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, [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.*, [19, 20] and the references therein) sustains the modeling assumptions of [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.*, [12, 26]), 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. Ambiguity sets of these three classes are parsimonious, that is, they require only limited prior distributional information such as knowledge of the minimal, maximal and most likely (or expected) types of all agents. Indeed, as the allocation problems studied in this paper are expected to occur infrequently, these indicators are the only properties of the unknown type distribution that can reasonable be extracted from scarce data, expert knowledge or common sense reasoning. Support-only and Markov ambiguity sets are widely studied in contexts where data is scarce (see, *e.g.*, [6, 12]). In particular, they are successfully used in other mechanism design problems (see, *e.g.*, [3, 8, 17, 22, 25]).

Pareto robust optimality is an important solution concept in robust optimization [16]. 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 [16]. This phenomenon is particularly pronounced in *adjustable* robust optimization [7]. 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. While adjustable robust optimization problems are generically NP-hard, we will show that the mechanism design problem at hand can be solved in closed form. However, we will also see that this problem suffers 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 establish a new sufficient condition for Pareto robust optimality, which does not depend on any specifics of the mechanism design problem at hand and may therefore be useful for other applications, and we develop a new *spatial induction* technique for checking this condition. Spatial induction 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 that a judiciously chosen favorite-agent mechanism satisfies our new sufficient condition for Pareto robust optimality.

On a high level, our proof strategy can be explained 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 establish a new sufficient condition for Pareto robust optimality, which may be useful beyond distributionally robust mechanism design. We also develop a new spatial induction technique for proving that the above favored-agent mechanisms satisfy our sufficient condition and are therefore Pareto robustly optimal. This technique crucially exploits the non-locality of the incentive compatibility constraints of the distributionally robust 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 [23], 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 [20] 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 [19] accounts both for costly verification and for limited penalties, thereby unifying the models in [5] and [20]. Chua et al. [11] 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, 8, 10, 18, 21, 25]), auction design (see, *e.g.*, [1, 3, 14, 17, 22]) or contracting (see, *e.g.*, [24]). 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. [9], 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, [9] 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. Section 6 assesses the performance of the proposed mechanisms numerically. Conclusions are drawn in Section 7, and all proofs are relegated to the online 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.*, [13]). 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 $\tilde{\boldsymbol{t}}=\boldsymbol{t}$ must grow proportionally to $\text{prob}_i(a_i;\boldsymbol{t},\boldsymbol{a}_{-i})$ under *any* non-decreasing utility function. In 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 principal 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. In Section A, we show, by restricting attention to mechanisms that admit an ex-post Nash equilibrium, that the principal also hedges against uncertainty about the agents' attitude towards 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

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

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

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

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.

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

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 $\mathcal{P}$. Recall also that a maximization (minimization) problem is called *solvable* if its supremum (infimum) is attained by a feasible solution. Thus, problem (MDP) is solvable if there is $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)\in\mathcal{X}$ with worst-case expected payoff $z^\star$.

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 optimal but perform also well under *all* type distributions in the ambiguity set $\mathcal{P}$. More precisely, we hope to identify an 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 [16].

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. Figure 1 illustrates the allocations of a favored-agent mechanism in the special case when there are only two agents.

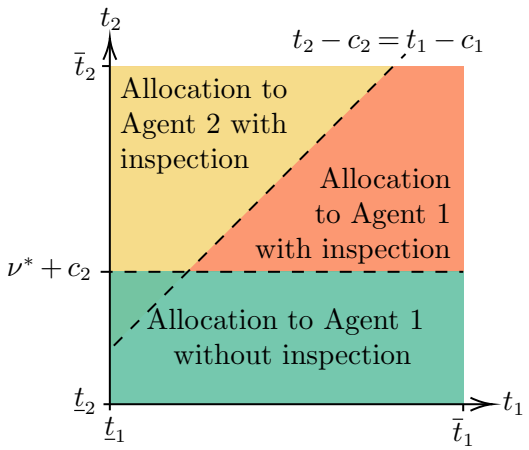

Figure 1 A favored-agent mechanism for two agents with favored agent $i^\star = 1$ and threshold value $\nu^\star$.

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 Ben-Porath et al. [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

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

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. Before studying specific ambiguity sets, we formally introduce the basics of spatial induction, which is our main tool for proving Pareto robust optimality. To facilitate a concise presentation, we first define the concept of unilateral reachability.

Definition 4 (Unilateral Reachability). For any fixed agent $i\in\mathcal{I}$, a scenario $\boldsymbol{t}'\in\mathcal{T}$ is called *i-unilaterally reachable* from another scenario $\boldsymbol{t}\in\mathcal{T}$ if $\boldsymbol{t}'_{-i}=\boldsymbol{t}_{-i}$.

Note that $\boldsymbol{t}'\in\mathcal{T}$ is i-unilaterally reachable from $\boldsymbol{t}\in\mathcal{T}$ if $\boldsymbol{t}'$ is obtained by changing the type t_i of agent i to t_i' while keeping the types $\boldsymbol{t}_{-i}$ of all other agents fixed.

Lemma 1. *The following hold for any $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$, $i\in\mathcal{I}$ and $\boldsymbol{t}\in\mathcal{T}$.*

(i) If $p_i(\boldsymbol{t})-q_i(\boldsymbol{t})=1$, then the mechanism $(\boldsymbol{p},\boldsymbol{q})$ allocates the good to agent i with probability 1 in any scenario that is i-unilaterally reachable from $\boldsymbol{t}$, i.e., $p_i(t_i',\boldsymbol{t}_{-i})=1$ for all $t_i'\in\mathcal{T}_i$.

(ii) If $p_i(\boldsymbol{t})=0$, then the mechanism $(\boldsymbol{p},\boldsymbol{q})$ inspects agent i whenever he wins the good in any scenario that is i-unilaterally reachable from $\boldsymbol{t}$, i.e., $p_i(t_i',\boldsymbol{t}_{-i})=q_i(t_i',\boldsymbol{t}_{-i})$ for all $t_i'\in\mathcal{T}_i$.

The proof of Lemma 1 follows directly from the (IC) and (FC) constraints. Note that, by (FC), the condition $p_i(\boldsymbol{t})-q_i(\boldsymbol{t})=1$ in Lemma 1(i) is satisfied if and only if $p_i(\boldsymbol{t})=1$ and $q_i(\boldsymbol{t})=0$. Thus, Lemma 1(i) asserts that if we allocate the good to agent i without inspection in scenario $\boldsymbol{t}$, then we should also allocate the good to agent i in any i-unilaterally reachable scenario $\boldsymbol{t}'$. However, we may or may not inspect agent i in scenario $\boldsymbol{t}'$. Lemma 1(ii) asserts that if agent i does not receive the good in scenario $\boldsymbol{t}$ (which implies via (FC) that agent i is *not* inspected in scenario $\boldsymbol{t}$), then we must inspect agent i in any i-unilaterally reachable scenario $\boldsymbol{t}'$ in which he receives the good.

The spatial induction technique to be developed in this paper critically relies on the following proposition, which introduces a sufficient condition for the Pareto robust optimality of a mechanism.

Proposition 1. *An optimal mechanism $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)\in\mathcal{X}$ for problem (MDP) is Pareto robustly optimal if there exists a partition $\mathcal{S}_1,\ldots,\mathcal{S}_m$ of the type space $\mathcal{T}$ such that the following conditions hold for any index $k\in\{1,\ldots,m\}$ and for any scenario $\boldsymbol{t}\in\mathcal{S}_k$.*

(i) There exists a probability distribution $\mathbb{P}\in\mathcal{P}\cap\mathcal{P}_0(\cup_{l=1}^k\mathcal{S}_l)$ with $\mathbb{P}(\tilde{\boldsymbol{t}}=\boldsymbol{t})>0$.

(ii) The mechanism $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves the following auxiliary scenario problem.

$$\begin{array}{ll}\max\limits_{(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}} & \sum\limits_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\\ \text{s.t.} & \sum\limits_{i\in\mathcal{I}}(p_i(\boldsymbol{t}')t_i'-q_i(\boldsymbol{t}')c_i)=\sum\limits_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t}')t_i'-q_i^\star(\boldsymbol{t}')c_i)\quad\forall\boldsymbol{t}'\in\cup_{l=1}^{k-1}\mathcal{S}_l\end{array}\tag{SP$_k(\boldsymbol{t})$}$$

Note that problem SP$_k(\boldsymbol{t})$ not only depends on k and $\boldsymbol{t}$ but also on the prescribed optimal mechanism $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. However, this dependence is notationally suppressed to avoid clutter.

Proposition 1 establishes a new sufficient condition for Pareto robust optimality that does not exploit any structural properties of the ambiguity set $\mathcal{P}$. We also emphasize that the proof of Proposition 1 does not exploit any specifics of the mechanism design problem (MDP) and thus readily extends to generic distributionally robust optimization problems. Thanks to Proposition 1, the Pareto robust optimality of a given optimal mechanism can be proved by identifying a partition $\mathcal{S}_1, \ldots, \mathcal{S}_m$ of $\mathcal{T}$ that satisfies two simple conditions. First, for any fixed $k \in \{1, \ldots, m\}$ and $\boldsymbol{t} \in \mathcal{S}_k$, there must exist an admissible distribution that assigns $\boldsymbol{t}$ a strictly positive probability, while assigning zero probability to $\mathcal{S}_l$ for every $l > k$. Second, the given mechanism should solve the scenario problems $\mathrm{SP}_k(\boldsymbol{t})$ simultaneously for all $k \in \{1, \ldots, m\}$ and $\boldsymbol{t} \in \mathcal{S}_k$. Conditions (i) and (ii) imply that the given mechanism cannot be Pareto robustly dominated by any other mechanism. We emphasize, however, that a mechanism satisfying the conditions (i) and (ii) of Proposition 1 is not necessarily robustly optimal. Thus, the conditions (i) and (ii) are not sufficient to guarantee Pareto robust optimality unless we restrict attention to robustly optimal mechanisms.

In each of the subsequent sections, we will leverage Proposition 1 to show that a given candidate mechanism is Pareto robustly optimal for a particular ambiguity set. Specifically, we will choose a partition $\mathcal{S}_1, \ldots, \mathcal{S}_m$ of the type space tailored to the given ambiguity set such that condition (i) is satisfied. We will then apply the following spatial induction technique. First, we will show that the given candidate mechanism solves the scenario problems $\mathrm{SP}_1(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{S}_1$. Next, we will exploit the non-locality of the incentive compatibility constraints—as manifested through Lemma 1—to relate any scenario in $\mathcal{S}_2$ to a scenario in $\mathcal{S}_1$. This will allow us to prove that the given candidate mechanism solves the scenario problems $\mathrm{SP}_2(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{S}_2$. We then iterate through the remaining subsets of the type space one by one and apply each time a similar induction step.

The following corollary shows that the sufficient condition of Proposition 1 implies a payoff equivalence principle, which will be useful to prove the main results of this paper. The proof of this corollary follows immediately from that of Proposition 1 and is thus omitted.

Corollary 1 **(Payoff Equivalence)**. *Assume that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star) \in \mathcal{X}$ is an optimal mechanism satisfying the conditions (i) and (ii) of Proposition 1. If any other mechanism $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, then it generates the same payoff as $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in every scenario $\boldsymbol{t} \in \mathcal{T}$.*

Recall that if $(\boldsymbol{p}, \boldsymbol{q})$ weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, then the *expected* payoff of $(\boldsymbol{p}, \boldsymbol{q})$ is at least as large as that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ under any distribution $\mathbb{P} \in \mathcal{P}$. Corollary 1 additionally asserts that, if $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ satisfies the sufficient condition of Proposition 1, then the *actual* payoff of $(\boldsymbol{p}, \boldsymbol{q})$ is at least as large as (in fact, exactly equal to) that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ under any scenario $\boldsymbol{t} \in \mathcal{T}$.

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

$$\begin{array}{rll} z^\star = & \sup\limits_{\boldsymbol{p},\boldsymbol{q}} & \inf\limits_{\boldsymbol{t}\in\mathcal{T}} \sum\limits_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \\ & \text{s.t.} & p_i, q_i \in \mathcal{L}(\mathcal{T}, [0,1]) \;\; \forall i \in \mathcal{I} \\ & & \text{(IC), (FC).} \end{array} \tag{3}$$

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

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

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 2. *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 2, 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 $\nu^\star > \max_{i\in\mathcal{I}} \overline{t}_i - c_i$. 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 $\hat{t}_{i^\star} = \overline{t}_{i^\star}$, and note that $\hat{\boldsymbol{t}} \neq \underline{\boldsymbol{t}}$ because $\underline{t}_{i^\star} < \overline{t}_{i^\star}$. 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

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

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 2 below that $(\boldsymbol{p}, \boldsymbol{q}')$ is also *not* a convex combination of favored-agent mechanisms.

LEMMA 2. *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 $\nu^\star = \max_{i\in\mathcal{I}} \underline{t}_i = \underline{t}_{i^\star}$ 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 $\mathcal{T}$.

$$\begin{aligned}
\mathcal{S}_1 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 \leq \underline{t}_1 \text{ and } t_1 = \bar{t}_1\} \\
\mathcal{S}_2 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 \leq \underline{t}_1 \text{ and } t_1 < \bar{t}_1\} \\
\mathcal{S}_3 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 > \underline{t}_1 \text{ and } t_1 = \underline{t}_1\} \\
\mathcal{S}_4 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 > \underline{t}_1 \text{ and } t_1 > \underline{t}_1\}
\end{aligned} \tag{4}$$

The sets $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$ are visualized in Figure 2. 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.

Denote by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ the favored-agent mechanism of type (i) with favored agent 1 and threshold value $\nu^\star = \underline{t}_1$. By definition, the principal's payoff in scenario $\boldsymbol{t}$ under $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ thus amounts to t_1 when $t_2 - c_2 \leq \underline{t}_1$ (*i.e.*, when $\boldsymbol{t} \in \mathcal{S}_1 \cup \mathcal{S}_2$) and to $\max_{i\in\mathcal{I}} t_i - c_i$ when $t_2 - c_2 > \underline{t}_1$ (*i.e.*, when $\boldsymbol{t} \in \mathcal{S}_3 \cup \mathcal{S}_4$). In the following, we will prove that this mechanism is Pareto robustly optimal in problem (3). To this end, we will leverage Proposition 1, which provides a sufficient condition for the Pareto robust optimality of robustly optimal mechanisms. From Theorem 1 we already know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is robustly optimal. To show that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is Pareto robustly optimal, it thus suffices to verify the conditions (i) and (ii) in Proposition 1 for the partition (4). Note first that condition (i) trivially holds because the support-only ambiguity set contains all Dirac point distributions concentrating unit mass at some scenario $\boldsymbol{t} \in \mathcal{T}$. It thus remains to verify condition (ii). To this end, we will show by induction on k that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{S}_k$. The induction steps exploit the non-locality of the incentive compatibility constraint (IC), which led to Lemma 1.

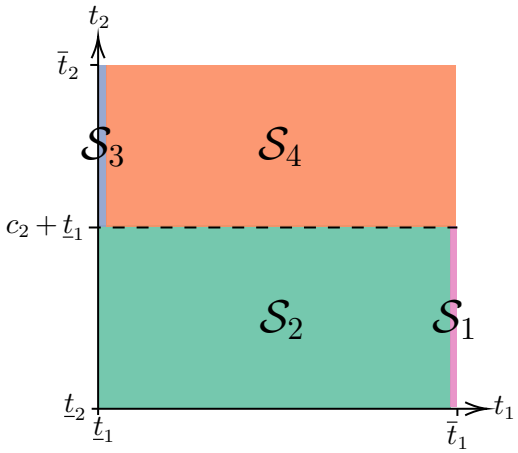

Figure 2 Partition (4) of the type space $\mathcal{T}$.

As for the base step corresponding to $k=1$, note that any $(\boldsymbol{p},\boldsymbol{q})\in\mathcal{X}$ is feasible in $\mathrm{SP}_1(\boldsymbol{t})$ for any $\boldsymbol{t}\in\mathcal{S}_1$. The objective function value of $(\boldsymbol{p},\boldsymbol{q})$ is dominated by that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ in $\mathrm{SP}_1(\boldsymbol{t})$ because

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

where the first inequality holds because $c_i>0$ for all $i\in\mathcal{I}$, while the second inequality follows from our simplifying assumption that $\bar{t}_1>c_2+\underline{t}_1$ and the definition of $\mathcal{S}_1$, which imply that $t_2\leq\underline{t}_1+c_2<\bar{t}_1=t_1$. Thus, $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_1(\boldsymbol{t})$. As a byproduct, we have shown that any mechanism $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_1(\boldsymbol{t})$ can match the payoff t_1 of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ only if $p_1(\boldsymbol{t})=1$ and $q_1(\boldsymbol{t})=0$ because $t_2<t_1$, $c_i>0$ and $(\boldsymbol{p},\boldsymbol{q})$ satisfies the (FC) constraints $\sum_{i\in\mathcal{I}}p_i(\boldsymbol{t})\leq 1$ and $q_i(\boldsymbol{t})\geq 0$.

As for the first induction step corresponding to $k=2$, consider any $\boldsymbol{t}\in\mathcal{S}_2$ and $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_2(\boldsymbol{t})$. From the base step we know that $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_1(\boldsymbol{t}')$ for all $\boldsymbol{t}'\in\mathcal{S}_1$. The constraints of $\mathrm{SP}_2(\boldsymbol{t})$ thus ensure that $(\boldsymbol{p},\boldsymbol{q})$ solves $\mathrm{SP}_1(\boldsymbol{t}')$ for all $\boldsymbol{t}'\in\mathcal{S}_1$, too. The reasoning in the base step further implies that $p_1(\boldsymbol{t}')=1$ and $q_1(\boldsymbol{t}')=0$ for all $\boldsymbol{t}'\in\mathcal{S}_1$. As any scenario $\boldsymbol{t}\in\mathcal{S}_2$ is 1-unilaterally reachable from $(\bar{t}_1,t_2)\in\mathcal{S}_1$, and as $p_1(\bar{t}_1,t_2)-q_1(\bar{t}_1,t_2)=1$, Lemma 1(i) implies that $p_1(\boldsymbol{t})=1$. The objective function value of $(\boldsymbol{p},\boldsymbol{q})$ in $\mathrm{SP}_2(\boldsymbol{t})$ thus amounts to $t_1-q_1(\boldsymbol{t})c_1$ and is bounded above by $t_1=\sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i-q_i^\star(\boldsymbol{t})c_i)$ because $c_1\geq 0$. Therefore, $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_2(\boldsymbol{t})$. In addition, as $c_1>0$, a mechanism $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_2(\boldsymbol{t})$ can attain the optimal payoff t_1 only if $p_1(\boldsymbol{t})=1$ and $q_1(\boldsymbol{t})=0$.

