| = n+1$ gives us the set $\cup_{k=1}^{n+1} \mathcal{T}^k_{IV}$, and $\mathcal{T}^{n+1}_{IV}(\mathcal{I}') \subseteq \mathcal{T}'(\mathcal{I}')$. As the set $\mathcal{T}'(\mathcal{I}')$ satisfies the assumptions (i) and (ii) in Lemma 7, $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generate the same payoff throughout $\mathcal{T}^{n+1}_{IV}(\mathcal{I}')$. As we have $\max_{j\in\mathcal{I}} t_j - c_j > \underline{\mu}_{i^\star} \ge t_i$ for all $i \notin \mathcal{I}'$ and $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}'$ for any $\boldsymbol{t} \in \mathcal{T}^{n+1}_{IV}(\mathcal{I}')$, mechanism $(\boldsymbol{p}, \boldsymbol{q})$ can match the payoff $\max_{i\in\mathcal{I}} t_i - c_i$ of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ only if (18) holds in $\mathcal{T}^{n+1}_{IV}(\mathcal{I}')$. As $\mathcal{I}'$ is chosen arbitrarily, (18) holds throughout $\mathcal{T}^{n+1}_{IV}$. This observation completes the induction step.

Step 5 ($\mathcal{T}_V$). In analogy to Step 4, we will show that (18) holds for every $\boldsymbol{t} \in \mathcal{T}_V$. This immediately implies that $(\boldsymbol{p}, \boldsymbol{q})$ generates the same payoff as $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in $\mathcal{T}_V$ and, consequently, throughout $\mathcal{T}$. To this end, define the set-valued function $\mathcal{I}(\boldsymbol{t}) = \{i \in \mathcal{I} : t_i > \underline{\mu}_{i^\star}\}$ for $\boldsymbol{t} \in \mathcal{T}_V$. Note that $|\mathcal{I}(\boldsymbol{t})| \ge 1$ for any $\boldsymbol{t} \in \mathcal{T}_V$ thanks to the definition of $\mathcal{T}_V$, which implies that $\max_{i\neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}$. We will prove by induction that (18) holds for all type profiles in $\mathcal{T}^n_V = \{\boldsymbol{t} \in \mathcal{T}_V : |\mathcal{I}(\boldsymbol{t}) \setminus \{i^\star\}| = n\}$ for all $n = 1, 2, \ldots, I-1$. Note that in any $\boldsymbol{t} \in \mathcal{T}^n_V$ there are n agents, each of which is different from $i^\star$, whose types exceed $\underline{\mu}_{i^\star}$. Agent $i^\star$'s type may or may not take a value above $\underline{\mu}_{i^\star}$.

As for the base step, set $n = 1$ and fix any scenario $\boldsymbol{t} \in \mathcal{T}^1_V$. Thus, there is exactly one agent $i^\circ \neq i^\star$ that satisfy $t_{i^\circ} > \underline{\mu}_{i^\star}$. Incentive compatibility ensures that $p_{i^\circ}(\boldsymbol{t}) - q_{i^\circ}(\boldsymbol{t}) \le p_{i^\circ}(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) = 0$, where the equality follows from that $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$ and from Steps 1, 2 and 3. Similarly for agent $i^\star$, we have $p_{i^\star}(\boldsymbol{t}) - q_{i^\star}(\boldsymbol{t}) \le p_{i^\star}(\underline{\mu}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 0$, where the equality follows from that $(\underline{\mu}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_{IV}$ and Step 4. We thus have $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t})$ and $i^\star$, which may or may not be in $\mathcal{I}(\boldsymbol{t})$. Then, we have

$$\begin{aligned}\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) &\le \sum_{i\in\mathcal{I}\setminus\mathcal{I}(\boldsymbol{t})\cup\{i^\star\}} p_i(\boldsymbol{t})t_i + \sum_{i\in\mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}} p_i(\boldsymbol{t})(t_i - c_i) + p_{i^\star}(\boldsymbol{t})(t_{i^\star} - c_{i^\star}) \\ &\le \max_{i\in\mathcal{I}} t_i - c_i = \sum_{i\in\mathcal{I}}(p^\star_i(\boldsymbol{t})t_i - q^\star_i(\boldsymbol{t})c_i),\end{aligned}$$

