# SCREENING THROUGH COORDINATION

NEMANJA ANTIĆ[*] AND KAI STEVERSON[†]

Abstract. In various screening environments agents have preferences that are independent of their type, making standard techniques infeasible. We show how a principal who faces multiple agents and whose preferences exhibit complementarities can benefit by **coordinating** her actions. Coordination delivers a payoff gain to the principal in states with many high-quality agents and accepts a payoff loss in states with many low-quality agents. Coordination sometimes results in **strategic favoritism**: the principal may favor agents who are unambiguously inferior. While this behavior is often attributed to biases, we show that a rational principal optimally pursues this strategy if complementarities are sufficiently high.

JEL Codes: D82, D23

*Date*: February 23, 2017.

*Key words and phrases.* Mechanism Design, Screening, Coordination.

[*] Nemanja Antic: Kellogg School of Management, Managerial Economics and Decision Sciences Department, Northwestern University, Evanston, IL 60208. Email: *nemanja.antic@kellogg.northwestern.edu,* http://www.kellogg.northwestern.edu/faculty/antic/

[†] Kai Steverson: Institute for the Interdisciplinary Study of Decision Making, New York University. Email: *ksteverson@nyu.edu.* Website: https://sites.google.com/site/ksteverson/. Financial support from NIH grant R01DA038063 is gratefully acknowledged.

Thanks to Eddie Dekel, Tim Fedderson, Faruk Gul, Peter Klibanoff, Stephen Morris, Wolfgang Pesendorfer, Nicola Persico, Ariel Rubinstein, Doron Ravid, Karl Schlag and Asher Wolinsky for helpful discussions.

# 1. Introduction

Under standard screening techniques, the principal tailors her action to each agent's preferences by offering a menu of options. For example, in the classic monopoly screening setting, agents who report a high taste for the good are offered a high quantity at a high price, and agents who report a low taste for the good are offered a low quantity at a low price. This technique allows the principal to leverage the **structure of the agents' preferences** in order to price discriminate and increase her own profits.

However, in a number of natural economic situations, agents can have preferences that do not vary with their type, which makes the tailoring technique infeasible. Examples of such situations include: (1) a manager downsizing a division and deciding which employees to keep, but without the authorization to renegotiate salary levels; (2) an executive assembling a team from within an organization to work on a prestigious project; (3) a regulatory agency deciding which companies to investigate based on the company's own regulatory filings; (4) a grant administrator choosing which research applications to fund.

In these examples, the principal makes a binary decision relevant to each agent (e.g., keep or lay off) where one of the outcomes is clearly preferred by the agent. Since none of the examples allow for side payments or transfers, the intensity with which an agent desires his preferred outcome is irrelevant to his behavior: regardless of whether an agent prefers to keep his job a lot or a little, he always wants to maximizes his chance of doing so. Therefore, agents' preferences do not vary with their type, which makes the standard screening technique infeasible, i.e., the principal can gain no leverage from the structure of the agents' preferences.

Our contribution in this paper is to show how a principal who faces this situation can instead leverage the **structure of her own preferences**. The key feature of our model that makes this possible is that there are multiple agents across whom the principal's preferences exhibits **complementarities**. For example, in the case of an executive assembling a team, the complementarities would mean that adding a high-quality agent to the team has more value to the principal when there are other high-quality agents already present to work with him. Given this, we show how a principal can exploit a **coordination mechanism** to optimally screen the agents. A coordination mechanism works by performing well when there are strong complementarities to benefit from, and performing poorly when the complementarities are weak. For example, coordination would add high-quality agents to the team when there are many of them to work together and enhance each other's productivity through collaboration. Conversely, when the overall pool of agents is mediocre, coordination would accept placing low-quality agents on the team. Sometimes rewarding the low quality agent is necessary for truthful revelation: if only high-quality agents are placed on the team, then no agent would admit to being low quality. In this way, coordination accepts the failure necessary for incentive compatibility when the complementarities are weak, in order to succeed when there are strong complementarities to benefit from.

A counter-intuitive result of coordination is it can sometimes lead to favoring an agent who the principal knows to be inferior, a phenomenon we refer to as **strategic favoritism**. For example, this would mean an agent with a worse type distribution being placed on the team more often than an agent with a better type distribution, while assuming the agents are symmetric in all other ways. In other words, **strategic favoritism** involves rewarding the agent who is inferior for the principal's own aims. In practice, favoring an inferior agent is often assumed to be inefficient and to arise from the principal's biases. A key insight of our work is that favoritism can instead be strategic and optimal for an unbiased principal.