Next, set $k=3$, and consider any $\boldsymbol{t}\in\mathcal{S}_3$ and $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_3(\boldsymbol{t})$. The constraints of $\mathrm{SP}_3(\boldsymbol{t})$ ensure that $(\boldsymbol{p},\boldsymbol{q})$ solves $\mathrm{SP}_1(\boldsymbol{t}')$ for all $\boldsymbol{t}'\in\mathcal{S}_1$ as well as $\mathrm{SP}_2(\boldsymbol{t}')$ for all $\boldsymbol{t}'\in\mathcal{S}_2$. Our earlier reasoning also implies that $p_1(\boldsymbol{t}')=1$ and $q_1(\boldsymbol{t}')=0$ for all $\boldsymbol{t}'\in\mathcal{S}_1\cup\mathcal{S}_2$. As any $\boldsymbol{t}\in\mathcal{S}_3$ is 2-unilaterally reachable from $(\underline{t}_1,\underline{t}_2)\in\mathcal{S}_2$, and as $p_1(\underline{t}_1,\underline{t}_2)=1$ and $p_2(\underline{t}_1,\underline{t}_2)=0$, Lemma 1(ii) ensures that $p_2(\boldsymbol{t})=q_2(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p},\boldsymbol{q})$ in $\mathrm{SP}_3(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ because

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

where the first inequality holds because $c_1\geq 0$. The second inequality exploits (FC) and the definition of $\mathcal{S}_3$, which implies that $t_2-c_2>\underline{t}_1=t_1$. Thus, $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_3(\boldsymbol{t})$. As $t_2-c_2>\underline{t}_1=t_1$, the mechanism $(\boldsymbol{p},\boldsymbol{q})$ can attain the optimal payoff t_2-c_2 of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ only if $p_2(\boldsymbol{t})=q_2(\boldsymbol{t})=1$.

Finally, set $k=4$, and consider any $\boldsymbol{t}\in\mathcal{S}_4$ and $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_4(\boldsymbol{t})$. The constraints of $\mathrm{SP}_4(\boldsymbol{t})$ ensure that $(\boldsymbol{p},\boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for all $\boldsymbol{t}'\in\mathcal{S}_l$ and $l=1,2,3$. Our earlier reasoning also implies that $p_1(\boldsymbol{t}')=1$ and $q_1(\boldsymbol{t}')=0$ for all $\boldsymbol{t}'\in\mathcal{S}_1\cup\mathcal{S}_2$ and that $p_2(\boldsymbol{t}')=q_2(\boldsymbol{t}')=1$ for all $\boldsymbol{t}'\in\mathcal{S}_3$. As $p_1(\boldsymbol{t}')=1$ and thus $p_2(\boldsymbol{t}')=0$ for all $\boldsymbol{t}'\in\mathcal{S}_1\cup\mathcal{S}_2$, we can use Lemma 1(ii) and similar arguments as before to

conclude that $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$. Also, as any scenario $\boldsymbol{t} \in \mathcal{S}_4$ is 1-unilaterally reachable from $(\underline{t}_1, t_2) \in \mathcal{S}_3$, and as $p_2(\underline{t}_1, t_2) = 1$, which implies that $p_1(\underline{t}_1, t_2) = 0$, Lemma 1(ii) ensures that $p_1(\boldsymbol{t}) = q_1(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_4(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

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

Hence, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_4(\boldsymbol{t})$. In summary, we have thus verified condition (ii) of Proposition 1 by using spatial induction. This establishes the claim that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is Pareto robustly optimal.

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

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{T}) : \mathbb{E}_{\mathbb{P}}[\tilde{t}_i] \in [\underline{\mu}_i, \overline{\mu}_i] \ \forall i \in \mathcal{I} \right\}, \tag{5}$$

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 3. *If $\mathcal{P}$ is a Markov ambiguity set of the form (5), then problem (MDP) is solvable, and $z^\star = \max_{i\in\mathcal{I}} \underline{\mu}_i$.*

Contrasting Proposition 3 with Proposition 2 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 (5), 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 3, 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 (5), 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 (5). 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 two preliminary lemmas. We stress that, even though 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.

LEMMA 3. *If $\mathcal{P}$ is a Markov ambiguity set of the form (5) 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 $\max_{i\neq i^\star} \hat{t}_i < \underline{\mu}_{i^\star}$ 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) $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$.*

In the next lemma we show that if $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, then problem (MDP) admits a Pareto robustly optimal favored-agent mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Specifically, we leverage Lemmas 1 and 3 as well as the payoff equivalence principle of Corollary 1 to show 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* $\boldsymbol{t} \in \mathcal{T}$.

LEMMA 4. *Assume that $\mathcal{P}$ is a Markov ambiguity set of the form (5) 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*

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

The assumption that $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton will be relaxed in Theorem 4 below. To gain some intuition, we first sketch the proof of Lemma 4 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 3.

$$\begin{aligned}
\mathcal{S}_1 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 < \bar{t}_1,\ t_2 < \underline{\mu}_1 \text{ and } t_1 = \underline{\mu}_1\} \\
\mathcal{S}_2 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 < \bar{t}_1,\ t_2 < \underline{\mu}_1 \text{ and } t_1 \neq \underline{\mu}_1\} \\
\mathcal{S}_3 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 < \bar{t}_1,\ t_2 \geq \underline{\mu}_1 \text{ and } t_1 = \bar{t}_1\} \\
\mathcal{S}_4 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 < \bar{t}_1,\ t_2 \geq \underline{\mu}_1 \text{ and } t_1 < \bar{t}_1\} \\
\mathcal{S}_5 &= \{\boldsymbol{t} \in \mathcal{T} : t_2 - c_2 \geq \bar{t}_1\}
\end{aligned} \tag{6}$$

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.

Denote now by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ the type (ii) favored-agent mechanism with favored agent 1 and threshold value $\nu^\star = \bar{t}_1$. By construction, the principal's payoff in scenario $\boldsymbol{t}$ under $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ amounts to t_1 when $t_2 - c_2 < \bar{t}_1$ (*i.e.*, when $\boldsymbol{t} \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4$) and to $\max_{i\in\mathcal{I}} t_i - c_i$ when $t_2 - c_2 \geq \bar{t}_1$ (*i.e.*, when $\boldsymbol{t} \in \mathcal{S}_5$). To prove Lemma 4, it suffices to show that the conditions (i) and (ii) of Proposition 1 are satisfied for the partition (6). The claim then follows from Corollary 1. We will exploit Lemma 3

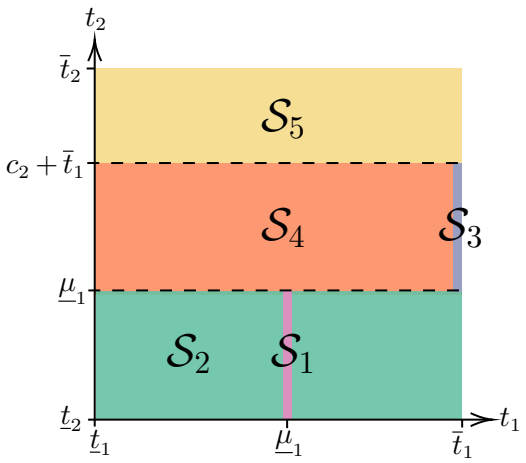

Figure 3 Partition (6) of the type space $\mathcal{T}$.

to verify condition (i). To verify condition (ii), we will use induction on k to show that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves the scenario problem $\mathrm{SP}_k(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{S}_k$. The induction step exploits Lemma 1.

We first show that condition (i) is satisfied, that is, for any $\boldsymbol{t} \in \mathcal{S}_k$ and $k \in \{1, \ldots, 5\}$ there exists $\mathbb{P} \in \mathcal{P}$ supported on $\cup_{l=1}^k \mathcal{S}_l$ with $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. Set first $k = 1$, and fix any $\boldsymbol{t} \in \mathcal{S}_1$. By Lemma 5, there exists 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 $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] = \underline{\mu}_i$ for all $i \in \mathcal{I}$, $\mathbb{P}(\tilde{\boldsymbol{t}} \in \{\boldsymbol{t}, \hat{\boldsymbol{t}}\}) = 1$, and $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. By the definition of $\mathcal{S}_1$, we have $t_{i^\star} = \underline{\mu}_{i^\star}$, and thus $\mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}] = \underline{\mu}_{i^\star}$ can hold only if $\hat{t}_{i^\star} = \underline{\mu}_{i^\star}$. As $\max_{i \neq i^\star} \hat{t}_i < \underline{\mu}_{i^\star}$, this means that $\hat{\boldsymbol{t}} \in \mathcal{S}_1$ and $\mathbb{P} \in \mathcal{P}_0(\mathcal{S}_1)$. Condition (i) thus holds for $k = 1$ and any $\boldsymbol{t} \in \mathcal{S}_1$. For any $k \in \{2, \ldots, 5\}$ and $\boldsymbol{t} \in \mathcal{S}_k$, condition (i) can easily be verified by using Lemma 3 and by noting that $\hat{\boldsymbol{t}} \in \mathcal{S}_1 \cup \mathcal{S}_2 = \{\boldsymbol{t} \in \mathcal{T} : \max_{i \neq i^\star} t_i < \underline{\mu}_{i^\star}\}$, which implies that the discrete distribution $\mathbb{P}$ is supported on $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_k$.

Next, we prove condition (ii) by induction on k. As for the base step corresponding to $k = 1$, note that any $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$ is feasible in $\mathrm{SP}_1(\boldsymbol{t})$ for any $\boldsymbol{t} \in \mathcal{S}_1$. As $t_2 < t_1$ and $c_i > 0$, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_1(\boldsymbol{t})$ is bounded above by $t_1 = \sum_{i \in \mathcal{I}} (p_i^\star(\boldsymbol{t}) t_i - q_i^\star(\boldsymbol{t}) c_i)$. Therefore, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_1(\boldsymbol{t})$. Also, a mechanism $(\boldsymbol{p}, \boldsymbol{q})$ can attain this bound only if $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$ because $t_2 < t_1$, $c_i > 0$ and $(\boldsymbol{p}, \boldsymbol{q})$ satisfies the (FC) constraints $\sum_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1$ and $q_i(\boldsymbol{t}) \geq 0$.

As for the first induction step corresponding to $k = 2$, consider any $\boldsymbol{t} \in \mathcal{S}_2$ and $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_2(\boldsymbol{t})$. From the base step we know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_1(\boldsymbol{t}')$ for all $\boldsymbol{t}' \in \mathcal{S}_1$. The constraints of $\mathrm{SP}_2(\boldsymbol{t})$ thus ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_1(\boldsymbol{t}')$ for all $\boldsymbol{t}' \in \mathcal{S}_1$, too. From the proof of the base step we further know that $p_1(\boldsymbol{t}') = 1$ and $q_1(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \mathcal{S}_1$. As any scenario $\boldsymbol{t} \in \mathcal{S}_2$ is 1-unilaterally reachable from $(\underline{\mu}_1, t_2) \in \mathcal{S}_1$, and as $p_1(\underline{\mu}_1, t_2) - q_1(\underline{\mu}_1, t_2) = 1$, Lemma 1(i) implies that $p_1(\boldsymbol{t}) = 1$. The objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_2(\boldsymbol{t})$ thus amounts to $t_1 - q_1(\boldsymbol{t}) c_1$ and is bounded above by t_1 because $c_1 > 0$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_2(\boldsymbol{t})$. In addition, as $c_1 > 0$, a mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_2(\boldsymbol{t})$ can attain the optimal payoff t_1 only if $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$.

Next, set $k = 3$, and consider any $\boldsymbol{t} \in \mathcal{S}_3$ and $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_3(\boldsymbol{t})$. The constraints of $\mathrm{SP}_3(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_1(\boldsymbol{t}')$ for all $\boldsymbol{t}' \in \mathcal{S}_1$ as well as $\mathrm{SP}_2(\boldsymbol{t}')$ for all $\boldsymbol{t}' \in \mathcal{S}_2$. Our earlier reasoning also implies that $p_1(\boldsymbol{t}') = 1$ and $q_1(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \mathcal{S}_1 \cup \mathcal{S}_2$. As any $\boldsymbol{t} \in \mathcal{S}_3$ is 2-unilaterally reachable from $(\bar{t}_1, \underline{t}_2) \in \mathcal{S}_2$, and as $p_1(\bar{t}_1, t_2) = 1$ and $p_2(\bar{t}_1, \underline{t}_2) = 0$, Lemma 1(ii) implies that $p_2(\boldsymbol{t}) = q_2(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_3(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

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

where the first inequality holds because $c_1 \geq 0$. The second inequality exploits (FC) and the definition of $\mathcal{S}_3$, which implies that $t_2 - c_2 < \bar{t}_1 = t_1$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_3(\boldsymbol{t})$. As $t_2 - c_2 < \bar{t}_1 = t_1$, the mechanism $(\boldsymbol{p}, \boldsymbol{q})$ can attain the optimal payoff t_1 only if $p_1(\boldsymbol{t}) = 1$ and $q_1(\boldsymbol{t}) = 0$.

Next, set $k=4$, and consider any $\boldsymbol{t}\in\mathcal{S}_4$ and $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_4(\boldsymbol{t})$. The constraints of $\mathrm{SP}_4(\boldsymbol{t})$ ensure that $(\boldsymbol{p},\boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for all $\boldsymbol{t}'\in\mathcal{S}_l$ and $l=1,2,3$. The previous arguments imply that $p_1(\boldsymbol{t}')=1$ and $q_1(\boldsymbol{t}')=0$ for all $\boldsymbol{t}'\in\mathcal{S}_1\cup\mathcal{S}_2\cup\mathcal{S}_3$. We can then use Lemma 1(i) and similar arguments as before to conclude that $p_1(\boldsymbol{t})=1$. Thus, the objective function value of $(\boldsymbol{p},\boldsymbol{q})$ in $\mathrm{SP}_4(\boldsymbol{t})$ is bounded above by $t_1-q_1(\boldsymbol{t})c_1\leq t_1$, and $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_4(\boldsymbol{t})$. As $c_1>0$, the mechanism $(\boldsymbol{p},\boldsymbol{q})$ can attain the same payoff only if $p_1(\boldsymbol{t})=1$ and $q_1(\boldsymbol{t})=0$.

Finally, set $k=5$, and consider any $\boldsymbol{t}\in\mathcal{S}_5$ and $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_5(\boldsymbol{t})$. The constraints of $\mathrm{SP}_5(\boldsymbol{t})$ ensure that $(\boldsymbol{p},\boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for all $\boldsymbol{t}'\in\mathcal{S}_l$ and $l=1,2,3,4$. Our earlier reasoning implies that $p_1(\boldsymbol{t}')=1$ and $q_1(\boldsymbol{t}')=0$ for all $\boldsymbol{t}'\in\cup_{l=1}^4\mathcal{S}_l$. As any $\boldsymbol{t}\in\mathcal{S}_5$ is 2-unilaterally reachable from $(t_1,\underline{t}_2)\in\mathcal{S}_1\cup\mathcal{S}_2$, and as $p_1(t_1,\underline{t}_2)=1$ and $p_2(t_1,\underline{t}_2)=0$, Lemma 1(ii) ensures that $p_2(\boldsymbol{t})=q_2(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p},\boldsymbol{q})$ in $\mathrm{SP}_5(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ because

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

where the first inequality holds because $p_2(\boldsymbol{t})=q_2(\boldsymbol{t})$ and $c_1\geq 0$. The second inequality exploits (FC) and the definition of $\mathcal{S}_5$, which implies that $t_2-c_2\geq\bar{t}_1\geq t_1$. Thus, $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_5(\boldsymbol{t})$. In summary, we have thus verified that $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ satisfies condition (ii) of Proposition 1 for the partition (6).

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 (5), 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 (5). 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 4 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 (5) 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

$$\mathcal{P}=\left\{\mathbb{P}\in\mathcal{P}_0(\mathcal{T})\;:\;\begin{array}{l}\mathbb{E}_{\mathbb{P}}[\tilde{t}_i]\in[\underline{\mu}_i,\overline{\mu}_i]\;\;\forall i\in\mathcal{I},\\ \tilde{t}_1,\ldots,\tilde{t}_I\text{ are mutually independent under }\mathbb{P}\end{array}\right\}. \tag{7}$$

By replacing the Markov ambiguity set (5) with its subset (7), 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 3 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 4. *If $\mathcal{P}$ is a Markov ambiguity set with independent types of the form (7), 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 4 because it is similar to that of Proposition 3. However, the proof idea can be summarized as follows. Proposition 3 implies that $\max_{i\in\mathcal{I}} \underline{\mu}_i$ provides a lower bound on $z^\star$. Proposition 4 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 3 and 4 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 (7), 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 (5) remain optimal under a Markov ambiguity set of the form (7). 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 (5) 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 (7). Hence, by Proposition 4, 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 (7), 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 (7). 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

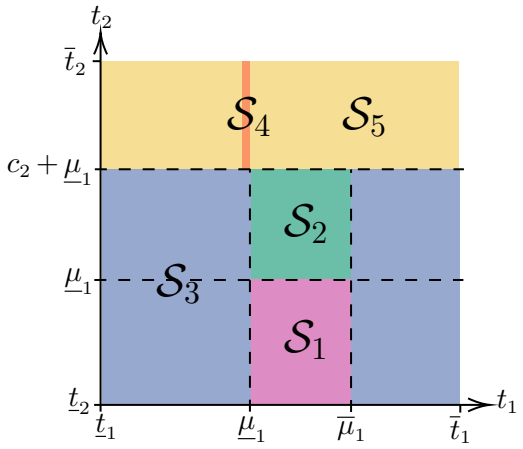

Figure 4 Partition (8) of the type space $\mathcal{T}$.

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 (7) 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 5. *If $\mathcal{P}$ is a Markov ambiguity set of the form (7) 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 exists 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) $\mathbb{P}(\tilde{\boldsymbol{t}}=\boldsymbol{t})>0$.*

In the next lemma we show that if $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton, then problem (MDP) admits a Pareto robustly optimal favored-agent mechanism $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$.