where the first equality holds because $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}(\boldsymbol{t}) \cup \{i^\star\}$ and $q_i(\boldsymbol{t})$ and c_i are non-negative, the second inequality from (FC) and that $\max_{j\in\mathcal{I}} t_j - c_j = \max_{j\in\mathcal{I}(\boldsymbol{t})\cup\{i^\star\}} t_j - c_j > \underline{\mu}_{i^\star} \ge t_i$ for all $i \in \mathcal{I} \setminus \mathcal{I}(\boldsymbol{t}) \cup \{i^\star\}$, and the equality follows from the definition of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Thus, $(\boldsymbol{p}, \boldsymbol{q})$ cannot generate a payoff higher than that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ throughout $\mathcal{T}^1_V$. Recalling the findings in Steps 1–4, this relation between the payoffs holds true for $\mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}_{IV} \cup \mathcal{T}^1_V$.

For an arbitrary $i^\circ \in \mathcal{I} \setminus \{i^\star\}$, we define the set $\mathcal{T}'(i^\circ) = \prod_{i\in\mathcal{I}} \mathcal{T}'_i$, where $\mathcal{T}'_i = \mathcal{T}_i$ for all $i \in \{i^\star, i^\circ\}$, and $\mathcal{T}'_i = [\underline{t}_i, \underline{\mu}_{i^\star}]$ for all $i \in \mathcal{I} \setminus \{i^\star, i^\circ\}$. Note that $\mathcal{T}'(i^\circ) \subseteq \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}^1_{IV} \cup \mathcal{T}^1_V$, an therefore the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ cannot be higher than that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ throughout $\mathcal{T}'(i^\circ)$. Denote by $\mathcal{T}^1_V(i^\circ)$ the subset of $\mathcal{T}^1_V$ where i° is the only agent among $\mathcal{I} \setminus \{i^\star\}$ with type $t_{i^\circ} > \underline{\mu}_{i^\star}$. Note that $\cup_{i^\circ\in\mathcal{I}\setminus\{i^\star\}} \mathcal{T}^1_V(i^\circ) = \mathcal{T}^1_V$, and $\mathcal{T}^1_V(i^\circ) \subseteq \mathcal{T}'(i^\circ)$. As the payoff of $(\boldsymbol{p}, \boldsymbol{q})$ cannot be higher than that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ throughout

$\mathcal{T}'(i^\circ)$, and the set $\mathcal{T}'$ satisfies the assumptions (i) and (ii) from Lemma 7, the payoffs of $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must coincide throughout $\mathcal{T}_V^1(i^\circ)$. As we know that $\max_{i\in\{i^\circ,i^\star\}} t_i - c_i > \underline{\mu}_{i^\star} \geq t_j$ for all $j\notin\{i^\circ,i^\star\}$ and $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i\in\{i^\circ,i^\star\}$ in any $\boldsymbol{t}\in\mathcal{T}_V^1(i^\circ)$, mechanism $(\boldsymbol{p},\boldsymbol{q})$ can match the payoff $\max_{i\in\{i^\circ,i^\star\}} t_i - c_i$ of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ only if (18) holds. As agent i° was chosen arbitrarily, we conclude that (18) holds throughout $\mathcal{T}_V^1$.

As for the induction step, assume that (18) holds throughout $\mathcal{T}_V^n$ for some $n\geq 1$ and fix any scenario $\boldsymbol{t}\in\mathcal{T}_V^{n+1}$. For any agent $i\in\mathcal{I}(\boldsymbol{t})\setminus\{i^\star\}$, incentive compatibility implies that $p_i(\boldsymbol{t}) - q_i(\boldsymbol{t}) \leq p_i(\underline{t}_i,\boldsymbol{t}_{-i}) = 0$, where the equality holds because $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}\cup\mathcal{T}_V^n$. Indeed, if $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}$, then $p_i(\underline{t}_i,\boldsymbol{t}_{-i})=0$ follows from Steps 1, 2 and 3, and if $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{T}_V^n$, then the equality follows from the induction hypothesis. If $i^\star\in\mathcal{I}(\boldsymbol{t})$, then incentive compatibility implies $p_{i^\star}(\boldsymbol{t}) - q_{i^\star}(\boldsymbol{t}) \leq p_{i^\star}(\underline{\mu}_{i^\star},\boldsymbol{t}_{-i^\star}) = 0$, where the equality follows from that $(\underline{\mu}_{i^\star},\boldsymbol{t}_{-i^\star})\in\mathcal{T}_{IV}$ and Step 4. We thus have $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i\in\mathcal{I}(\boldsymbol{t})$, and by the definition of $\mathcal{T}_V$, in particular for all $i\in\arg\max_{j\in\mathcal{I}} t_j - c_j$. Then, the principal's payoff in $\boldsymbol{t}$ can be written as:

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq \sum_{i\notin\mathcal{I}(\boldsymbol{t})} p_i(\boldsymbol{t})t_i + \sum_{i\in\mathcal{I}(\boldsymbol{t})} p_i(\boldsymbol{t})(t_i - c_i) \leq \max_{i\in\mathcal{I}} t_i - c_i = \sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the first inequality follows because $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i\in\mathcal{I}(\boldsymbol{t})$ and $c_i>0$ for all $i\notin\mathcal{I}(\boldsymbol{t})$. The second inequality holds because the two sums represent a weighted average of $t_i - c_i$ for $i\in\mathcal{I}(\boldsymbol{t})$ and t_i for $i\notin\mathcal{I}(\boldsymbol{t})$. All these terms are smaller or equal to $\max_{i\in\mathcal{I}} t_i - c_i$. In particular, the definition of $\mathcal{T}_V$ ensures that $\max_{j\in\mathcal{I}} t_j - c_j > \underline{\mu}_{i^\star} \geq t_i$ for all $i\notin\mathcal{I}(\boldsymbol{t})$. This reasoning shows that the payoff from $(\boldsymbol{p},\boldsymbol{q})$ cannot be higher than $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in $\mathcal{T}_V^{n+1}$.

For an arbitrary $\mathcal{I}'\subseteq\mathcal{I}$ with $\mathcal{I}'\ni i^\star$ and $|\mathcal{I}'\setminus\{i^\star\}| = n+1$, we define the set $\mathcal{T}'(\mathcal{I}') = \prod_{i\in\mathcal{I}}\mathcal{T}_i'$, where $\mathcal{T}_i' = \mathcal{T}_i$ for all $i\in\mathcal{I}'$ and $\mathcal{T}_i' = [\underline{t}_i,\underline{\mu}_{i^\star}]$ for all $i\in\mathcal{I}\setminus\mathcal{I}'$. Note that $\mathcal{T}'(\mathcal{I}')\subseteq\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}\cup\mathcal{T}_{IV}^{n+1}\cup\mathcal{T}_V^{n+1}$, and therefore the payoff of $(\boldsymbol{p},\boldsymbol{q})$ cannot be higher than that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ throughout $\mathcal{T}'(\mathcal{I}')$. Denote by $\mathcal{T}_V^{n+1}(\mathcal{I}')$ the subset of $\mathcal{T}_V$ where $t_i\leq\underline{\mu}_{i^\star}$ for all $i\notin\mathcal{I}'$. Note that the union of $\mathcal{T}_V^{n+1}(\mathcal{I}')$ over all $\mathcal{I}'\subseteq\mathcal{I}$ with $\mathcal{I}'\ni i^\star$ and $|\mathcal{I}'\setminus\{i^\star\}| = n+1$ gives us the set $\cup_{k=1}^{n+1}\mathcal{T}_V^k$, and $\mathcal{T}_V^{n+1}(\mathcal{I}')\subseteq\mathcal{T}'(\mathcal{I}')$. The set $\mathcal{T}'(\mathcal{I}')$ satisfies the assumptions (i) and (ii) in Lemma 7, which implies that the payoffs from $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ must coincide throughout $\mathcal{T}_V^{n+1}(\mathcal{I}')$. As we know that $\max_{i\in\mathcal{I}} t_i - c_i > \underline{\mu}_{i^\star} \geq t_j$ for all $j\notin\mathcal{I}'$ and $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i\in\mathcal{I}'$ for any $\boldsymbol{t}\in\mathcal{T}_V^{n+1}(\mathcal{I}')$, mechanism $(\boldsymbol{p},\boldsymbol{q})$ can match the payoff $\max_{i\in\mathcal{I}} t_i - c_i$ only if (18) holds in $\mathcal{T}_V^{n+1}(\mathcal{I}')$. As $\mathcal{I}'$ was chosen arbitrarily, (18) holds throughout $\mathcal{T}_V^{n+1}$. This observation completes the induction step.

The above reasoning shows that the principal's payoff from $(\boldsymbol{p},\boldsymbol{q})$ and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ coincide throughout the entire type space $\mathcal{T}$. □