To more precisely motivate and discuss coordination and strategic favoritism, we now describe the mechanism design environment we study in this paper. A principal faces a number of agents without the use of transfers. The principal decides to take either action 1 or action 0 on each of the agents. Each agent prefer action 1 on himself, and cares about nothing else. Therefore, we can think of action 1 as the agent keeping his job, being placed on the project etc. Each agent has a production level that depends on the action taken on him. Production from action 0 is fixed and known to all players, while production from action 1 depends on the agent's private type. Agents draw their private types from a distribution specific to them, and the draws are assumed to be independent across agents. The principal's payoff is any increasing and supermodular function of the agent-specific production levels. The supermodularity is what gives the principal's preferences complementarities across the agents.

We say the principal's choice of action 0 or 1 is the **correct response** to an agent's type if it maximizes that agent's production level given his type. The **incorrect response** to an agent's type does the opposite. We call a type a high type if action 1 is the correct response and a low type if action 0 is the correct response. In the example of assembling a team, the principal would like to include high-type agents and exclude low-type agents. Since action 0 provides the same production regardless of type, an incorrect response on a high type provides the same production as a correct response on a low type. Therefore, high types always produce more than low types.

Before describing our general results, it will be useful to consider a special case that starkly demonstrates the logic of coordination. Specifically, consider the environment in which the agents are fully symmetric and have only two possible types: a high type and a low type. The optimal mechanism in this setting, as described in Theorem 1, correctly responds to every agent's type when the number of high-type agents exceeds a fixed threshold. Conversely, when the number of high-type agents falls below a lower fixed threshold, the optimal mechanism incorrectly responds to every agent's type. If the number of high-type agents falls between these two thresholds, then the optimal mechanism either takes action 1 on every agent or action 0 on every agent.

This gives a stark demonstration of how coordination works: by grouping the correct responses on states with more high-type agents, and grouping incorrect responses on states with more low-types agents. The complementarities rewards the principal for grouping high-output outcomes of the agents together, which coordination achieves by grouping correct responses on states with more high types. Conversely, sometimes responding incorrectly is necessary for truthful revelation: if the principal only responds correctly, then no agent would admit to being a low type. And coordination places those required incorrect responses where they do the least damage: on states with more low types. Therefore, coordination benefits the principal by succeeding when there are strong complementarities to benefit from, and accepting the failure necessary for incentive compatibility when the complementarities are weak.

Our general coordination result extends this logic to our full model with asymmetric agents who may have any number of types. We introduce a partial order on the states that measures the quality of the states and that has the feature that states higher in the partial order have more high types. We then define a coordination property which involves the mechanism grouping its correct responses at states higher in this partial order, and grouping its incorrect responses at states lower in the partial order. Theorem 2 establishes that this coordination property always occurs in an optimal mechanism, which demonstrates that it is the optimal way to leverage the structure of the principal's preferences.

A surprising implication of coordination is that it can lead the optimal mechanism to take action 1 more often on a less productive agent than on a more productive agent, a phenomenon we call **strategic favoritism**. By "less productive agent" we mean an agent with an unambiguously worse distribution of types, while assuming the agents are symmetric in all ways other than their type distribution. Hence, strategic favoritism rewards the agent who is inferior for the principal's own aims. In practice, giving preferential treatment to inferior agents is often attributed to biases of the principal (e.g., nepotism). We show instead how favoritism can be optimal for an unbiased, rational principal. More specifically, our Theorem 3 establishes that there always exist beliefs at which every optimal mechanism displays strategic favoritism, provided that the principal's preferences have enough complementarities. We characterize the requirement for "enough complementarities" using a simple, closed-form inequality. Moreover, when the complementarities are sufficiently high, the magnitude of the favoritism can be made quite large: the less productive agent can receive action 1 with as much as a fifty percent higher probability than the more productive agent.

One way in which coordination could manifest in practice is in making use of self-evaluations. Many organizations ask workers to self-evaluate; this can be part of deciding whom to lay off or part of a review process that determines internal promotions or placement. When self-evaluations have content for which verification is impractical, such content must not be trivially manipulable: it cannot be the case that a worker who evaluates himself favorably is always rewarded. The logic of coordination suggests a solution through conditioning the response to the self-evaluation of one worker on the collective self-evaluations of all workers, e.g., placing agents with favorable (unfavorable) self-evaluations on a prestigious project when the overall quality of self-evaluations is favorable (unfavorable). In the presence of complementarities, coordinating in this way benefits the organization by performing well when there are strong complementarities from which the principal can benefit.