LEMMA 6. *Assume that $\mathcal{P}$ is a Markov ambiguity set of the form (7) 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 6 is reminiscent of Lemma 4. The assumption that $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$ is a singleton will be relaxed in Theorem 6 below. To gain some intuition into this result, we sketch the proof Lemma 6 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 $\overline{t}_2 > c_2+\underline{\mu}_1$. 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 4.

$$\begin{aligned}
\mathcal{S}_1 &= \{\boldsymbol{t}\in\mathcal{T} : t_2 - c_2 \leq \underline{\mu}_1,\ t_2\leq\underline{\mu}_1 \text{ and } t_1\in(\underline{\mu}_1,\overline{\mu}_1]\}\\
\mathcal{S}_2 &= \{\boldsymbol{t}\in\mathcal{T} : t_2 - c_2 \leq \underline{\mu}_1,\ t_2>\underline{\mu}_1 \text{ and } t_1\in(\underline{\mu}_1,\overline{\mu}_1]\}\\
\mathcal{S}_3 &= \{\boldsymbol{t}\in\mathcal{T} : t_2 - c_2 \leq \underline{\mu}_1, \text{ and } t_1\notin(\underline{\mu}_1,\overline{\mu}_1]\}\\
\mathcal{S}_4 &= \{\boldsymbol{t}\in\mathcal{T} : t_2 - c_2 > \underline{\mu}_1, \text{ and } t_1=\underline{\mu}_1\}\\
\mathcal{S}_5 &= \{\boldsymbol{t}\in\mathcal{T} : t_2 - c_2 > \underline{\mu}_1, \text{ and } t_1\neq\underline{\mu}_1\}
\end{aligned} \tag{8}$$

Under our simplifying assumptions about $\underline{\mu}_1$, $\underline{\mu}_2$, $\bar{t}_2$ and c_2, one can show that all of these sets are nonempty. We emphasize, however, that the formal proof of Lemma 6 in the online appendix does not rely on any of the simplifying assumptions imposed here.

Denote now by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ the type (i) favored-agent mechanism with favored agent $i^\star$ and threshold value $\nu^\star = \underline{\mu}_{i^\star}$. To prove Lemma 6, it suffices to show that the conditions (i) and (ii) of Proposition 1 are satisfied for the partition (8). The claim then follows from Corollary 1. We will exploit Lemma 5 to verify condition (i). The arguments needed to verify condition (ii) closely parallel those used in the proof of Lemma 4. Thus, we omit this part of the proof for brevity.

We now show that condition (i) is satisfied, that is, for any $k \in \{1, \ldots, 5\}$ and $\boldsymbol{t} \in \mathcal{S}_k$ there exists $\mathbb{P} \in \mathcal{P}$ supported on $\cup_{l=1}^k \mathcal{S}_l$ with $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. By Lemma 5, given any $\boldsymbol{t} \in \mathcal{T}$ and $\mu_1 \in [\underline{\mu}_1, \overline{\mu}_1]$ there exists a scenario $\hat{\boldsymbol{t}} \in \mathcal{T}$ with $\hat{t}_2 < \underline{\mu}_1$ and a discrete distribution $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{E}_{\mathbb{P}}[\tilde{t}_1] = \mu_1$, $\mathbb{P}(\tilde{t}_i \in \{t_i, \hat{t}_i\}) = 1$ for all $i \in \mathcal{I}$, and $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. We will show that $\mathbb{P}$ is supported on $\cup_{l=1}^k \mathcal{S}_l$ and therefore satisfies condition (i) for a judiciously chosen $\mu_1 \in [\underline{\mu}_1, \overline{\mu}_1]$. To this end, we first study some implications of Lemma 5 on the support $\mathbb{S}$ of $\mathbb{P}$. Clearly, we have $\mathbb{S} \subseteq \{\boldsymbol{t}' \in \mathcal{T} : t'_i \in \{t_i, \hat{t}_i\} \text{ for all } i \in \mathcal{I}\}$, and as $\hat{t}_2 < \underline{\mu}_1$, we also have

$$\mathbb{S} \subseteq \left\{\boldsymbol{t}' \in \mathcal{T} : t'_2 \leq \max\{t_2, \underline{\mu}_1\} \text{ and } t'_2 - c_2 \leq \max\{t_2 - c_2, \underline{\mu}_1\}\right\}. \tag{9}$$

In the special case when $t_1 = \mu_1$, then the valid relations $\mathbb{E}[\tilde{t}_1] = \mu_1$ and $\mathbb{P}(\tilde{t}_1 \in \{t_1, \hat{t}_1\}) = 1$ imply that $\hat{t}_1 = t_1$. Hence, if $t_1 = \mu_1$, then we additionally have

$$\mathbb{S} \subseteq \left\{\boldsymbol{t}' \in \mathcal{T} : t'_2 \leq \max\{t_2, \underline{\mu}_1\} \text{ and } t'_2 - c_2 \leq \max\{t_2 - c_2, \underline{\mu}_1\} \text{ and } t'_1 = t_1\right\}. \tag{10}$$

To prove condition (i), set first $k = 1$, and fix any $\boldsymbol{t} \in \mathcal{S}_1$. By the definition of $\mathcal{S}_1$, we have $\max\{t_2, \underline{\mu}_1\} = \max\{t_2 - c_2, \underline{\mu}_1\} = \underline{\mu}_1$ and $t_1 \in (\underline{\mu}_1, \overline{\mu}_1]$. In this case, we choose $\mu_1 = t_1$, in which case the support of $\mathbb{P}$ satisfies (10). As $\max\{t_2, \underline{\mu}_1\} = \max\{t_2 - c_2, \underline{\mu}_1\} = \underline{\mu}_1$, we obtain

$$\mathbb{S} \subseteq \left\{\boldsymbol{t}' \in \mathcal{T} : t'_1 = t_1 \text{ and } t'_2 - c_2 \leq \underline{\mu}_1 \text{ and } t'_2 \leq \underline{\mu}_1\right\} \subseteq \mathcal{S}_1.$$

This implies that $\mathbb{P} \in \mathcal{P}_0(\mathcal{S}_1)$. Hence, we have shown that condition (i) holds for any $\boldsymbol{t} \in \mathcal{S}_1$.

Consider next $k = 2$, and fix any $\boldsymbol{t} \in \mathcal{S}_2$. By the definition of $\mathcal{S}_2$ we have $\max\{t_2 - c_2, \underline{\mu}_1\} = \underline{\mu}_1$ and $t_1 \in (\underline{\mu}_1, \overline{\mu}_1]$. In this case, we choose again $\mu_1 = t_1$. As the support of $\mathbb{P}$ satisfies (10) and as $\max\{t_2 - c_2, \underline{\mu}_1\} = \underline{\mu}_1$, we find $\mathbb{S} \subseteq \{\boldsymbol{t}' \in \mathcal{T} : t'_2 - c_2 \leq \underline{\mu}_1 \text{ and } t'_1 = t_1\} \subseteq \mathcal{S}_1 \cup \mathcal{S}_2$, which implies that $\mathbb{P} \in \mathcal{P}_0(\mathcal{S}_1 \cup \mathcal{S}_2)$. Hence, we have shown that condition (i) holds for any $\boldsymbol{t} \in \mathcal{S}_2$.

Next, set $k = 3$, and fix any $\boldsymbol{t} \in \mathcal{S}_3$. By the definition of $\mathcal{S}_3$, we again have $\max\{t_2 - c_2, \underline{\mu}_1\} = \underline{\mu}_1$. In this case, we choose $\mu_1 \in [\underline{\mu}_1, \overline{\mu}_1]$ arbitrarily. As the support of $\mathbb{P}$ satisfies (9) and as $\max\{t_2 - c_2, \underline{\mu}_1\} = \underline{\mu}_1$, we find $\mathbb{S} \subseteq \{\boldsymbol{t}' \in \mathcal{T} : t'_2 - c_2 \leq \underline{\mu}_1\} \subseteq \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$, which implies that $\mathbb{P} \in \mathcal{P}_0(\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3)$. Hence, we have shown that condition (i) holds for any $\boldsymbol{t} \in \mathcal{S}_3$.

Next, set $k = 4$, and fix any $\boldsymbol{t} \in \mathcal{S}_4$. By the definition of $\mathcal{S}_4$, we have $t_1 = \underline{\mu}_1 \in [\underline{\mu}_1, \overline{\mu}_1]$. In this case, we choose $\mu_1 = t_1$. As the support of $\mathbb{P}$ satisfies (10), we find $\mathbb{S} \subseteq \{\boldsymbol{t}' \in \mathcal{T} : t'_1 = t_1 = \underline{\mu}_1\} \subseteq \cup_{l=1}^4 \mathcal{S}_l$, which implies that $\mathbb{P} \in \mathcal{P}_0(\cup_{l=1}^4 \mathcal{S}_l)$. Hence, we have shown that condition (i) holds for any $\boldsymbol{t} \in \mathcal{S}_4$. For $k = 5$, condition (i) trivially holds for any $\boldsymbol{t} \in \mathcal{S}_5$ thanks to Lemma 5.

In summary, we have verified that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ satisfies condition (i) of Proposition 1 for the partition (8). To prove that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ also satisfies condition (ii), one can proceed as in the proof of Lemma 4. Details are omitted for brevity.

The next theorem shows that the Pareto robust optimality result of Lemma 6 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 (7), 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 (5) is replaced with its subset (7) 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 (5). 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 (7) with independent types. To this end, consider an arbitrary distribution $\mathbb{P}$ in the Markov ambiguity set (7). The expected payoff of $(\boldsymbol{p}, \boldsymbol{q})$ is then given by

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

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

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

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

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 $\bar{t}_2 - c_2 \geq \bar{t}_1 > \overline{\mu}_1$. Indeed, we can define $\mathbb{P}^\star$ as the two-point distribution that assigns probability $\alpha = (\underline{\mu}_2 - \underline{t}_2)/((\bar{t}_1 + \overline{\mu}_1)/2 + c_2 - \underline{t}_2)$ to scenario $(\underline{\mu}_1, (\bar{t}_1 + \overline{\mu}_1)/2 + c_2)$ and probability $1 - \alpha$ to scenario $(\underline{\mu}_1, \underline{t}_2)$. One readily verifies that this distribution belongs to the ambiguity set (7) 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 $\nu \geq \bar{t}_1$ if $\mathcal{P}$ is a Markov ambiguity set of the form (7) 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 (5) and (7). 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 (5) and (7). 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 $\nu < \bar{t}_{i^\star}$ is strictly suboptimal (and thus fails to be Pareto robustly optimal) under the ambiguity set (5). 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 $\nu \geq \bar{t}_{i^\star}$ fails to be Pareto robustly optimal under the ambiguity set (7). This implies that it is crucial for the principal to know whether or not the agents' types are independent.

6. Numerical Illustration

We now assess the Pareto robustly optimal mechanisms designed for different ambiguity sets against the optimal mechanism tailored to a crisp (but possibly misspecified) type distribution. For better comparability, we report the expected payoffs generated by different mechanisms relative to the maximum expected payoff achievable with full knowledge of the true type distribution.

Throughout this section we assume that there are two agents with $\mathcal{T}_1 = [0.2, 2]$, $\mathcal{T}_2 = [0, 1.5]$, $[\underline{\mu}_1, \overline{\mu}_1] = [0.75, 1.25]$, $[\underline{\mu}_2, \overline{\mu}_2] = [0.5, 1]$, $c_1 = 0.1$, and $c_2 = 0.4$. We also assume that the true type distribution is of the form $\mathbb{P}_\epsilon = (1-\epsilon)\mathbb{P}^{\mathrm{N}} + \epsilon\mathbb{P}^{\mathrm{E}}$, where $\mathbb{P}^{\mathrm{N}} \in \mathcal{P}_0(\mathcal{T})$ represents a nominal distribution that captures the seller's best guess for the unknown true distribution, and $\mathbb{P}^{\mathrm{E}} \in \mathcal{P}_0(\mathcal{T})$ represents an extremal distribution to be specified later. The weight $\epsilon \in [0, 1]$ determines the contamination level of $\mathbb{P}^{\mathrm{N}}$. Indeed, as ϵ increases, the true distribution $\mathbb{P}_\epsilon$ approaches $\mathbb{P}^{\mathrm{E}}$, thus becoming increasingly different from $\mathbb{P}^{\mathrm{N}}$. In the following we use $\mathcal{N}_{\mathcal{T}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote the truncated normal distribution obtained by conditioning the (untruncated) normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ on the event $\tilde{\boldsymbol{t}} \in \mathcal{T}$. We then define $\mathbb{P}^{\mathrm{N}} = \mathcal{N}_{\mathcal{T}}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ as the truncated normal distribution with $\boldsymbol{\mu}_0 = \bar{\boldsymbol{t}}/2$ and $\boldsymbol{\Sigma}_0 = \mathrm{diag}(\bar{\boldsymbol{t}})$.

We use three different models for the extremal distribution, that is, we set $\mathbb{P}^{\mathrm{E}}_1 = \mathcal{N}_{\mathcal{T}}(\underline{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$, $\mathbb{P}^{\mathrm{E}}_2 = \mathcal{N}_{\mathcal{T}}(\overline{\boldsymbol{\mu}}, \boldsymbol{\Sigma})$ and $\mathbb{P}^{\mathrm{E}}_3 = \frac{1}{2}\mathcal{N}_{\mathcal{T}}(\underline{\boldsymbol{t}}, \boldsymbol{\Sigma}) + \frac{1}{2}\mathcal{N}_{\mathcal{T}}(\bar{\boldsymbol{t}}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \frac{1}{100}\boldsymbol{I}$. By construction, $\mathbb{P}^{\mathrm{E}}_1$ and $\mathbb{P}^{\mathrm{E}}_2$ concentrate all probability mass near the lower and upper bounds on the expected value of $\tilde{\boldsymbol{t}}$, respectively, and $\mathbb{P}^{\mathrm{E}}_3$ concentrates half of the probability mass near $\underline{\boldsymbol{t}}$ and the other half near $\bar{\boldsymbol{t}}$. Note that $\mathbb{P}^{\mathrm{E}}_1$ represents a blurred version of the worst-case distribution that minimizes the optimal mechanism's expected payoff over the Markov ambiguity sets with or without independent types. In contrast, $\mathbb{P}^{\mathrm{E}}_2$ represents a blurred version of the corresponding best-case distribution. Contaminating $\mathbb{P}^{\mathrm{N}}$ with $\mathbb{P}^{\mathrm{E}}_1$ or $\mathbb{P}^{\mathrm{E}}_2$ thus captures two complementary extreme scenarios. Lastly, note that the agent's types are perfectly correlated under $\mathbb{P}^{\mathrm{E}}_3$, in which case the independence assumption is in some sense maximally violated. This distribution facilitates a more nuanced comparison between the Pareto robustly optimal mechanisms for Markov ambiguity sets with and without independent types.

For every $\epsilon \in [0, 1]$, we use the optimal mechanism under perfect distributional information as a baseline. This mechanism solves problem (MDP) in view of the singleton ambiguity set $\mathcal{P} = \{\mathbb{P}_\epsilon\}$. We emphasize that this mechanism may *not* be a favored-agent mechanism because the types of the two agents fail to be independent under $\mathbb{P}_\epsilon$. Thus, the infinite-dimensional linear program (MDP) cannot be solved exactly. This prompts us to approximate $\mathcal{T}$ by the grid $\hat{\mathcal{T}} = \mathcal{T} \cap \delta \cdot \mathbb{Z}^I$ with $\delta = 0.05$ and to approximate $\mathbb{P}^{\mathrm{N}}$ by a discrete distribution $\hat{\mathbb{P}}^{\mathrm{N}}$ supported on $\hat{\mathcal{T}}$. Specifically, the probabilities of $\hat{\mathbb{P}}^{\mathrm{N}}$ are obtained by evaluating the probability density function of the normal distribution $\mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ at all grid points in $\hat{\mathcal{T}}$ and by normalizing them so that they sum to 1. Similarly, the three extremal distributions are approximated by discrete distributions $\hat{\mathbb{P}}^{\mathrm{E}}_1$, $\hat{\mathbb{P}}^{\mathrm{E}}_2$ and $\hat{\mathbb{P}}^{\mathrm{E}}_3$. The optimal mechanism is then approximated by the solution of a discretized version of problem (MDP) with $\mathcal{P} = \{\hat{\mathbb{P}}_\epsilon\}$, which enforces the robust constraints only on $\hat{\mathcal{T}}$. Here, $\hat{\mathbb{P}}_\epsilon$ is defined in the obvious way as a mixture of the discretized nominal and extremal distributions. As $\hat{\mathcal{T}} \subseteq \mathcal{T}$, the discretized mechanism design problem relaxes (MDP) and thus overestimates the expected payoff that can be earned with the exact optimal mechanism. In addition, as $\hat{\mathcal{T}}$ is finite, the discretized mechanism design problem constitutes a finite-dimensional linear program. All linear programs are implemented in MATLAB R2022a using the YALMIP interface and solved with GUROBI.

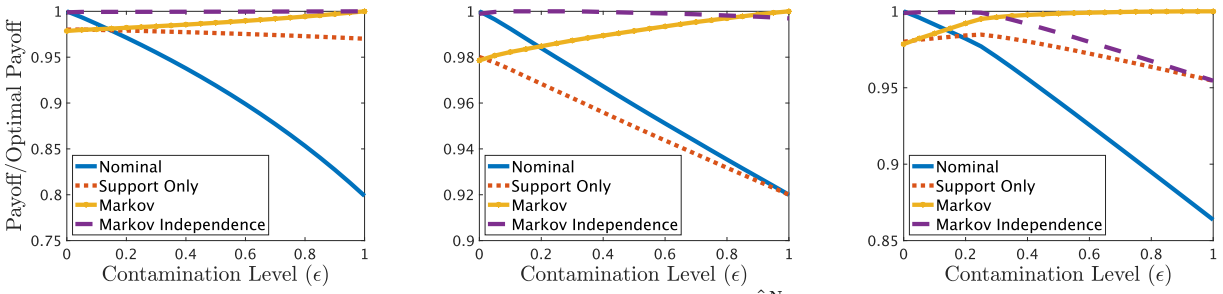

Figure 5 Expected payoffs of the optimal mechanism tailored to $\hat{\mathbb{P}}^{\mathrm{N}}$ and the proposed Pareto robustly optimal mechanisms for different contamination levels ϵ and for $\hat{\mathbb{P}}_1^{\mathrm{E}}$ (left), $\hat{\mathbb{P}}_2^{\mathrm{E}}$ (middle) and $\hat{\mathbb{P}}_3^{\mathrm{E}}$ (right). All results are normalized by the expected payoff of the optimal mechanism with full distributional information.

In the first experiment we compare the different mechanisms in view of their *worst-case* payoffs across all scenarios in $\boldsymbol{t} \in \hat{\mathcal{T}}$. While the worst-case payoff of the nominal mechanism tailored to $\hat{\mathbb{P}}^{\mathrm{N}}$ amounts to 0, the Pareto robustly optimal mechanisms tailored to support-only and Markov ambiguity sets with and without independent types all generate a worst-case payoff equal to 0.2. When evaluating the worst case only over the scenarios $\boldsymbol{t} \in \hat{\mathcal{T}} \cap ([\underline{\mu}_1, \overline{\mu}_1] \times [\underline{\mu}_2, \overline{\mu}_2])$, the nominal mechanism, which generates a worst-case payoff of 0.5, is still inferior to all Pareto robustly optimal mechanisms, which generate worst-case payoffs equal to 0.65 (for the support-only ambiguity set) and 0.75 (for the Markov ambiguity sets with or without independent types).

In the second experiment we compare the different mechanisms in view of their *expected* payoffs under $\hat{\mathbb{P}}_\epsilon$ normalized by the maximum expected payoff achievable when $\hat{\mathbb{P}}_\epsilon$ is known. Figure 5 visualizes the resulting relative payoffs as a function of the contamination level $\epsilon \in [0,1]$ for the three different contaminating distributions $\hat{\mathbb{P}}_1^{\mathrm{E}}$, $\hat{\mathbb{P}}_2^{\mathrm{E}}$ and $\hat{\mathbb{P}}_3^{\mathrm{E}}$. We observe that the Pareto robustly optimal mechanisms tailored to Markov ambiguity sets consistently outperform the nominal mechanism tailored to $\hat{\mathbb{P}}^{\mathrm{N}}$ even for small contamination levels. This outperformance becomes more pronounced as ϵ increases. Notably, the Pareto robustly optimal mechanism for Markov ambiguity set with independent types is only marginally suboptimal at $\epsilon = 0$ (in which case the nominal mechanism is optimal) and performs exceptionally well across all contamination levels when the contaminating distribution is set to $\hat{\mathbb{P}}_1^{\mathrm{E}}$ or $\hat{\mathbb{P}}_2^{\mathrm{E}}$. This is perhaps expected because the agents' types are (almost) independent under these two distributions. If the contaminating distribution is set to $\hat{\mathbb{P}}_3^{\mathrm{E}}$, under which the types are highly correlated, then this mechanism's performance deteriorates for $\epsilon \gtrsim 0.25$, and it is outperformed by the Pareto robustly optimal mechanism for Markov ambiguity sets *without* independent types. The Pareto robustly optimal mechanism for support-only ambiguity sets uses only minimal distributional information and thus displays a worse performance than the Pareto robustly optimal mechanisms for Markov ambiguity sets. Nevertheless, its performance is better or similar to that of the nominal mechanism. We thus conclude that the Pareto robustly optimal mechanisms not only provide the best possible worst-case guarantees but are also likely to outperform optimal mechanisms tailored to crisp distributions that are only marginally contaminated.

7. 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.

The allocation problem with costly verification addressed in this paper generalizes the stochastic model by Ben-Porath et al. [5], which posits that the type distribution is commonly known. Different variants of this stochastic model have been investigated in the recent literature. Li [19] studies the impact of limited penalties by assuming that the principal can recover the good only *partially* when agents misreport their types. Hu [15] assumes that the principal can only observe *signals* correlated with the types and that the cost of inspection increases with the accuracy of the signal, and Chua et al. [11] assume that there are *multiple* homogeneous goods that need to be allocated to *different* agents. All of these problem variants are addressed with fundamentally different techniques than the baseline problem by [5]. In addition, the corresponding optimal mechanisms are structurally different from favored-agent mechanisms studied here. More research is needed to study these problem variants under distributional ambiguity.

References

[1] Anunrojwong J, Balseiro S, Besbes O (2022) On the robustness of second-price auctions in prior-independent mechanism design. *Proceedings of the 23rd ACM Conference on Economics and Computation*, 151–152.

[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: A 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] Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization* (Princeton University Press).

[7] Bertsimas D, ten Eikelder SC, den Hertog D, Trichakis N (2023) Pareto adaptive robust optimality via a Fourier–Motzkin elimination lens. *Mathematical Programming* 1–54.

[8] 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.

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

[10] Chen Z, Hu Z, Wang R (2021) Screening with limited information: The minimax theorem and a geometric approach. *International Conference on Web and Internet Economics*, 549.

[11] Chua GA, Hu G, Liu F (2023) Optimal multi-unit allocation with costly verification. *Social Choice and Welfare* 61(3):455–488.

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

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

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

[15] Hu G (2024) Screening by (in)accurate inspection. *SSRN preprint* 4797356.

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

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

[18] Koçyiğit Ç, Rujeerapaiboon N, Kuhn D (2022) Robust multidimensional pricing: Separation without regret. *Mathematical Programming* 196(1–2):841–874.

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

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

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

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

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

[24] Walton D, Carroll G (2022) A general framework for robust contracting models. *Econometrica* 90(5):2129–2159.

[25] Wang S, Liu S, Zhang J (2024) Minimax regret robust screening with moment information. *Manufacturing & Service Operations Management* (Forthcoming).

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

Appendix.

A. Robustness Against the Agents' Attitude Towards Ambiguity

Incentive compatibility constraints are meant to ensure that all agents prefer to reveal their true types instead of lying. However, the probability that agent i receives the good, which amounts to $p_i(t_i, \boldsymbol{t}_{-i})$ under truthful reporting and to $p_i(t_i', \boldsymbol{t}_{-i}) - q_i(t_i', \boldsymbol{t}_{-i})$ under untruthful reporting, depends on the other agents' types and is thus uncertain. As agents may use different decision criteria to rank their choices under uncertainty, there is no canonical way of enforcing incentive compatibility. For instance, if an agent believes that the type distribution is $\mathbb{P} \in \mathcal{P}_0(\mathcal{T})$, he might rank different choices by their expected utility under $\mathbb{P}$. Alternatively, if an agent believes that the type distribution falls within an ambiguity set $\mathcal{P} \subseteq \mathcal{P}_0(\mathcal{T})$, he might rank different choices by their worst-case expected utility, their best-case expected utility, or a convex combination of these, in view of all distributions in $\mathcal{P}$. Different decision criteria give rise to different incentive compatibility constraints.

Definition 5. A mechanism $(\boldsymbol{p}, \boldsymbol{q})$ is called

- *Bayesian incentive compatible* with respect to a distribution $\mathbb{P} \in \mathcal{P}_0(\mathcal{T})$ if for all $i \in \mathcal{I}$, $t_i, t_i' \in \mathcal{T}_i$,

$$\mathbb{E}_{\mathbb{P}}[p_i(t_i, \tilde{\boldsymbol{t}}_{-i}) \mid \tilde{t}_i = t_i] \geq \mathbb{E}_{\mathbb{P}}[p_i(t_i', \tilde{\boldsymbol{t}}_{-i}) - q_i(t_i', \tilde{\boldsymbol{t}}_{-i}) \mid \tilde{t}_i = t_i],$$

- *Hurwicz incentive compatible* with respect to $\alpha \in [0,1]$ and $\mathcal{P} \subseteq \mathcal{P}_0(\mathcal{T})$ if for all $i \in \mathcal{I}$, $t_i, t_i' \in \mathcal{T}_i$,

$$\begin{aligned}&\alpha \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[p_i(t_i, \tilde{\boldsymbol{t}}_{-i}) \mid \tilde{t}_i = t_i] + (1-\alpha) \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[p_i(t_i, \tilde{\boldsymbol{t}}_{-i}) \mid \tilde{t}_i = t_i] \\ &\geq \alpha \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[p_i(t_i', \tilde{\boldsymbol{t}}_{-i}) - q_i(t_i', \tilde{\boldsymbol{t}}_{-i}) \mid \tilde{t}_i = t_i] + (1-\alpha) \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[p_i(t_i', \tilde{\boldsymbol{t}}_{-i}) - q_i(t_i', \tilde{\boldsymbol{t}}_{-i}) \mid \tilde{t}_i = t_i].\end{aligned}$$

The parameter α in the Hurwicz incentive compatibility constraints encodes the agent's risk aversion. Indeed, if $\alpha = 1$, the agent ranks outcomes based on their worst-case expected utility, which reflects risk-averse behavior. On the other hand, if $\alpha = 0$, the agent ranks outcomes based on the best-case expected payoff, which reflects risk-seeking behavior. The following proposition shows that the incentive compatibility constraints (IC) in problem (MDP) ensure that even agents with Bayesian or Hurwicz preferences have no incentive to misreport their true types.

Proposition 5. *If $(\boldsymbol{p}, \boldsymbol{q})$ satisfies (IC), then it is Bayesian incentive compatible with respect to all $\mathbb{P} \in \mathcal{P}_0(\mathcal{T})$ and Hurwicz incentive compatible with respect to all $\alpha \in [0,1]$ and $\mathcal{P} \subseteq \mathcal{P}_0(\mathcal{T})$.*

Proof. By taking expectations with respect to any $\mathbb{P} \in \mathcal{P}_0(\mathcal{T})$ on both sides of the (IC) constraints, it is easy to see that (IC) implies Bayesian incentive compatibility. Next, consider any risk aversion parameter $\alpha \in [0,1]$ and ambiguity set $\mathcal{P} \subseteq \mathcal{P}_0(\mathcal{T})$. We already know that Bayesian incentive compatibility holds for every distribution $\mathbb{P} \in \mathcal{P}$. Maximizing or minimizing both sides of an inequality depending on $\mathbb{P}$ over all $\mathbb{P} \in \mathcal{P}$ preserves the inequality. In addition, taking convex combinations of two valid inequalities (of the same direction) yields another valid inequality. Therefore, one readily verifies that $(\boldsymbol{p}, \boldsymbol{q})$ satisfies Hurwicz incentive compatibility. □

Proposition implies that the (IC) constraints enforced in problem (MDP) protect the principal against uncertainty about the agents' preferences.

B. Proofs

Proof of Lemma 1. Fix any $i \in \mathcal{I}$, $\boldsymbol{t} \in \mathcal{T}$ and $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$, and suppose that $p_i(\boldsymbol{t}) - q_i(\boldsymbol{t}) = 1$. We then have $1 \geq p_i(t_i', \boldsymbol{t}_{-i}) \geq p_i(\boldsymbol{t}) - q_i(\boldsymbol{t}) = 1$ for any $t_i' \in \mathcal{T}_i$, where the two inequalities follow from (FC) and (IC), respectively. This implies that $p_i(t_i', \boldsymbol{t}_{-i}) = 1$, and thus assertion (i) follows.

Next, fix any $i \in \mathcal{I}$, $\boldsymbol{t} \in \mathcal{T}$ and $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$, and suppose that $p_i(\boldsymbol{t}) = 0$. We then have $0 = p_i(\boldsymbol{t}) \geq p_i(t_i', \boldsymbol{t}_{-i}) - q_i(t_i', \boldsymbol{t}_{-i}) \geq 0$ for any $t_i' \in \mathcal{T}_i$, where the two inequalities follow from (IC) and (FC), respectively. This implies that $p_i(t_i', \boldsymbol{t}_{-i}) = q_i(t_i', \boldsymbol{t}_{-i})$, and thus assertion (ii) follows. □

Proof of Proposition 1. Fix any optimal mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star) \in \mathcal{X}$, and suppose that there exists a partition $\mathcal{S}_1, \ldots, \mathcal{S}_m$ of $\mathcal{T}$ satisfying the conditions (i) and (ii) in the proposition statement. Consider now any feasible mechanism $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$ that Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. In the following, we will use induction to show that $(\boldsymbol{p}, \boldsymbol{q})$ must generate the same payoff as $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in every scenario $\boldsymbol{t} \in \mathcal{T}$. Thus, $(\boldsymbol{p}, \boldsymbol{q})$ cannot generate a strictly higher expected payoff than $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ under any distribution, which contradicts our assumption that $(\boldsymbol{p}, \boldsymbol{q})$ Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$.

As for the base step of the induction, we aim to show that $(\boldsymbol{p}, \boldsymbol{q})$ generates the same payoff as $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in every scenario $\boldsymbol{t} \in \mathcal{S}_1$. To this end, note first that $(\boldsymbol{p}, \boldsymbol{q})$ is feasible in $\mathrm{SP}_1(\boldsymbol{t})$. As $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_1(\boldsymbol{t})$ thanks to assumption (ii), the mechanism $(\boldsymbol{p}, \boldsymbol{q})$ must generate a weakly lower payoff than $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in scenario $\boldsymbol{t}$. The same argument applies for every scenario $\boldsymbol{t} \in \mathcal{S}_1$, and thus we have

$$\sum_{i \in \mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \leq \sum_{i \in \mathcal{I}} (p_i^\star(\boldsymbol{t}) t_i - q_i^\star(\boldsymbol{t}) c_i) \quad \forall \boldsymbol{t} \in \mathcal{S}_1. \tag{11}$$

Assume now that the inequality in (11) is strict for some $\boldsymbol{t}' \in \mathcal{S}_1$. By assumption (i), there exists a distribution $\mathbb{P} \in \mathcal{P} \cap \mathcal{P}_0(\mathcal{S}_1)$ that assigns a strictly positive probability mass to scenario $\boldsymbol{t}'$. Taking expectations with respect to $\mathbb{P}$ on both sides of (11) thus yields

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

which contradicts our initial assumption that $(\boldsymbol{p}, \boldsymbol{q})$ Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Thus, the inequality in (11) must be an equality for every $\boldsymbol{t} \in \mathcal{S}_1$, that is, the mechanisms $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ must indeed generate the same payoff throughout $\mathcal{S}_1$.

The induction step corresponding to $k \geq 2$ is based on the hypothesis that $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generate the same payoff on $\cup_{l=1}^{k-1} \mathcal{S}_l$. Our goal is to show that the two mechanisms also generate the same payoff in every scenario $\boldsymbol{t} \in \mathcal{S}_k$. To this end, note first that $(\boldsymbol{p}, \boldsymbol{q})$ is feasible in $\mathrm{SP}_k(\boldsymbol{t})$ due to the induction hypothesis. As $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$ thanks to assumption (ii), $(\boldsymbol{p}, \boldsymbol{q})$ generates a weakly lower payoff than $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in scenario $\boldsymbol{t}$. As this is true for every scenario $\boldsymbol{t} \in \mathcal{S}_k$, we have

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

As in the base case, one can now show by contradiction that the inequality in the above expression is never strict. Hence, $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generate the same payoff throughout $\mathcal{S}_k$.

By induction, the two mechanisms $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generate the same payoff on all of $\mathcal{T}$ and thus the same *expected* payoff under every $\mathbb{P} \in \mathcal{P}$. This contradicts the assumption that $(\boldsymbol{p}, \boldsymbol{q})$ Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. As $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is robustly optimal by assumption, and as $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is *not* Pareto robustly dominated by any feasible mechanism, it is indeed Pareto robustly optimal.

Note that our arguments above also imply that any mechanism $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$ that weakly Pareto robustly dominates $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generates the same payoff as $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in every scenario $\boldsymbol{t} \in \mathcal{T}$. □

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

$$\begin{array}{rll} z^\star \leq & \sup\limits_{\boldsymbol{p}, \boldsymbol{q}} & \inf\limits_{\boldsymbol{t} \in \mathcal{T}} \sum\limits_{i \in \mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \\ & \text{s.t.} & p_i, q_i \in \mathcal{L}(\mathcal{T}, [0,1]) \ \forall i \in \mathcal{I} \\ & & \sum\limits_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1 \ \forall \boldsymbol{t} \in \mathcal{T} \\ = & \sup\limits_{\boldsymbol{p}} & \inf\limits_{\boldsymbol{t} \in \mathcal{T}} \sum\limits_{i \in \mathcal{I}} p_i(\boldsymbol{t}) t_i \\ & \text{s.t.} & p_i \in \mathcal{L}(\mathcal{T}, [0,1]) \ \forall i \in \mathcal{I}, \ \sum\limits_{i \in \mathcal{I}} p_i(\boldsymbol{t}) \leq 1 \ \forall \boldsymbol{t} \in \mathcal{T}, \end{array}$$

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 2 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 2. □

Proof of Lemma 2. 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 $(\boldsymbol{p}^k, \boldsymbol{q}^k)$, $k \in \mathcal{K} = \{1, \ldots, K\}$. 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 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 $\nu^\star = \underline{t}_{i^\star}$. We also use the following partition of the type space $\mathcal{T}$.

$$\begin{aligned}
\mathcal{T}_I &= \{\boldsymbol{t} \in \mathcal{T} : \max_{i\neq i^\star} t_i - c_i \leq \underline{t}_{i^\star} \text{ and } t_{i^\star} = \bar{t}_{i^\star}\} \\
\mathcal{T}_{II} &= \{\boldsymbol{t} \in \mathcal{T} : \max_{i\neq i^\star} t_i - c_i \leq \underline{t}_{i^\star} \text{ and } t_{i^\star} < \bar{t}_{i^\star}\} \\
\mathcal{T}_{III} &= \{\boldsymbol{t} \in \mathcal{T} : \max_{i\neq i^\star} t_i - c_i > \underline{t}_{i^\star}\}
\end{aligned}$$

The sets $\mathcal{T}_I$ and $\mathcal{T}_{II}$ are nonempty and contain at least $(\bar{t}_{i^\star}, \underline{\boldsymbol{t}}_{-i^\star})$ and $\underline{\boldsymbol{t}}$, respectively, because $\underline{t}_{i^\star} < \bar{t}_{i^\star}$ and $c_i > 0$ for all $i \in \mathcal{I}$. However, $\mathcal{T}_{III}$ can be empty if $\underline{t}_{i^\star}$ or c_i, $i \neq i^\star$, are sufficiently large.

To establish the Pareto robust optimality of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, we leverage Proposition 1. From Theorem 1 we already know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is robustly optimal. To prove that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is also Pareto robustly optimal, it thus suffices to prove that the conditions (i) and (ii) in Proposition 1 hold for some

partition of the type space $\mathcal{T}$. Specifically, we will show that it holds for a refinement of the partition $\mathcal{T}_I, \mathcal{T}_{II}, \mathcal{T}_{III}$. Condition (i) trivially holds because the support-only ambiguity set contains all Dirac point distributions supported on any point $\boldsymbol{t} \in \mathcal{T}$. It thus remains to verify condition (ii).

The rest of the proof is divided into three steps focusing on the three subsets $\mathcal{T}_I$, $\mathcal{T}_{II}$ and $\mathcal{T}_{III}$. In this process, we will construct a refined partition $\mathcal{S}_1, \ldots, \mathcal{S}_m$ of $\mathcal{T}$ with $m = 2I+1$, where $\mathcal{S}_1, \ldots, \mathcal{S}_I$ forms a partition of $\mathcal{T}_I$, $\mathcal{S}_{I+1}$ coincides with $\mathcal{T}_{II}$, and $\mathcal{S}_{I+2}, \ldots, \mathcal{S}_{2I+1}$ forms a partition of $\mathcal{T}_{III}$. We will also exploit the non-locality of the incentive compatibility constraint (IC) via Lemma 1 to show iteratively that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves the scenario problem $\mathrm{SP}_k(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{S}_k$ and $k = 1, \ldots, m$.