Making use of self-evaluations is also important for investigative bodies, e.g., a government regulator that decides which firms to investigate, based in part on each firm's own regularity filings. Just as in the previous discussion, if such regulatory filings are to be useful, they cannot be trivially manipulable: even the most innocuous filing must sometimes be investigated. And in the presence of complementarities, coordination provides a useful way to proceed. Here, we are interpreting action 1 to mean leaving the firm alone and action 0 to mean investigating the firm. Therefore, complementarities can arise from the fact that failing to investigate a regulation-violating firm is damaging to competing firms.

Strategic favoritism in self-evaluations could take the form of holding higher-quality agents to a higher standard. For example, a more talented employee would have to submit a glowing self-evaluation to be placed on a valuable project, while a low-ability agent would only have to submit an average report. Similarly, firms less likely to be involved in regulatory violations need to submit more innocuous regulatory filings to avoid investigation. Holding higher-quality agents to a higher standard does not *necessarily* create favoritism, but can lead to it when done to such an extreme that the lower-ability employee becomes more likely to be assigned to the project or avoid the layoff.

We present two extensions to our model. First, we consider the robustness of our results to allowing correlation between agent types. We restrict ourselves to the case where each agent has only two possible types, and we establish a continuity result regarding the optimal coordination mechanism with certain kinds of positive correlation. In our baseline setting with independence, incentive compatibility requires all incentive constraints to bind. Introducing correlation relaxes this requirement, and the robustness result demonstrates that our main insights do not depend in a knife-edge way on everywhere-binding incentive constraints.

We also consider an extension that allows the agents to collude against the principal. Since coordination tends to take action 1 on high type agents when there are many of them, the agents can collude to all report being the highest possible type. However, we show that the principal can slightly modify the optimal mechanism to become coalition proof while achieving virtually the same payoff.

The rest of the paper is organized as follows. Section 1.1 discusses related literature. Section 2 provides a simple example that demonstrates our key ideas of coordination and strategic favoritism. Section 3 presents the formal model. Section 4 discusses how and why the optimal mechanism uses coordination. Section 5 discusses the occurrence of strategic favoritism in the optimal mechanism. Section 6 gives the extensions to the model. Section 7 concludes.

## 1.1. Literature Review

There is a large literature on screening with agents who have type-dependent preferences. Broadly speaking, this literature uses differences in the preferences of the agent to screen, and optimal screening is based on balancing the rewards from separating agents with the cost of doing so (this cost could be a transfer, if this is available, or an inefficiency in allocation).

As far as we are aware, all the papers on screening with type-independent preferences differ from our own work by providing the principal some way to "check" the agents' reports: this can take the form of (1) direct verification, where either the principal, or a third party, has the (possibly costly) ability to see the agent's type; (2) using repeated interactions to check reports against the probability law governing type distributions; or (3) checking reports across agents by exploiting correlation between agents' types. The papers in each of these three categories do not allow for our main channel of coordination due to either having a single agent or assuming the principal's preferences are separable across agents. In contrast, our paper does not assume the principal has any way to check the agents' reports.

Closely related worked by Ben-Porath, Dekel, and Lipman (2014) falls in the first category of direct verification of each agent's report at a cost. In their paper, the principal allocates a single, indivisible good among $n$ agents, who all want the good regardless of their type. The principal's payoff depends only on the type of the agent who receives the good, which rules out complementarities and hence our coordination channel. Rahman (2012) and Chassang and Padro i Miquel (2014) consider the presence of a monitor or whistle-blower who can verify the report of the agent for the principal. The mechanisms in these papers rely on (threats of) verification to incentivize truthful reporting.

Work by Guo and Hörner (2015), Li, Matouschek, and Powell (2015) and Lipnowski and Ramos (2015) falls into the second category of verifying the agents' reports using repeated interaction. These papers consider settings that resemble ours in that there are no transfers, the principal decides on a binary action, and agents want action 1 regardless of their type. However, these papers consider only a single agent, which rules out the possibility of coordination. The repeated interactions allow the type-dependent intensity of the agent's desire for action 1 to matter. The optimal mechanisms in these papers use a dynamic "budget" that limits how often the agent can receive action 1 across repeated interactions. Therefore, the agent has a greater incentive to ask for action 1 when his preference intensity is high, which can benefit a principal who prefers taking action 1 on types with higher preference intensity. In contrast, since we consider a single interaction with each agent, we cannot use such an inter-temporal trade-off to incentivize our agents. Furthermore, we do not require any assumption on how the principal's preferences align with the preference intensity of the agents.