The subsequent reasoning critically relies on our assumption that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is a favored-agent mechanism of type (i) with favored agent $i^\star \in \arg\max_{i \in \mathcal{I}} \underline{t}_i$ and threshold value $\nu^\star = \underline{t}_{i^\star}$. By Definition 3, this means that the principal's payoff in scenario $\boldsymbol{t}$ 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 $\boldsymbol{t} \in \mathcal{T}_{III}$).

Step 1 ($\mathcal{T}_I$). We partition $\mathcal{T}_I$ into I subsets of the form

$$\mathcal{S}_k = \{\boldsymbol{t} \in \mathcal{T}_I : |\{i \in \mathcal{I} : t_i \geq t_{i^\star}\}| = k\}$$

for $k = 1, \ldots, I$. We will use induction on k to prove that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is optimal in $\mathrm{SP}_k(\boldsymbol{t})$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_k(\boldsymbol{t})$ must satisfy $\boldsymbol{p}(\boldsymbol{t}) = \boldsymbol{p}^\star(\boldsymbol{t})$ and $\boldsymbol{q}(\boldsymbol{t}) = \boldsymbol{q}^\star(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{S}_k$.

As for the base step, note that any mechanism $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$ is feasible in $\mathrm{SP}_1(\boldsymbol{t})$ for any $\boldsymbol{t} \in \mathcal{S}_1$. In addition, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ is dominated by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ in $\mathrm{SP}_1(\boldsymbol{t})$ because

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

where the first inequality holds because $c_i > 0$ for all $i \in \mathcal{I}$, and the second inequality follows from the definition of $\mathcal{S}_1$, which implies that $t_i < t_{i^\star}$ for all $i \neq i^\star$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves indeed $\mathrm{SP}_1(\boldsymbol{t})$. As $t_i < t_{i^\star}$ for all $i \neq i^\star$, the above arguments also reveal that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_1(\boldsymbol{t})$ must coincide with $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ on $\mathcal{T}_1$, that is, $\boldsymbol{p}(\boldsymbol{t}) = \boldsymbol{p}^\star(\boldsymbol{t})$ and $\boldsymbol{q}(\boldsymbol{t}) = \boldsymbol{q}^\star(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{T}_1$.

As for the induction step, assume that for all $l < k$ we know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_l(\boldsymbol{t})$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_l(\boldsymbol{t})$ must satisfy $\boldsymbol{p}(\boldsymbol{t}) = \boldsymbol{p}^\star(\boldsymbol{t})$ and $\boldsymbol{q}(\boldsymbol{t}) = \boldsymbol{q}^\star(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{S}_l$. Fix now any $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints in $\mathrm{SP}_k(\boldsymbol{t})$ ensure that

$$\sum_{i \in \mathcal{I}} (p_i(\boldsymbol{t}') t_i' - q_i(\boldsymbol{t}') c_i) = \sum_{i \in \mathcal{I}} (p_i^\star(\boldsymbol{t}') t_i' - q_i^\star(\boldsymbol{t}') c_i) \quad \forall \boldsymbol{t}' \in \cup_{l=1}^{k-1} \mathcal{S}_l.$$

As $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is optimal in $\mathrm{SP}_l(\boldsymbol{t}')$ thanks to the induction hypothesis, this equality implies that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. By the second part of the induction hypothesis, this in turn implies that $\boldsymbol{p}(\boldsymbol{t}') = \boldsymbol{p}^\star(\boldsymbol{t}')$ and $\boldsymbol{q}(\boldsymbol{t}') = \boldsymbol{q}^\star(\boldsymbol{t}')$ for all $\boldsymbol{t}' \in \mathcal{S}_l$ and for all $l < k$. By the definition of $\mathcal{S}_k$, there are exactly $k-1$ agents $i \neq i^\star$ with types $t_i \geq t_{i^\star}$. For any such agent i, scenario $\boldsymbol{t}$ is i-unilaterally reachable from $(\underline{t}_i, \boldsymbol{t}_{-i})$. Note that $t_{i^\star} = \bar{t}_{i^\star} > \underline{t}_{i^\star} \geq \underline{t}_i$ for all $i \in \mathcal{I}$, where the equality follows from the definition of $\mathcal{T}_I$, and the second inequality follows from the definition of $i^\star$. This implies that $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{S}_{k-1}$. Therefore, we know from the induction hypothesis that $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = p_i^\star(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$. By Lemma 1(ii), we then have $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \neq i^\star$ with $t_i \geq t_{i^\star}$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

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

where the first inequality holds because $c_i > 0$ for all $i \in \mathcal{I}$ and because $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \neq i^\star$ with $t_i \geq t_{i^\star}$. The second inequality holds because $\boldsymbol{t} \in \mathcal{T}_I$, which implies that $t_i - c_i \leq t_{i^\star}$ for all $i \in \mathcal{I}$. Similarly, the equality holds because $\boldsymbol{t} \in \mathcal{T}_I$, in which case the payoff generated by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ amounts

to $t_{i^\star}$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. In addition, as $t_{i^\star} = \bar{t}_{i^\star}$ and $t_i - c_i \leq \underline{t}_{i^\star} < \bar{t}_{i^\star}$ for all $i \neq i^\star$, the two inequalities in the above expression can collapse to equalities only if $p_i(\boldsymbol{t}) = 0$ for every $i \neq i^\star$, $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$. Put differently, $(\boldsymbol{p}, \boldsymbol{q})$ can only be optimal in $\mathrm{SP}_k(\boldsymbol{t})$ if $\boldsymbol{p}(\boldsymbol{t}) = \boldsymbol{p}^\star(\boldsymbol{t})$ and $\boldsymbol{q}(\boldsymbol{t}) = \boldsymbol{q}^\star(\boldsymbol{t})$. As $\boldsymbol{t} \in \mathcal{S}_k$ was chosen arbitrarily, this observation completes the induction step.

Step 2 ($\mathcal{T}_{II}$). Define $\mathcal{S}_{I+1} = \mathcal{T}_{II}$, and set $k = I+1$. Next, fix an arbitrary scenario $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. From Step 1, we thus know that $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{\ell=1}^{k-1} \mathcal{S}_\ell = \mathcal{T}_I$. Clearly, $\boldsymbol{t}$ is $i^\star$-unilaterally reachable from $(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star})$. As $(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_I$ and $p_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) - q_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 1$, Lemma 1(i) implies that $p_{i^\star}(\boldsymbol{t}) = 1$. Thus, we have $t_{i^\star} - q_{i^\star}(\boldsymbol{t}) c_{i^\star} \leq t_{i^\star}$, that is, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, which implies that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. Also, as $c_{i^\star} > 0$, the mechanism $(\boldsymbol{p}, \boldsymbol{q})$ can attain the optimal value $t_{i^\star}$ of $\mathrm{SP}_k(\boldsymbol{t})$ only if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$, that is, only if $\boldsymbol{p}(\boldsymbol{t}) = \boldsymbol{p}^\star(\boldsymbol{t})$ and $\boldsymbol{q}(\boldsymbol{t}) = \boldsymbol{q}^\star(\boldsymbol{t})$.

Step 3 ($\mathcal{T}_{III}$). We partition $\mathcal{T}_{III}$ into I subsets of the form

$$\mathcal{S}_k = \{\boldsymbol{t} \in \mathcal{T}_{III} : |\{i \in \mathcal{I} : t_i > \underline{t}_{i^\star}\}| = k - I - 1\},$$

for $k = I+2, \ldots, 2I+1$. We will use induction on k to show that, for any $\boldsymbol{t} \in \mathcal{S}_k$, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$ with optimal value $\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_k(\boldsymbol{t})$ must satisfy

$$\sum_{i \in \arg\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}} p_i(\boldsymbol{t}) = 1 \quad \text{and} \quad p_i(\boldsymbol{t}) = q_i(\boldsymbol{t}) \quad \forall i \in \arg\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}. \tag{12}$$

As for the base step corresponding to $k = I+2$, fix any $\boldsymbol{t} \in \mathcal{S}_{I+2}$, and consider a mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < I+2$. From Steps 1 and 2 we thus know that $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{l=1}^{I+1} \mathcal{S}_l = \mathcal{T}_I \cup \mathcal{T}_{II}$. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k - I - 1 = 1$, there exists exactly one agent $i^\circ \neq i^\star$ with $t_{i^\circ} > \underline{t}_{i^\star}$. By the definition of $\mathcal{T}_{III}$, agent i° is the only agent whose adjusted type satisfies $t_{i^\circ} - c_{i^\circ} > \underline{t}_{i^\star}$. Note that $\boldsymbol{t}$ is i°-unilaterally reachable from $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ})$. As $\underline{t}_{i^\circ} - c_{i^\circ} \leq \underline{t}_{i^\star}$ by the definition of the favored agent $i^\star$, we further have $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I \cup \mathcal{T}_{II}$. The reasoning in Steps 1 and 2 thus implies that $p_{i^\circ}(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, which in turn implies via Lemma 1(ii) that $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Indeed, we have

$$\sum_{i \in \mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \leq p_{i^\circ}(\boldsymbol{t})(t_{i^\circ} - c_{i^\circ}) + \sum_{i \in \mathcal{I} \setminus \{i^\circ\}} p_i(\boldsymbol{t}) t_i \leq \max_{i \in \mathcal{I}} t_i - c_i = \sum_{i \in \mathcal{I}} (p^\star_i(\boldsymbol{t}) t_i - q^\star_i(\boldsymbol{t}) c_i),$$

where the first inequality holds because $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$ and because $c_i > 0$ for all $i \in \mathcal{I}$, whereas the second inequality holds because $t_{i^\circ} - c_{i^\circ} = \max_{i \neq i^\star} t_i - c_i > \underline{t}_{i^\star}$ and because $t_i \leq \underline{t}_{i^\star}$ for all $i \neq i^\circ$. Finally, the equality holds because $\boldsymbol{t} \in \mathcal{T}_{III}$, in which case $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generates a payoff of $\max_{i \in \mathcal{I}} t_i - c_i$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. In addition, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ can equal $\max_{i \neq i^\star} t_i - c_i = t_{i^\circ} - c_{i^\circ}$ only if $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t}) = 1$. As $\boldsymbol{t} \in \mathcal{S}_k$ was chosen freely, this completes the base step.

As for the induction step, fix any $k \in \{I+3, \ldots, 2I+1\}$. Assume that for all $l \in \{I+2, \ldots, k-1\}$ we know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_l(\boldsymbol{t})$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_l(\boldsymbol{t})$ satisfies (12) for all $\boldsymbol{t} \in \mathcal{S}_l$. Fix now any $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints in $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. Thus, $(\boldsymbol{p}, \boldsymbol{q})$ satisfies $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \mathcal{T}_I \cup \mathcal{T}_{II}$, which follows from Steps 1 and 2, and it satisfies (12) for all $\boldsymbol{t}' \in \cup_{l=I+2}^{k-1} \mathcal{S}_l$, which follows from the induction hypothesis. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k \in \{I+3, \ldots, 2I+1\}$, there are exactly $k - I - 1$ agents $i \in \mathcal{I}$ with types $t_i > \underline{t}_{i^\star}$. For any such agent i, scenario $\boldsymbol{t}$ is i-unilaterally reachable from $(\underline{t}_i, \boldsymbol{t}_{-i})$. Note that either $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II}$ or $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \cup_{l=I+2}^{k-1} \mathcal{S}_l$ because $\underline{t}_{i^\star} \geq \underline{t}_i$ for all $i \in \mathcal{I}$. We now show that $p_i(\underline{t}_i, \boldsymbol{t}_{-i})$ must vanish in both cases. If $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II}$, then we have $i \neq i^\star$, in which case $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = p^\star_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$. If $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \cup_{l=I+2}^{k-1} \mathcal{S}_l$, on the other hand, then the definition of $i^\star$ implies that $\underline{t}_i - c_i \leq \underline{t}_{i^\star}$, and the definition of $\mathcal{T}_{III}$ implies

that $\max_{i'\in\mathcal{I}\setminus\{i^\star\}} t_{i'} - c_{i'} > \underline{t}_{i^\star}$. Hence, i is no element of $\arg\max_{i'\in\mathcal{I}\setminus\{i^\star\}} t_{i'} - c_{i'}$, implying that $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$ thanks to (12). Lemma 1(ii) now ensures that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}$ with $t_i > \underline{t}_{i^\star}$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \le \sum_{i\in\mathcal{I}:\, t_i > \underline{t}_{i^\star}} p_i(\boldsymbol{t})(t_i - c_i) + \sum_{i\in\mathcal{I}:\, t_i \le \underline{t}_{i^\star}} p_i(\boldsymbol{t})t_i \le \max_{i\in\mathcal{I}} t_i - c_i = \sum_{i\in\mathcal{I}} (p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the second inequality holds because $\max_{i\neq i^\star} t_i - c_i > \underline{t}_{i^\star}$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$, and $(\boldsymbol{p}, \boldsymbol{q})$ can solve $\mathrm{SP}_k(\boldsymbol{t})$ only if it obeys (12). This observation completes the induction step. □

Proof of Proposition 3. Relaxing the incentive compatibility constraints and the first inequality in (FC) yields

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

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 $p_i \ge 0$ and $\sum_{i\in\mathcal{I}} p_i(\boldsymbol{t}) \le 1$ for all $\boldsymbol{t} \in \mathcal{T}$, we further have

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

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

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

where the first inequality holds because $\delta_{\underline{\boldsymbol{\mu}}} \in \mathcal{P}$, the second inequality holds because $\max_{i\in\mathcal{I}} t_i \ge t_{i^\star}$ 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 $\nu^\star \ge \bar{t}_{i^\star}$. 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 3 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 $\max_{i\in\mathcal{I}} t_i - c_i > \nu^\star \ge t_{i^\star}$, 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 $\max_{i\in\mathcal{I}} t_i - c_i \ge \nu^\star \ge t_{i^\star}$, 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 3. □

Proof of Lemma 3. For any $\boldsymbol{t}\in\mathcal{T}$, we will show that there exists a scenario $\hat{\boldsymbol{t}}\in\mathcal{T}$ that satisfies $\max_{i\neq i^\star}\hat{t}_i<\underline{\mu}_{i^\star}$ 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 $\max_{i\neq i^\star}\hat{t}_i<\underline{\mu}_{i^\star}$. 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 $\max_{i\neq i^\star}\hat{t}_i<\underline{\mu}_{i^\star}$. 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 $\prod_{i\in\mathcal{I}}[\underline{\mu}_i-\varepsilon,\underline{\mu}_i+\varepsilon]\subseteq\{\boldsymbol{t}\in\mathcal{T}:\max_{i\neq i^\star}t_i<\underline{\mu}_{i^\star}\}$. 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

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

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 $\prod_{i\in\mathcal{I}}[\underline{\mu}_i-\varepsilon,\underline{\mu}_i+\varepsilon]\subseteq\{\boldsymbol{t}\in\mathcal{T}:\max_{i\neq i^\star}t_i<\underline{\mu}_{i^\star}\}$ 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 4. Throughout the proof, we denote by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ the favored-agent mechanism of type (ii) with favored agent $i^\star$, where $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$, and with threshold value $\nu^\star = \bar{t}_{i^\star}$. We will use the following partition of the type space.

$$\begin{aligned}
\mathcal{T}_I &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i < \underline{\mu}_{i^\star} \text{ and } t_{i^\star} = \underline{\mu}_{i^\star}\}\\
\mathcal{T}_{II} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i < \underline{\mu}_{i^\star} \text{ and } t_{i^\star} \neq \underline{\mu}_{i^\star}\}\\
\mathcal{T}_{III} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i \geq \underline{\mu}_{i^\star} \text{ and } t_{i^\star} = \bar{t}_{i^\star}\}\\
\mathcal{T}_{IV} &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star} \text{ and } \max_{i\neq i^\star} t_i \geq \underline{\mu}_{i^\star} \text{ and } t_{i^\star} < \bar{t}_{i^\star}\}\\
\mathcal{T}_V &= \{\boldsymbol{t}\in\mathcal{T} : \max_{i\neq i^\star} t_i - c_i \geq \bar{t}_{i^\star}\}
\end{aligned} \tag{13}$$

Note that some of the conditions in the above definitions are redundant and are only introduced for ease of readability. The sets $\mathcal{T}_I$ and $\mathcal{T}_{II}$ are nonempty and contain at least $\underline{\boldsymbol{\mu}}$ and $(\bar{t}_{i^\star}, \underline{\boldsymbol{\mu}}_{-i^\star})$, respectively, because $\underline{\mu}_i < \underline{\mu}_{i^\star} < \bar{t}_{i^\star}$ for all $i \neq i^\star$. However, the sets $\mathcal{T}_{III}, \mathcal{T}_{IV}$ and $\mathcal{T}_V$ can be empty if $\underline{\mu}_{i^\star}$, $\bar{t}_{i^\star}$ or c_i, $i \neq i^\star$, are sufficiently large.

From Theorem 3 we already know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is robustly optimal. To prove Lemma 4, we will leverage Corollary 1 of Proposition 1, which provides sufficient conditions for the claim made. In particular, to show that 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 $\boldsymbol{t} \in \mathcal{T}$, it suffices to verify the conditions (i) and (ii) in Proposition 1 for some partition of the type space $\mathcal{T}$. We will show that they holds for a refinement of the partition $\mathcal{T}_I - \mathcal{T}_V$. Specifically, we will construct a refined partition $\mathcal{S}_1, \ldots, \mathcal{S}_m$ of $\mathcal{T}$ with $m = 2I + 2$, where $\mathcal{S}_1$ coincides with $\mathcal{T}_I$, $\mathcal{S}_2$ coincides with $\mathcal{T}_{II}$, $\mathcal{S}_3, \ldots, \mathcal{S}_{I+1}$ forms a partition of $\mathcal{T}_{III}$, $\mathcal{S}_{I+2}$ coincides with $\mathcal{T}_{IV}$ and $\mathcal{S}_{I+3}, \ldots, \mathcal{S}_{2I+2}$ forms a partition of $\mathcal{T}_V$.