The idea of imposing a budget across repeated interaction is also used in the older literature on "linking decisions" (see Jackson and Sonnenschein (2007), Rubinstein and Yaari (1983), and Fang and Norman (2006)). Differing from the papers discussed in the previous paragraph, this literature does not use discounted flows for the payoffs of players, but instead focuses on various kinds of limit results. The use of limit results allows for simpler budgets: in Jackson and Sonnenschein (2007), how often each agent can report a given type is determined by how likely that type is in the underlying type distribution and thus the reports are "verified" against the law of large numbers.

In our final category of using correlation to check agents' reports, we are unaware of any application to screening problems with type-independent preferences. Of course, exploiting such correlation is common in the broader mechanism design literature, for example in Cremer and Mclean (1988) as well as in the classic implementation literature (Maskin, 1999). In principle, those techniques could be applied to screening problems similar to ours.

## 2. Simple Example

We now present a simple example that demonstrates our key ideas of **coordination** and **strategic favoritism**. Suppose a principal is deciding whether to delegate a project to two agents whom we will call agent 1 and agent 2. The principal can delegate the project to one of the agents individually, to both agents jointly, or to neither agent and do the project herself. Each agent has a private type, which can be either low ($L$) or high ($H$), and that determines his suitability for the project. The agents draw their types independently of one another, and agents 1 and 2 have a $\frac{2}{3}$ and $\frac{1}{2}$ probability of having a high type, respectively. The agents are working on a fixed salary, so the principal does not have the use of transfers. Instead the principal can only decide whether to include (action 1) or not include (action 0) each agent on the project. Both agents always want to be included on the project regardless of their type, while the principal cares only about the overall value of the project.

Project value is normalized to zero if the principal does the project herself. Each high type agent assigned to the project increases its value by $\frac{1}{2}$. However, including even one low type agent sets the value back to zero, regardless of the other agent's type and whether or not he was included. In this way, the project value exhibits complementarities across the two agents, since working with a low type agent prevents a high type agent from adding value to the project.

We use notation such as $HL$ to denote the state where agent 1 is a high type and agent 2 is a low type. We also use notation such as 10 to indicate the principal including agent 1 on the project but not agent 2. Project value as a function of the state and action is summarized in table 1.

| Project Value | State and Action |
|---|---|
| 1 | if state = $HH$ and action = 11 |
| $\frac{1}{2}$ | if state = $HH$ and action = 10 or 01 |
| $\frac{1}{2}$ | if state = $HL$ and action = 10 |
| $\frac{1}{2}$ | if state = $LH$ and action = 01 |
| 0 | in all other cases |

Table 1: Principal's payoffs in the simple example

Before talking to the agents, the principal is able to commit to a mechanism specifying a contingent plan of action based on the agents' self-reports of their own type. The mechanism used by the principal must be incentive compatible, which requires that truthful reporting by all agents is an equilibrium.[1]

The principal faces the challenge that the agents have type-independent preferences, which raises the question of whether a mechanism helps at all, i.e., whether a principal can improve upon making the best decisions ex-ante (before the agents make reports). Without a mechanism, the principal's best course of action is to always include both agents on the project. This strategy yields an expected payoff of $\frac{1}{3}$ and is described in table 2a.

[1] More precisely, if all agents report truthfully, then no agent can strictly raise his probability of being included by lying.

| State | $HH$ | $LH$ | $HL$ | $LL$ |
|---|---|---|---|---|
| Action | 11 | 11 | 11 | 11 |

Table 2a: Best ex-ante strategy

| State | $HH$ | $LH$ | $HL$ | $LL$ |
|---|---|---|---|---|
| Action | 11 | 01 | 01 | 11 |

Table 2b: Coordination mechanism

Under this strategy, at state $LH$ the high type of agent 2 is included on the project but adds no value because he is paired with a low type of agent 1, which sets the project value to zero. And the inclusion of a high type always has an implicit cost since, for incentive compatibility to hold, it must be balanced by including a low type of the same agent at a different state. Therefore, the inclusion of both agents at state $LH$ is inefficient since it provides no value but imposes a cost. By committing to a mechanism, the principal can fix this inefficiency by coordinating her correct and incorrect responses to the agents' types. By "correct response" we mean including high type agents and excluding low type agents.