We first exploit Lemma 3 to show that condition (i) holds. The subsequent arguments only depend on the definitions of $\mathcal{S}_1 = \mathcal{T}_I$ and $\mathcal{S}_2 = \mathcal{T}_{II}$. In contrast, the definitions of $\mathcal{S}_l$ for $l \in \{3, \ldots, m\}$ do not play a role for proving condition (i). Condition (i) of Proposition 1 requires that, for any index $k \in \{1, \ldots, m\}$ and for any scenario $\boldsymbol{t} \in \mathcal{S}_k$, there exists $\mathbb{P} \in \mathcal{P} \cap \mathcal{P}_0(\cup_{l=1}^k \mathcal{S}_l)$ with $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. Given any scenario $\boldsymbol{t} \in \mathcal{T}$, by Lemma 3, there exists 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 $\mathbb{E}_{\mathbb{P}}[\tilde{t}_i] = \underline{\mu}_i$ for all $i \in \mathcal{I}$, $\mathbb{P}(\tilde{\boldsymbol{t}} \in \{\boldsymbol{t}, \hat{\boldsymbol{t}}\}) = 1$, and $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. Consider first any scenario $\boldsymbol{t} \in \mathcal{S}_1$. By the definition of $\mathcal{S}_1$, we have $t_{i^\star} = \underline{\mu}_{i^\star}$. Hence, the identity $\mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}] = \underline{\mu}_{i^\star}$ can be satisfied only if $\hat{t}_{i^\star} = \underline{\mu}_{i^\star}$. As $\max_{i\neq i^\star} \hat{t}_i < \underline{\mu}_{i^\star}$, this means that $\hat{\boldsymbol{t}} \in \mathcal{S}_1$ and $\mathbb{P} \in \mathcal{P} \cap \mathcal{P}_0(\mathcal{S}_1)$. Condition (i) thus holds for $k = 1$ and for any $\boldsymbol{t} \in \mathcal{S}_1$. For any $k \in \{2, \ldots, m\}$ and $\boldsymbol{t} \in \mathcal{S}_k$, condition (i) can be easily verified thanks to Lemma 3 and because $\hat{\boldsymbol{t}} \in \mathcal{S}_1 \cup \mathcal{S}_2 = \{\boldsymbol{t} \in \mathcal{T} : \max_{i\neq i^\star} t_i < \underline{\mu}_{i^\star}\}$, which implies that the discrete distribution $\mathbb{P}$ is supported on $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_k$.

Next, we prove condition (ii). To this end, the rest of the proof is divided into five steps focusing on the five subsets $\mathcal{T}_I - \mathcal{T}_V$ and, in particular, on their respective refined partitions. We will exploit the non-locality of the incentive compatibility constraint (IC) via Lemma 1 to show iteratively that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves the scenario problem $\mathrm{SP}_k(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathcal{S}_k$ and $k = 1, \ldots, m$. The subsequent reasoning critically relies on the definition of the mechanism $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, under which the principal's payoff in scenario $\boldsymbol{t}$ amounts to $t_{i^\star}$ when $\max_{i\neq i^\star} t_i - c_i < \bar{t}_{i^\star}$ (*i.e.*, when $\boldsymbol{t} \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}_{IV}$) and to $\max_{i\in\mathcal{I}} t_i - c_i$ when $\max_{i\neq i^\star} t_i - c_i \geq \bar{t}_{i^\star}$ (*i.e.*, when $\boldsymbol{t} \in \mathcal{T}_V$).

Step 1 ($\mathcal{T}_I$). Define $\mathcal{S}_1 = \mathcal{T}_I$, and set $k = 1$. Fix an arbitrary scenario $\boldsymbol{t} \in \mathcal{S}_1$ and mechanism $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$. Note that any $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{X}$ is feasible in $\mathrm{SP}_1(\boldsymbol{t})$. In addition, the objective value of $(\boldsymbol{p}, \boldsymbol{q})$ is dominated by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

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

where the first inequality holds because $c_i > 0$ for all $i \in \mathcal{I}$, and the second inequality follows from the definition of $\mathcal{S}_1$, which implies that $t_i < t_{i^\star}$ for all $i \neq i^\star$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_1(\boldsymbol{t})$. As $t_i < t_{i^\star}$

for all $i \neq i^\star$, the above arguments also reveal that any mechanism $(\boldsymbol{p},\boldsymbol{q})$ that solves $\mathrm{SP}_1(\boldsymbol{t})$ must satisfy $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$ for any $\boldsymbol{t}\in\mathcal{S}_1=\mathcal{T}_I$.

Step 2 ($\mathcal{T}_{II}$). Define $\mathcal{S}_2=\mathcal{T}_{II}$. Fix an arbitrary scenario $\boldsymbol{t}\in\mathcal{S}_2$ and mechanism $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_2(\boldsymbol{t})$. The constraints of $\mathrm{SP}_2(\boldsymbol{t})$ ensure that $(\boldsymbol{p},\boldsymbol{q})$ solves $\mathrm{SP}_1(\boldsymbol{t}')$ for every $\boldsymbol{t}'\in\mathcal{S}_1=\mathcal{T}_I$. From Step 1, we know that any such solution satisfies $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$ for all $\boldsymbol{t}'\in\mathcal{T}_I$. Clearly, $\boldsymbol{t}$ is $i^\star$-unilaterally reachable from $(\underline{\mu}_{i^\star},\boldsymbol{t}_{-i^\star})$. As $(\underline{\mu}_{i^\star},\boldsymbol{t}_{-i^\star})\in\mathcal{T}_I$ and $p_{i^\star}(\underline{\mu}_{i^\star},\boldsymbol{t}_{-i^\star})-q_{i^\star}(\underline{\mu}_{i^\star},\boldsymbol{t}_{-i^\star})=1$, Lemma 1(i) implies that $p_{i^\star}(\boldsymbol{t})=1$. Thus, we have $t_{i^\star}-q_{i^\star}(\boldsymbol{t})c_{i^\star}\leq\bar{t}_{i^\star}$, that is, the objective function value of $(\boldsymbol{p},\boldsymbol{q})$ in $\mathrm{SP}_2(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$, which implies that $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_2(\boldsymbol{t})$. Also, as $c_{i^\star}>0$, the mechanism $(\boldsymbol{p},\boldsymbol{q})$ can attain the optimal value $t_{i^\star}$ only if $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$.

Step 3 ($\mathcal{T}_{III}$). We partition $\mathcal{T}_{III}$ into $I-1$ subsets of the form

$$\mathcal{S}_k=\left\{\boldsymbol{t}\in\mathcal{T}_{III}:|\{i\in\mathcal{I}:t_i\geq\underline{\mu}_{i^\star}\}|=k-1\right\},$$

for $k=3,\ldots,I+1$. Note that $|\{i\in\mathcal{I}:t_i\geq\underline{\mu}_{i^\star}\}|\geq 2$ for all $\boldsymbol{t}\in\mathcal{T}_{III}$ because $t_{i^\star}=\bar{t}_{i^\star}>\underline{\mu}_{i^\star}$ and $\max_{i\neq i^\star}t_i\geq\underline{\mu}_{i^\star}$ by the definition of $\mathcal{T}_{III}$. We will use induction on k to show that, for any $\boldsymbol{t}\in\mathcal{S}_k$, the mechanism $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$, and any mechanism $(\boldsymbol{p},\boldsymbol{q})$ that solves $\mathrm{SP}_k(\boldsymbol{t})$ must satisfy $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$.

As for the base step corresponding to $k=3$, fix any $\boldsymbol{t}\in\mathcal{S}_3$ and consider a mechanism $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p},\boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}'\in\mathcal{S}_l$ and $l<k$. From Steps 1 and 2 we thus know that $p_{i^\star}(\boldsymbol{t}')=1$ and $q_{i^\star}(\boldsymbol{t}')=0$ for all $\boldsymbol{t}'\in\cup_{l=1}^2\mathcal{S}_l=\mathcal{T}_I\cup\mathcal{T}_{II}$. As $\boldsymbol{t}\in\mathcal{S}_k$ and $k-1=2$, there exists exactly one agent $i^\circ\neq i^\star$ with $t_{i^\circ}\geq\underline{\mu}_{i^\star}$. Note that $\boldsymbol{t}$ is i°-unilaterally reachable from $(\underline{t}_{i^\circ},\boldsymbol{t}_{-i^\circ})$. As $\underline{t}_{i^\circ}<\underline{\mu}_{i^\circ}<\underline{\mu}_{i^\star}$ by the definition of the favored agent $i^\star$, we further have $(\underline{t}_{i^\circ},\boldsymbol{t}_{-i^\circ})\in\mathcal{T}_{II}$. The reasoning in Step 2 thus implies that $p_{i^\circ}(\underline{t}_i,\boldsymbol{t}_{-i})=0$, which in turn implies via Lemma 1(ii) that $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p},\boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$. Indeed, we have

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

where the first inequality holds because $p_{i^\circ}(\boldsymbol{t})=q_{i^\circ}(\boldsymbol{t})$ and because $c_i>0$ for all $i\in\mathcal{I}$, whereas the second inequality holds because $t_{i^\circ}-c_{i^\circ}<\bar{t}_{i^\star}=t_{i^\star}$ and because $t_i<\underline{\mu}_{i^\star}$ for all $i\neq i^\circ,i^\star$. Thus, $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. In addition, the objective function value of $(\boldsymbol{p},\boldsymbol{q})$ can equal $t_{i^\star}$ only if $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$. As $\boldsymbol{t}\in\mathcal{S}_k$ was chosen freely, this completes the base step.

As for the induction step, fix any $k\in\{4,\ldots,I+1\}$. Assume that for all $l\in\{3,\ldots,k-1\}$ we know that $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_l(\boldsymbol{t})$ and that any mechanism $(\boldsymbol{p},\boldsymbol{q})$ that solves $\mathrm{SP}_l(\boldsymbol{t})$ satisfies $p_{i^\star}(\boldsymbol{t})=q_{i^\star}(\boldsymbol{t})=1$ for all $\boldsymbol{t}\in\mathcal{S}_l$. Fix now any $\boldsymbol{t}\in\mathcal{S}_k$ and mechanism $(\boldsymbol{p},\boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints in $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p},\boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}'\in\mathcal{S}_l$ and $l<k$. Thus, $(\boldsymbol{p},\boldsymbol{q})$ satisfies $p_{i^\star}(\boldsymbol{t}')=1$ and $q_{i^\star}(\boldsymbol{t}')=0$ for all $\boldsymbol{t}'\in\mathcal{T}_I\cup\mathcal{T}_{II}\cup(\cup_{l=3}^{k-1}\mathcal{S}_l)$, which follows from Steps 1 and 2 and from the induction hypothesis. As $\boldsymbol{t}\in\mathcal{S}_k$ and $k\in\{4,\ldots,I+1\}$, there are exactly $k-1$ agents $i\in\mathcal{I}$ with types $t_i\geq\underline{\mu}_{i^\star}$. For any such agent $i\neq i^\star$, scenario $\boldsymbol{t}$ is i-unilaterally reachable from $(\underline{t}_i,\boldsymbol{t}_{-i})$. As $(\underline{t}_i,\boldsymbol{t}_{-i})\in\mathcal{S}_{k-1}$, we have $p_i(\underline{t}_i,\boldsymbol{t}_{-i})=0$ thanks to our induction hypothesis. Lemma 1(ii) now ensures that $p_i(\boldsymbol{t})=q_i(\boldsymbol{t})$ for all $i\in\mathcal{I}$ with $t_i\geq\underline{\mu}_{i^\star}$ and $i\neq i^\star$. Thus, the objective function value of any feasible $(\boldsymbol{p},\boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$, that is,

$$\sum_{i\in\mathcal{I}}(p_i(\boldsymbol{t})t_i-q_i(\boldsymbol{t})c_i)\leq\sum_{\substack{i\in\mathcal{I}\setminus\{i^\star\}:\\ t_i\geq\underline{\mu}_{i^\star}}}p_i(\boldsymbol{t})(t_i-c_i)+\sum_{\substack{i\in\mathcal{I}:\\ t_i<\underline{\mu}_{i^\star}}}p_i(\boldsymbol{t})t_i+p_{i^\star}(\boldsymbol{t})t_{i^\star}\leq\bar{t}_{i^\star}=\sum_{i\in\mathcal{I}}(p_i^\star(\boldsymbol{t})t_i-q_i^\star(\boldsymbol{t})c_i),$$

where the second inequality holds because $\max_{i\neq i^\star}t_i-c_i<\bar{t}_{i^\star}$. Thus, $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. Also, $(\boldsymbol{p},\boldsymbol{q})$ can solve $\mathrm{SP}_k(\boldsymbol{t})$ only if it satisfies $p_{i^\star}(\boldsymbol{t})=1$ and $q_{i^\star}(\boldsymbol{t})=0$. This observation completes the induction step.

Step 4 ($\mathcal{T}_{IV}$). Define $\mathcal{S}_{I+2} = \mathcal{T}_{IV}$, and set $k = I+2$. Next, fix an arbitrary scenario $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. From Steps 1, 2, and 3, we thus know that $p_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{\ell=1}^{k-1} \mathcal{S}_\ell = \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$. Clearly, $\boldsymbol{t}$ is $i^\star$-unilaterally reachable from $(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_{III}$. As $p_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) - q_{i^\star}(\bar{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 1$, by Lemma 1(i), we have $p_{i^\star}(\boldsymbol{t}) = 1$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by $t_{i^\star} - q_{i^\star}(\boldsymbol{t}) c_{i^\star} \leq t_{i^\star}$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. Also, as $c_{i^\star} > 0$, the mechanism $(\boldsymbol{p}, \boldsymbol{q})$ can attain the optimal value $t_{i^\star}$ of $\mathrm{SP}_k(\boldsymbol{t})$ only if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$.

Step 5 ($\mathcal{T}_V$). We partition $\mathcal{T}_V$ into I subsets of the form

$$\mathcal{S}_k = \left\{ \boldsymbol{t} \in \mathcal{T}_V \,:\, |\{i \in \mathcal{I} : t_i \geq \bar{t}_{i^\star}\}| = k - I - 2 \right\},$$

for $k = I+3, \ldots, 2I+2$. We will use induction on k to show that, for any $\boldsymbol{t} \in \mathcal{S}_k$, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$ with optimal value $\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}$, and any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_k(\boldsymbol{t})$ must satisfy

$$\sum_{i \in \arg\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}} p_i(\boldsymbol{t}) = 1 \quad \text{and} \quad p_i(\boldsymbol{t}) = q_i(\boldsymbol{t}) \quad \forall i \in \arg\max_{i' \in \mathcal{I}} t_{i'} - c_{i'}. \tag{14}$$

As for the base step corresponding to $k = I+3$, fix any $\boldsymbol{t} \in \mathcal{S}_{I+3}$, and consider a mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < I+3$. From Steps 1, 2, 3 and 4, we thus know that $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{l=1}^{I+1} \mathcal{S}_l = \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}_{IV}$. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k - I - 2 = 1$, there exists exactly one agent i° with $t_{i^\circ} \geq \bar{t}_{i^\star}$. By the definition of $\mathcal{T}_V$, agent i° is the only agent whose adjusted type satisfies $t_{i^\circ} - c_{i^\circ} \geq \bar{t}_{i^\star}$ so that we must have $i^\circ \neq i^\star$. Note that $\boldsymbol{t}$ is i°-unilaterally reachable from $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ})$. As $\underline{t}_{i^\circ} - c_{i^\circ} < \underline{\mu}_{i^\circ} < \underline{\mu}_{i^\star} < \bar{t}_{i^\star}$ by the definition of the favored agent $i^\star$, we further have $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}_{IV}$. The reasoning in Steps 1, 2, 3 and 4 thus implies that $p_{i^\circ}(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, which in turn implies via Lemma 1(ii) that $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Indeed, we have

$$\sum_{i \in \mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) \leq p_{i^\circ}(\boldsymbol{t})(t_{i^\circ} - c_{i^\circ}) + \sum_{i \in \mathcal{I} \setminus \{i^\circ\}} p_i(\boldsymbol{t}) t_i \leq \max_{i \in \mathcal{I}} t_i - c_i = \sum_{i \in \mathcal{I}} (p^\star_i(\boldsymbol{t}) t_i - q^\star_i(\boldsymbol{t}) c_i),$$

where the first inequality holds because $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$ and because $c_i > 0$ for all $i \in \mathcal{I}$, whereas the second inequality holds because $t_{i^\circ} - c_{i^\circ} = \max_{i \neq i^\star} t_i - c_i \geq \bar{t}_{i^\star}$ and because $t_i - c_i < t_i < \bar{t}_{i^\star}$ for all $i \neq i^\circ$. Finally, the equality holds because $\boldsymbol{t} \in \mathcal{T}_V$, in which case $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generates a payoff of $\max_{i \in \mathcal{I}} t_i - c_i$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. In addition, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ can equal $\max_{i \in \mathcal{I}} t_i - c_i = t_{i^\circ} - c_{i^\circ}$ only if $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t}) = 1$. As $\boldsymbol{t} \in \mathcal{S}_k$ was chosen freely, this completes the base step.

As for the induction step, fix any $k \in \{I+4, \ldots, 2I+2\}$. Assume that for all $l \in \{I+3, \ldots, k-1\}$ we know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_l(\boldsymbol{t})$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_l(\boldsymbol{t})$ satisfies (14) for all $\boldsymbol{t} \in \mathcal{S}_l$. Fix now any $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints in $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. Thus, $(\boldsymbol{p}, \boldsymbol{q})$ satisfies $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}_{IV}$, which follows from Steps 1, 2, 3 and 4, and it satisfies (14) for all $\boldsymbol{t}' \in \cup_{l=I+3}^{k-1} \mathcal{S}_l$, which follows from the induction hypothesis. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k \in \{I+4, \ldots, 2I+2\}$, there are exactly $k - I - 2$ agents $i \in \mathcal{I}$ with types $t_i \geq \bar{t}_{i^\star}$. For any such agent i, scenario $\boldsymbol{t}$ is i-unilaterally reachable from $(\underline{t}_i, \boldsymbol{t}_{-i})$. Note that either $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}_{IV}$ or $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \cup_{l=I+3}^{k-1} \mathcal{S}_l$ because $\underline{t}_i < \underline{\mu}_i < \underline{\mu}_{i^\star} < \bar{t}_{i^\star}$ for all $i \in \mathcal{I}$. We now show that $p_i(\underline{t}_i, \boldsymbol{t}_{-i})$ must vanish in both cases. If $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} \cup \mathcal{T}_{IV}$, then we have $i \neq i^\star$, in which case $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = p^\star_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$. If $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \cup_{l=I+3}^{k-1} \mathcal{S}_l$, on the other hand, then the definition of $i^\star$ implies that $\underline{t}_i - c_i < \bar{t}_{i^\star}$, and the definition of $\mathcal{T}_V$ implies that $\max_{i' \in \mathcal{I} \setminus \{i^\star\}} t_{i'} - c_{i'} \geq \bar{t}_{i^\star}$. Hence, i is no element of $\arg\max_{i' \in \mathcal{I} \setminus \{i^\star\}} t_{i'} - c_{i'}$, implying that $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$ thanks to (14). Lemma 1(ii)

now ensures that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}$ with $t_i \geq \bar{t}_{i^\star}$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

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

where the second inequality holds because $\max_{i\neq i^\star} t_i - c_i \geq \bar{t}_{i^\star}$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$, and $(\boldsymbol{p}, \boldsymbol{q})$ can solve $\mathrm{SP}_k(\boldsymbol{t})$ only if it obeys (14). This observation completes the induction step. □

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 4 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 4, 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 4 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

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i)\right] =& \mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i < \nu^\star\right) \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i) \,\middle|\, \max_{i\neq i^\star} \tilde{t}_i - c_i < \nu^\star\right] \\
&+ \mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i = \nu^\star\right) \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i) \,\middle|\, \max_{i\neq i^\star} \tilde{t}_i - c_i = \nu^\star\right] \\
&+ \mathbb{P}\left(\max_{i\neq i^\star} \tilde{t}_i - c_i > \nu^\star\right) \mathbb{E}_{\mathbb{P}}\left[\sum_{i\in\mathcal{I}}(p_i(\tilde{\boldsymbol{t}})\tilde{t}_i - q_i(\tilde{\boldsymbol{t}})c_i) \,\middle|\, \max_{i\neq i^\star} \tilde{t}_i - c_i > \nu^\star\right].
\end{aligned} \tag{15}$$

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 (15) 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 $\tilde{t}_{i^\star}$. 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 (15), 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 4. □

Proof of Lemma 5. 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

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

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 $\hat{t}_i \geq \underline{t}_i$. We thus showed that $\hat{\boldsymbol{t}}\in\mathcal{T}$.

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

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

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 6. Throughout the proof, we denote by $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ the favored-agent mechanism of type (i) with favored agent $i^\star$, where $\arg\max_{i\in\mathcal{I}} \underline{\mu}_i = \{i^\star\}$, and threshold value $\nu^\star = \underline{\mu}_{i^\star}$. We will use the following partition of $\mathcal{T}$.

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

Note that $\mathcal{T}_I$ and $\mathcal{T}_{III}$ are nonempty because they contain at least $(\overline{\mu}_{i^\star}, \underline{\boldsymbol{\mu}}_{-i^\star})$ and $\underline{\boldsymbol{\mu}}$, respectively, but the sets $\mathcal{T}_{II}$, $\mathcal{T}_{IV}$ and $\mathcal{T}_V$ can be empty if $\underline{\mu}_{i^\star}$ or c_i are sufficiently large for all $i\neq i^\star$.

From Theorem 5 we know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ is robustly optimal. To prove Lemma 6, we will leverage Corollary 1 of Proposition 1, which provides sufficient conditions for the claim made. In particular, to show that 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 $\boldsymbol{t}\in\mathcal{T}$, it suffices to verify the conditions (i) and (ii) in Proposition 1 for some partition of the type space $\mathcal{T}$. We will show that this holds for a refinement of the partition $\mathcal{T}_I - \mathcal{T}_V$. Specifically, we will construct a refined partition $\mathcal{S}_1, \ldots, \mathcal{S}_m$ of $\mathcal{T}$ with $m = 3I-1$, where $\mathcal{S}_1$ coincides with $\mathcal{T}_I$, $\mathcal{S}_2, \ldots, \mathcal{S}_I$ forms a partition of $\mathcal{T}_{II}$, $\mathcal{S}_{I+1}$ coincides with $\mathcal{T}_{III}$, $\mathcal{S}_{I+2}, \ldots, \mathcal{S}_{2I}$ forms a partition of $\mathcal{T}_{IV}$, and $\mathcal{S}_{2I+1}, \ldots, \mathcal{S}_{3I-1}$ forms a partition of $\mathcal{T}_V$.

We will formally define the refined partition $\mathcal{S}_1, \ldots, \mathcal{S}_m$ while proving condition (ii). We prove condition (ii) in five steps, focusing on the five subsets $\mathcal{T}_I, \ldots, \mathcal{T}_V$ and, in particular, on their respective refined partitions. We will exploit the non-locality of the incentive compatibility constraint (IC) via Lemma 1 to show iteratively that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves the scenario problem $\mathrm{SP}_k(\boldsymbol{t})$ for all $\boldsymbol{t}\in\mathcal{S}_k$ and $k = 1, \ldots, m$. The subsequent reasoning critically relies on the definition of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ under which the principal's payoff in scenario $\boldsymbol{t}$ amounts to $t_{i^\star}$ when $\max_{i\neq i^\star} t_i - c_i \leq \underline{\mu}_{i^\star}$ (*i.e.*, when $\boldsymbol{t}\in\mathcal{T}_I\cup\mathcal{T}_{II}\cup\mathcal{T}_{III}$) and to $\max_{i\in\mathcal{I}} t_i - c_i$ when $\max_{i\neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}$ (*i.e.*, when $\boldsymbol{t}\in\mathcal{T}_{IV}\cup\mathcal{T}_V$).

Step 1 ($\mathcal{T}_I$). Define $\mathcal{S}_1 = \mathcal{T}_I$, and set $k = 1$. Fix an arbitrary scenario $\boldsymbol{t}\in\mathcal{S}_1$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})\in\mathcal{X}$. Note that any $(\boldsymbol{p}, \boldsymbol{q})\in\mathcal{X}$ is feasible in $\mathrm{SP}_1(\boldsymbol{t})$. In addition, the objective value of $(\boldsymbol{p}, \boldsymbol{q})$ is dominated by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

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

where the first inequality holds because $c_i > 0$ for all $i\in\mathcal{I}$, and the second inequality follows from the definition of $\mathcal{S}_1$, which implies that $t_i < t_{i^\star}$ for all $i\neq i^\star$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_1(\boldsymbol{t})$. As $t_i < t_{i^\star}$ for all $i\neq i^\star$, the above arguments also reveal that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_1(\boldsymbol{t})$ must satisfy $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ for any $\boldsymbol{t}\in\mathcal{S}_1 = \mathcal{T}_I$.

Step 2 ($\mathcal{T}_{II}$). We partition $\mathcal{T}_{II}$ into $I-1$ subsets of the form

$$\mathcal{S}_k = \left\{\boldsymbol{t}\in\mathcal{T}_{II} : |\{i\in\mathcal{I} : t_i > \underline{\mu}_{i^\star}\}| = k\right\},$$

for $k = 2, \ldots, I$. Note that $|\{i\in\mathcal{I} : t_i > \underline{\mu}_{i^\star}\}| \geq 2$ for all $\boldsymbol{t}\in\mathcal{T}_{II}$ because $t_{i^\star}\in(\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$ and $\max_{i\neq i^\star} t_i > \underline{\mu}_{i^\star}$ by the definition of $\mathcal{T}_{II}$. We will use induction on k to show that, for any $\boldsymbol{t}\in\mathcal{S}_k$, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$, and any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_k(\boldsymbol{t})$ must satisfy $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$.

As for the base step corresponding to $k = 2$, fix any $\boldsymbol{t}\in\mathcal{S}_2$ and consider a mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}'\in\mathcal{S}_l$ and $l < k$. From Step 1 we thus know that $p_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}'\in\mathcal{S}_1 = \mathcal{T}_I$. As $\boldsymbol{t}\in\mathcal{S}_k$ and $k = 2$, there exists exactly one agent $i^\circ \neq i^\star$ with $t_{i^\circ} > \underline{\mu}_{i^\star}$. Note that $\boldsymbol{t}$ is i°-unilaterally

reachable from $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ})$. As $\underline{t}_{i^\circ} < \underline{\mu}_{i^\circ} < \underline{\mu}_{i^\star}$ by the definition of the favored agent $i^\star$, we further have $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I$. The reasoning in Step 1 thus implies that $p_{i^\circ}(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, which in turn implies via Lemma 1(ii) that $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Indeed, we have

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

where the first inequality holds because $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$ and because $c_i > 0$ for all $i \in \mathcal{I}$, whereas the second inequality holds because $t_{i^\circ} - c_{i^\circ} \leq \underline{\mu}_{i^\star} < t_{i^\star}$ and because $t_i \leq \underline{\mu}_{i^\star}$ for all $i \neq i^\circ, i^\star$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. In addition, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ can equal $t_{i^\star}$ only if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$. As $\boldsymbol{t} \in \mathcal{S}_k$ was chosen freely, this completes the base step.

As for the induction step, fix any $k \in \{3, \ldots, I\}$. Assume that for all $l \in \{2, \ldots, k-1\}$ we know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_l(\boldsymbol{t})$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_l(\boldsymbol{t})$ satisfies $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$ for all $\boldsymbol{t} \in \mathcal{S}_l$. Fix now any $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints in $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. Thus, $(\boldsymbol{p}, \boldsymbol{q})$ satisfies $p_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{l=1}^{k-1} \mathcal{S}_l$, which follows from Step 1 and the induction hypothesis. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k \in \{3, \ldots, I\}$, there are exactly k agents $i \in \mathcal{I}$ with types $t_i > \underline{\mu}_{i^\star}$. For any such agent $i \neq i^\star$, scenario $\boldsymbol{t}$ is i-unilaterally reachable from $(\underline{t}_i, \boldsymbol{t}_{-i})$. As $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{S}_{k-1}$, we have $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$ thanks to our induction hypothesis. Lemma 1(ii) now ensures that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}$ with $t_i \geq \underline{\mu}_{i^\star}$ and $i \neq i^\star$. Thus, the objective function value of any feasible $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$, that is,

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq \sum_{\substack{i\in\mathcal{I}\setminus\{i^\star\}:\\ t_i > \underline{\mu}_{i^\star}}} p_i(\boldsymbol{t})(t_i - c_i) + \sum_{\substack{i\in\mathcal{I}:\\ t_i \leq \underline{\mu}_{i^\star}}} p_i(\boldsymbol{t})t_i + p_{i^\star}(\boldsymbol{t})t_{i^\star} \leq t_{i^\star} = \sum_{i\in\mathcal{I}} (p_i^\star(\boldsymbol{t})t_i - q_i^\star(\boldsymbol{t})c_i),$$

where the second inequality holds because $\max_{i\neq i^\star} t_i - c_i < \underline{\mu}_{i^\star}$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. Also, $(\boldsymbol{p}, \boldsymbol{q})$ can solve $\mathrm{SP}_k(\boldsymbol{t})$ only if it satisfies $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$. This observation completes the induction step.

Step 3 ($\mathcal{T}_{III}$). Define $\mathcal{S}_{I+1} = \mathcal{T}_{III}$, and set $k = I + 1$. Next, fix an arbitrary scenario $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. From Steps 1 and 2 we thus know that $p_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{\ell=1}^{k-1} \mathcal{S}_\ell = \mathcal{T}_I \cup \mathcal{T}_{II}$. Clearly, $\boldsymbol{t}$ is $i^\star$-unilaterally reachable from $(\overline{\mu}_{i^\star}, \boldsymbol{t}_{-i^\star}) \in \mathcal{T}_I \cup \mathcal{T}_{II}$. Hence, we have $p_{i^\star}(\overline{\mu}_{i^\star}, \boldsymbol{t}_{-i^\star}) - q_{i^\star}(\overline{\mu}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 1$, which in turn implies via Lemma 1(i) that $p_{i^\star}(\boldsymbol{t}) = 1$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by $t_{i^\star} - q_{i^\star}(\boldsymbol{t})c_{i^\star} \leq t_{i^\star}$. Hence, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. Also, as $c_{i^\star} > 0$, the mechanism $(\boldsymbol{p}, \boldsymbol{q})$ can attain the optimal value $t_{i^\star}$ of $\mathrm{SP}_k(\boldsymbol{t})$ only if $p_{i^\star}(\boldsymbol{t}) = 1$ and $q_{i^\star}(\boldsymbol{t}) = 0$.

Step 4 ($\mathcal{T}_{IV}$). We partition $\mathcal{T}_{IV}$ into $I - 1$ subsets of the form

$$\mathcal{S}_k = \left\{ \boldsymbol{t} \in \mathcal{T}_{IV} \,:\, |\{i \in \mathcal{I} : t_i > \underline{\mu}_{i^\star}\}| = k - I - 1 \right\},$$

for $k = I+2, \ldots, 2I$. Note that, by the definition of $\mathcal{T}_{IV}$, we have $i^\star \notin \{i \in \mathcal{I} : t_i > \underline{\mu}_{i^\star}\}$ for any $\boldsymbol{t} \in \mathcal{T}_{IV}$ because $t_{i^\star} = \underline{\mu}_{i^\star}$. We will use induction on k to show that, for any $\boldsymbol{t} \in \mathcal{S}_k$, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$ with optimal value $\max_{i'\in\mathcal{I}} t_{i'} - c_{i'}$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_k(\boldsymbol{t})$ must satisfy

$$\sum_{i \in \arg\max_{i'\in\mathcal{I}} t_{i'} - c_{i'}} p_i(\boldsymbol{t}) = 1 \quad \text{and} \quad p_i(\boldsymbol{t}) = q_i(\boldsymbol{t}) \quad \forall i \in \arg\max_{i'\in\mathcal{I}} t_{i'} - c_{i'}. \tag{16}$$

As for the base step corresponding to $k = I + 2$, fix any $\boldsymbol{t} \in \mathcal{S}_{I+2}$, and consider a mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and

$l < I+2$. From Steps 1, 2 and 3, we thus know that $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{l=1}^{I+1} \mathcal{S}_l = \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k - I - 1 = 1$, there exists exactly one agent $i^\circ \neq i^\star$ with $t_{i^\circ} > \underline{\mu}_{i^\star}$. Note that $\boldsymbol{t}$ is i°-unilaterally reachable from $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ})$. As $\underline{t}_{i^\circ} - c_{i^\circ} < \underline{\mu}_{i^\circ} < \underline{\mu}_{i^\star}$ by the definition of the favored agent $i^\star$, we further have $(\underline{t}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$. The reasoning in Steps 1, 2 and 3 thus implies that $p_{i^\circ}(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$, which in turn implies via Lemma 1(ii) that $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Indeed, we have

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq p_{i^\circ}(\boldsymbol{t})(t_{i^\circ} - c_{i^\circ}) + \sum_{i\in\mathcal{I}\setminus\{i^\circ\}} p_i(\boldsymbol{t})t_i \leq \max_{i\in\mathcal{I}} t_i - c_i = \sum_{i\in\mathcal{I}} (p^\star_i(\boldsymbol{t})t_i - q^\star_i(\boldsymbol{t})c_i),$$

where the first inequality holds because $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t})$ and because $c_i > 0$ for all $i \in \mathcal{I}$, whereas the second inequality holds because $t_{i^\circ} - c_{i^\circ} = \max_{i\neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}$ and because $t_i \leq \underline{\mu}_{i^\star}$ for all $i \neq i^\circ$. Finally, the equality holds because $\boldsymbol{t} \in \mathcal{T}_{IV}$, in which case $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generates a payoff of $\max_{i\in\mathcal{I}} t_i - c_i$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. In addition, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ can equal $\max_{i\neq i^\star} t_i - c_i = t_{i^\circ} - c_{i^\circ}$ only if $p_{i^\circ}(\boldsymbol{t}) = q_{i^\circ}(\boldsymbol{t}) = 1$. As $\boldsymbol{t} \in \mathcal{S}_k$ was chosen freely, this completes the base step.

As for the induction step, fix any $k \in \{I+3, \ldots, 2I\}$. Assume that for all $l \in \{I+2, \ldots, k-1\}$ we know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_l(\boldsymbol{t})$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_l(\boldsymbol{t})$ satisfies (16) for all $\boldsymbol{t} \in \mathcal{S}_l$. Fix now any $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints in $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. Thus, $(\boldsymbol{p}, \boldsymbol{q})$ satisfies $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$, which follows from Steps 1, 2 and 3, and it satisfies (16) for all $\boldsymbol{t}' \in \cup_{l=I+2}^{k-1} \mathcal{S}_l$, which follows from the induction hypothesis. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k \in \{I+3, \ldots, 2I\}$, there are exactly $k - I - 1$ agents $i \in \mathcal{I} \setminus \{i^\star\}$ with types $t_i > \underline{\mu}_{i^\star}$. For any such agent i, scenario $\boldsymbol{t}$ is i-unilaterally reachable from $(\underline{t}_i, \boldsymbol{t}_{-i})$. Note that either $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$ or $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \cup_{l=I+2}^{k-1} \mathcal{S}_l$ because $\underline{t}_i < \underline{\mu}_i < \underline{\mu}_{i^\star}$ for all $i \in \mathcal{I} \setminus \{i^\star\}$. We now show that $p_i(\underline{t}_i, \boldsymbol{t}_{-i})$ must vanish in both cases. If $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$, we have $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = p^\star_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$ because $i \neq i^\star$. If $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \cup_{l=I+2}^{k-1} \mathcal{S}_l$, on the other hand, then the definition of $i^\star$ implies that $\underline{t}_i - c_i < \underline{\mu}_{i^\star}$, and the definition of $\mathcal{T}_{IV}$ implies that $\max_{i'\in\mathcal{I}\setminus\{i^\star\}} t_{i'} - c_{i'} > \underline{\mu}_{i^\star}$. Hence, i is no element of $\arg\max_{i'\in\mathcal{I}\setminus\{i^\star\}} t_{i'} - c_{i'}$, implying that $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$ thanks to (16). Lemma 1(ii) now ensures that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}$ with $t_i > \underline{\mu}_{i^\star}$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

$$\sum_{i\in\mathcal{I}} (p_i(\boldsymbol{t})t_i - q_i(\boldsymbol{t})c_i) \leq \sum_{i\in\mathcal{I}:\, t_i > \underline{\mu}_{i^\star}} p_i(\boldsymbol{t})(t_i - c_i) + \sum_{i\in\mathcal{I}:\, t_i \leq \underline{\mu}_{i^\star}} p_i(\boldsymbol{t})t_i \leq \max_{i\in\mathcal{I}} t_i - c_i = \sum_{i\in\mathcal{I}} (p^\star_i(\boldsymbol{t})t_i - q^\star_i(\boldsymbol{t})c_i),$$

where the second inequality holds because $\max_{i\neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$, and $(\boldsymbol{p}, \boldsymbol{q})$ can solve $\mathrm{SP}_k(\boldsymbol{t})$ only if it obeys (16). This observation completes the induction step.

Step 5 ($\mathcal{T}_V$). We partition $\mathcal{T}_V$ into $I - 1$ subsets of the form

$$\mathcal{S}_k = \left\{ \boldsymbol{t} \in \mathcal{T}_V \,:\, |\{i \in \mathcal{I} \setminus \{i^\star\} : t_i > \underline{\mu}_{i^\star}\}| = k - 2I \right\},$$

for $k = 2I+1, \ldots, 3I-1$. We will use induction on k to show that, for any $\boldsymbol{t} \in \mathcal{S}_k$, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$ with optimal value $\max_{i'\in\mathcal{I}} t_{i'} - c_{i'}$, that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_k(\boldsymbol{t})$ must satisfy (16).

As for the base step corresponding to $k = 2I+1$, fix any $\boldsymbol{t} \in \mathcal{S}_{2I+1}$, and consider a mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints of $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < 2I+1$. From Steps 1, 2, and 3, we thus know that $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{l=1}^{I+1} \mathcal{S}_l = \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$, and from Step 4, we know that $(\boldsymbol{p}, \boldsymbol{q})$ satisfies (16) for all $\boldsymbol{t}' \in \cup_{l=I+2}^{2I} \mathcal{S}_l = \mathcal{T}_{IV}$. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k - 2I = 1$, there exists exactly one agent $i^\circ \neq i^\star$ with $t_{i^\circ} > \underline{\mu}_{i^\star}$. Now consider any agent $i \in \{i^\circ, i^\star\}$. Note that $\boldsymbol{t}$ is i-unilaterally reachable from $(\underline{\mu}_i, \boldsymbol{t}_{-i})$. If $i = i^\star$, then we have $(\underline{\mu}_{i^\circ}, \boldsymbol{t}_{-i^\circ}) \in \mathcal{T}_{IV}$. Our reasoning in Step 4 implies that

$p_i(\underline{\mu}_i, \boldsymbol{t}_{-i}) = 0$ as $\underline{\mu}_{i^\star} < \max_{i' \in \mathcal{I}} t'_i - c'_i$. If $i \neq i^\star$, on the other hand, we must have $(\underline{\mu}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$, because, $\underline{\mu}_i - c_i < \underline{\mu}_{i^\star}$ by the definition of the favored agent $i^\star$, and i° is the only agent in $\mathcal{I} \setminus \{i^\star\}$ with $t_{i^\circ} - c_{i^\circ} > \underline{\mu}_{i^\star}$. The reasoning in Steps 1, 2 and 3 thus implies that $p_i(\underline{\mu}_i, \boldsymbol{t}_{-i}) = 0$. Then, for all $i \in \{i^\circ, i^\star\}$, Lemma 1(ii) implies that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$. Indeed, we have

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

where the first inequality holds because $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \{i^\circ, i^\star\}$ and because $c_i > 0$ for all $i \in \mathcal{I}$, whereas the second inequality holds because $t_{i^\circ} - c_{i^\circ} = \max_{i \neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}$ and because $t_i \leq \underline{\mu}_{i^\star}$ for all $i \notin \{i^\circ, i^\star\}$. Finally, the equality holds because $\boldsymbol{t} \in \mathcal{T}_V$, in which case $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ generates a payoff of $\max_{i \in \mathcal{I}} t_i - c_i$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$. In addition, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ can equal $\max_{i \neq i^\star} t_i - c_i = t_{i^\circ} - c_{i^\circ}$ only if $(\boldsymbol{p}, \boldsymbol{q})$ satisfies (16). As $\boldsymbol{t} \in \mathcal{S}_k$ was chosen freely, this completes the base step.

As for the induction step, fix any $k \in \{2I+2, \ldots, 3I-1\}$. Assume that for all $l \in \{2I+1, \ldots, k-1\}$ we know that $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_l(\boldsymbol{t})$ and that any mechanism $(\boldsymbol{p}, \boldsymbol{q})$ that solves $\mathrm{SP}_l(\boldsymbol{t})$ satisfies (16) for all $\boldsymbol{t} \in \mathcal{S}_l$. Fix now any $\boldsymbol{t} \in \mathcal{S}_k$ and mechanism $(\boldsymbol{p}, \boldsymbol{q})$ feasible in $\mathrm{SP}_k(\boldsymbol{t})$. The constraints in $\mathrm{SP}_k(\boldsymbol{t})$ ensure that $(\boldsymbol{p}, \boldsymbol{q})$ solves $\mathrm{SP}_l(\boldsymbol{t}')$ for every $\boldsymbol{t}' \in \mathcal{S}_l$ and $l < k$. Thus, $(\boldsymbol{p}, \boldsymbol{q})$ satisfies $p_{i^\star}(\boldsymbol{t}') = p^\star_{i^\star}(\boldsymbol{t}') = 1$ and $q_{i^\star}(\boldsymbol{t}') = q^\star_{i^\star}(\boldsymbol{t}') = 0$ for all $\boldsymbol{t}' \in \cup_{l=1}^{I+1} \mathcal{S}_l = \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$, which follows from Steps 1, 2 and 3, and it satisfies (16) for all $\boldsymbol{t}' \in \cup_{l=I+2}^{k-1} \mathcal{S}_l$, which follows from Step 4 and the induction hypothesis. As $\boldsymbol{t} \in \mathcal{S}_k$ and $k \in \{2I+2, \ldots, 3I-1\}$, there are exactly $k - 2I$ agents $i \in \mathcal{I} \setminus \{i^\star\}$ with types $t_i > \underline{\mu}_{i^\star}$. Let i denote any such agent or agent $i^\star$. Then, scenario $\boldsymbol{t}$ is i-unilaterally reachable from $(\underline{\mu}_i, \boldsymbol{t}_{-i})$. Note that we have $(\underline{\mu}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_{IV}$ if $i = i^\star$, and we have either $(\underline{\mu}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$ or $(\underline{\mu}_i, \boldsymbol{t}_{-i}) \in \cup_{l=2I+1}^{k-1} \mathcal{S}_l$ if $i \neq i^\star$, because $\underline{\mu}_i < \underline{\mu}_{i^\star}$ follows from the definition of $i^\star$. We now show that $p_i(\underline{\mu}_i, \boldsymbol{t}_{-i})$ must vanish in any case. If $(\underline{\mu}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III}$, then we have $i \neq i^\star$, in which case $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = p^\star_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$. If $(\underline{\mu}_i, \boldsymbol{t}_{-i}) \in \mathcal{T}_{IV}$, then we have $i = i^\star$, in which case $p_{i^\star}(\underline{t}_{i^\star}, \boldsymbol{t}_{-i^\star}) = 0$ follows from (16) and $\max_{i \neq i^\star} t_i - c_i > \underline{\mu}_{i^\star}$. If $(\underline{t}_i, \boldsymbol{t}_{-i}) \in \cup_{l=I+2}^{k-1} \mathcal{S}_l$, on the other hand, then the definition of $i^\star$ implies that $\underline{t}_i - c_i < \underline{\mu}_{i^\star}$, and the definition of $\mathcal{T}_V$ implies that $\max_{i' \in \mathcal{I} \setminus \{i^\star\}} t_{i'} - c_{i'} > \underline{\mu}_{i^\star}$. Hence, i is no element of $\arg\max_{i' \in \mathcal{I} \setminus \{i^\star\}} t_{i'} - c_{i'}$, implying that $p_i(\underline{t}_i, \boldsymbol{t}_{-i}) = 0$ thanks to (16). Lemma 1(ii) now ensures that $p_i(\boldsymbol{t}) = q_i(\boldsymbol{t})$ for all $i \in \mathcal{I}$ with $t_i > \underline{\mu}_{i^\star}$ and for $i = i^\star$. Thus, the objective function value of $(\boldsymbol{p}, \boldsymbol{q})$ in $\mathrm{SP}_k(\boldsymbol{t})$ is bounded above by that of $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ because

$$\begin{aligned}\sum_{i \in \mathcal{I}} (p_i(\boldsymbol{t}) t_i - q_i(\boldsymbol{t}) c_i) &\leq \sum_{i \in \mathcal{I}:\, t_i > \underline{\mu}_{i^\star}} p_i(\boldsymbol{t})(t_i - c_i) + p_{i^\star}(\boldsymbol{t})(t_{i^\star} - c_{i^\star}) + \sum_{\substack{i \in \mathcal{I} \setminus \{i^\star\}: \\ t_i \leq \underline{\mu}_{i^\star}}} p_i(\boldsymbol{t}) t_i \\ &\leq \max_{i \in \mathcal{I}} t_i - c_i = \sum_{i \in \mathcal{I}} (p^\star_i(\boldsymbol{t}) t_i - q^\star_i(\boldsymbol{t}) c_i),\end{aligned}$$

where the second inequality holds because $\max_{i \neq i^\star} t_i - c_i > \underline{\mu}_{i^\star} \geq t_j$ for all $j \in \mathcal{I} \setminus \{i^\star\} : t_j \leq \underline{\mu}_{i^\star}$. Thus, $(\boldsymbol{p}^\star, \boldsymbol{q}^\star)$ solves $\mathrm{SP}_k(\boldsymbol{t})$, and $(\boldsymbol{p}, \boldsymbol{q})$ can solve $\mathrm{SP}_k(\boldsymbol{t})$ only if it obeys (16). This observation completes the induction step.

Next, we exploit Lemma 5 to show that condition (i) holds. Condition (i) of Proposition 1 requires that, for any index $k \in \{1, \ldots, 3I-1\}$ and for any scenario $\boldsymbol{t} \in \mathcal{S}_k$, there exists $\mathbb{P} \in \mathcal{P} \cap \mathcal{P}_0(\cup_{l=1}^k \mathcal{S}_l)$ with $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. Given any scenario $\boldsymbol{t} \in \mathcal{T}$ and any $\mu_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$, by Lemma 5, there exists 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 $\mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}] = \mu_{i^\star}$, $\mathbb{P}(\tilde{t}_i \in \{t_i, \hat{t}_i\}) = 1$ for all $i \in \mathcal{I}$, and $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$. We will show that, for any $k \in \{1, \ldots, 3I-1\}$ and for any scenario $\boldsymbol{t} \in \mathcal{S}_k$, there exists $\mu_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$ such that $\mathbb{P}$, as defined in Lemma 5, satisfies condition (i). To this end, we first derive some useful implications of Lemma 5 for arbitrary $\boldsymbol{t} \in \mathcal{T}$ and $\mu_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$. In the following, we use $\mathbb{S}$ to denote the support of the distribution $\mathbb{P}$. Note

that $\mathbb{S} \subseteq \{\boldsymbol{t}' \in \mathcal{T} : t'_i \in \{t_i, \hat{t}_i\} \text{ for all } i \in \mathcal{I}\}$. Then, as $\max_{i \neq i^\star} \hat{t}_i < \underline{\mu}_{i^\star}$, we have $|\{i \in \mathcal{I} \setminus \{i^\star\} : t'_i > \underline{\mu}_{i^\star}\}| \leq |\{i \in \mathcal{I} \setminus \{i^\star\} : t_i > \underline{\mu}_{i^\star}\}|$ for any $\boldsymbol{t}' \in \mathbb{S}$. For the same reason, we may also conclude that $\max_{i \neq i^\star} t'_i - c_i \leq \max\{\max_{i \neq i^\star} t_i - c_i, \underline{\mu}_{i^\star}\}$. We thus obtain

$$\mathbb{S} \subseteq \left\{\boldsymbol{t}' \in \mathcal{T} : f(\boldsymbol{t}') \leq f(\boldsymbol{t}) \text{ and } \max_{i \neq i^\star} t'_i - c_i \leq g(\boldsymbol{t})\right\}, \tag{17}$$

where $f(\boldsymbol{t}) = |\{i \in \mathcal{I} \setminus \{i^\star\} : t_i > \underline{\mu}_{i^\star}\}|$ and $g(\boldsymbol{t}) = \max\{\max_{i \neq i^\star} t_i - c_i, \underline{\mu}_{i^\star}\}$.

If $t_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$ and $\mu_{i^\star} = t_{i^\star}$, then the identities $\mathbb{E}_{\mathbb{P}}[\tilde{t}_{i^\star}] = \mu_{i^\star}$ and $\mathbb{P}(\tilde{t}_{i^\star} \in \{t_{i^\star}, \hat{t}_{i^\star}\}) = 1$ imply that $\hat{t}_{i^\star} = \mu_{i^\star}$. Together with (17), this in turn implies that

$$\mathbb{S} \subseteq \left\{\boldsymbol{t}' \in \mathcal{T} : f(\boldsymbol{t}') \leq f(\boldsymbol{t}) \text{ and } \max_{i \neq i^\star} t'_i - c_i \leq g(\boldsymbol{t}) \text{ and } t'_{i^\star} = t_{i^\star}\right\} \quad \text{if} \quad t_{i^\star} = \mu_{i^\star}. \tag{18}$$

We are now ready to prove condition (i). Consider first $k = 1$, and fix any $\boldsymbol{t} \in \mathcal{S}_1 = \mathcal{T}_I$. By the definition of $\mathcal{T}_I$, we have $f(\boldsymbol{t}) = 0$ and $g(\boldsymbol{t}) = \underline{\mu}_{i^\star}$. Observing that $t_{i^\star} \in (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$, we may choose $\mu_{i^\star} = t_{i^\star}$ and define $\mathbb{P}$ as in Lemma 5. By (18) and as $f(\boldsymbol{t}) = 0$ and $g(\boldsymbol{t}) = \underline{\mu}_{i^\star}$, we thus obtain

$$\mathbb{S} \subseteq \left\{\boldsymbol{t}' \in \mathcal{T} : t'_{i^\star} = t_{i^\star} \text{ and } \max_{i \neq i^\star} t'_i - c_i \leq \underline{\mu}_{i^\star} \text{ and } f(\boldsymbol{t}') \leq 0\right\} \subseteq \mathcal{S}_1,$$

where the second inclusion holds because $\mathcal{S}_1 = \mathcal{T}_I$ and because of the definition of $\mathcal{T}_I$. We have therefore shown that $\mathbb{P} \in \mathcal{P}_0(\cup_{l=1}^k \mathcal{S}_l) = \mathcal{P}_0(\mathcal{S}_1)$. By Lemma 5, we further have $\mathbb{P} \in \mathcal{P}$ and $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$, which completes the proof for $k = 1$.

Next, consider any $k \in \{2, \ldots, I\}$ and $\boldsymbol{t} \in \mathcal{S}_k \subseteq \mathcal{T}_{II}$. By the definitions of $\mathcal{S}_k$ and $\mathcal{T}_{II}$, we have $f(\boldsymbol{t}) = k - 1$ and $g(\boldsymbol{t}) = \underline{\mu}_{i^\star}$. As $t_{i^\star} \in (\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$, we may again choose $\mu_{i^\star} = t_{i^\star}$ and define $\mathbb{P}$ as in Lemma 5. By (18) and the identities $f(\boldsymbol{t}) = k - 1$ and $g(\boldsymbol{t}) = \underline{\mu}_{i^\star}$, we then find

$$\begin{aligned}\mathbb{S} &\subseteq \left\{\boldsymbol{t}' \in \mathcal{T} : t'_{i^\star} = t_{i^\star} \text{ and } \max_{i \neq i^\star} t'_i - c_i \leq \underline{\mu}_{i^\star} \text{ and } f(\boldsymbol{t}') \leq k - 1\right\} \\ &\subseteq \{\boldsymbol{t}' \in \mathcal{T}_I \cup \mathcal{T}_{II} : f(\boldsymbol{t}') \leq k - 1\} \subseteq \cup_{l=1}^k \mathcal{S}_l,\end{aligned}$$

where the second and third inclusions follow from the definitions of $\mathcal{T}_I, \mathcal{T}_{II}$ and $\mathcal{S}_k$. Together with Lemma 5, this observation implies that $\mathbb{P} \in \mathcal{P} \cap \mathcal{P}_0(\cup_{l=1}^k \mathcal{S}_l)$ and $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$.

Next, set $k = I + 1$, and consider any $\boldsymbol{t} \in \mathcal{S}_{I+1} = \mathcal{T}_{III}$. In this case, we again have $g(\boldsymbol{t}) = \underline{\mu}_{i^\star}$. Consider any fixed $\mu_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$ and $\mathbb{P}$ as defined in Lemma 5. By (17) and that $g(\boldsymbol{t}) = \underline{\mu}_{i^\star}$, we have $\mathbb{S} \subseteq \{\boldsymbol{t}' \in \mathcal{T} : \max_{i \neq i^\star} t'_i - c_i \leq \underline{\mu}_{i^\star}\} \subseteq \mathcal{T}_I \cup \mathcal{T}_{II} \cup \mathcal{T}_{III} = \cup_{l=1}^k \mathcal{S}_l$. Thus, we again have $\mathbb{P} \in \mathcal{P} \cap \mathcal{P}_0(\cup_{l=1}^k \mathcal{S}_l)$ and $\mathbb{P}(\tilde{\boldsymbol{t}} = \boldsymbol{t}) > 0$.

Next, consider any $k \in \{I + 2, \ldots, 2I\}$ and $\boldsymbol{t} \in \mathcal{S}_k \subseteq \mathcal{T}_{IV}$. By the definition of $\mathcal{S}_k$ in Step 4, we have $f(\boldsymbol{t}) = k - I - 1$. As $t_{i^\star} = \underline{\mu}_{i^\star}$ by the definition of $\mathcal{T}_{IV}$, we can choose $\mu_{i^\star} = t_{i^\star} = \underline{\mu}_{i^\star}$ and define $\mathbb{P}$ as in Lemma 5. We can now leverage (18) to conclude that $\mathbb{S} \subseteq \{\boldsymbol{t}' \in \mathcal{T} : t'_{i^\star} = \underline{\mu}_{i^\star} \text{ and } f(\boldsymbol{t}') \leq k - I - 1\}$, which is a subset of $\mathcal{T}_{III} \cup (\cup_{l=I+2}^k \mathcal{S}_l) \subseteq \cup_{l=1}^k \mathcal{S}_l$. This completes the proof for $k \in \{I + 2, \ldots, 2I\}$.

Finally, consider any $k \in \{2I + 1, \ldots, 3I - 1\}$ and any $\boldsymbol{t} \in \mathcal{S}_k \subseteq \mathcal{T}_V$. Consider any fixed $\mu_{i^\star} \in [\underline{\mu}_{i^\star}, \overline{\mu}_{i^\star}]$, and define $\mathbb{P}$ as in Lemma 5. By the definition of $\mathcal{S}_k$ in Step 5, we now have $f(\boldsymbol{t}) = k - 2I$. Together with (17), this implies that $\mathbb{S} \subseteq \{\boldsymbol{t}' \in \mathcal{T} : f(\boldsymbol{t}') \leq k - 2I\}$, which is a subset of $\cup_{l=1}^k \mathcal{S}_l$ by the definition of $\mathcal{S}_k$. The claim thus holds for $k \in \{2I + 1, \ldots, 3I - 1\}$. This completes the proof of condition (i). Hence, the claim follows. □