Concretely, suppose the principal correctly responded at states $HH$ and $LH$ and incorrectly responded at states $HL$ and $LL$. This approach is an example of a **coordination mechanism** and is described in table 2b.

This coordination mechanism yields an expected payoff of $\frac{5}{12}$, which improves upon the no-mechanism payoff of $\frac{1}{3}$. This improvement comes through leveraging the complementarities in the principal's payoff function by grouping together the correct responses to the agents' types, and placing them in states with relatively more high types. Grouping the correct responses together avoids the inefficiency we saw earlier, since a high-type agent is never paired with a low-type agent. And placing the correct responses in states with more high types uses the correct responses where they have the most value. In contrast, correctly responding at state $LL$ has no value to the principal at all since the project value will always be zero at that state.

We now check whether the above coordination mechanism is incentive compatible. Agent 2 is always placed on the project and so can do no better than telling the truth. Agent 1 also has no incentive to lie since he has a 50-50 chance of being placed on the project regardless of what type he reports. Notice that the incentive constraint of both agents bind making them indifferent among all their reports. This is not an artifact of the current example; everywhere-binding incentive constraints will be a feature of our general setting.

The coordination mechanism also displays **strategic favoritism**. Specifically, agent 2 is placed on the project more often than agent 1 despite being less likely to be a high type. Hence, coordination resulted in rewarding the inferior agent by including him on the project more often. A key decision that led to strategic favoritism was that we grouped $LH$ with the "high states," where the mechanism responds correctly, and we grouped $HL$ with the "low states," where the mechanism responds incorrectly. This decision may seem arbitrary since $HL$ and $LH$ have the same number of high types. However, reversing this decision would have violated the incentive constraint of agent 2. Under this reversal, agent 2 is placed on the project if and only if he makes the same report as agent 1. And since agent 1 is more likely to be a high type than a low type, agent 2 would have a strict incentive to always report being a high type. This shows that the coordination mechanism was not arbitrarily constructed. In fact, the coordination mechanism is optimal in that it yields the highest payoff to the principal among all incentive-compatible mechanisms.[2]

## 3. Model

A single principal faces $n$ agents denoted by $i \in \{1, ..., n\}$. Agent $i$ draws a privately known type from the finite set $\Theta_i \subset \mathbb{R}$ according to distribution $\mu_i \in \Delta\Theta_i$. The total state space is $\Theta$, where $\Theta := \prod_{i=1}^{n} \Theta_i$. Agents draw their types independently, and the principal and agents share a common prior $\mu \in \Delta\Theta$ defined by $\mu(\theta) := \prod_{i=1}^{n} \mu_i(\theta_i)$, for all $\theta \in \Theta$. Without loss of generality we can suppose that $\mu$ is full support since each $\Theta_i$ can be edited to remove types that never occur. As a standard shorthand, we use $\Theta_{-i}$ and $\mu_{-i}$ to refer to the state space and prior with agent $i$ removed.[3]

For each agent $i$, the principal chooses a binary action, $a_i \in \{0, 1\}$. Therefore, the principal's total action space is is given by $A := \{0, 1\}^n$. Agent $i$ prefers $a_i = 1$ over $a_i = 0$ and cares about nothing else. The intensity of the preference for action 1 does not matter; agent $i$ will always seek to maximize the probability of $a_i = 1$.

Each agent has a production function given by $X_i : \Theta_i \times \{0, 1\} \to \mathbb{R}_{++}$, where $X_i(\theta_i, a_i)$ is the production of agent $i$ of type $\theta_i$ when receiving action $a_i$. The type of the agent is only relevant when action 1 is taken; action 0 always leads to the same production. In other

[2] Including stochastic mechanisms.

[3] More precisely, we set $\Theta_{-i} = \prod_{j \neq i} \Theta_j$ and we define $\mu_{-i} \in \Delta\Theta_{-i}$ by setting $\mu_{-i}(t) = \prod_{j \neq i} \mu_j(t_j)$ for any $t \in \Theta_{-i}$.

words, we assume that $X_i(\theta_i, 0) = X_i(\theta_i', 0)$ for all $\theta_i', \theta_i \in \Theta_i$. Without loss of generality, we consider the case where production from action 1 increases with type, i.e., $\theta_i' > \theta_i$ implies $X_i(\theta_i', 1) > X_i(\theta_i, 1)$.

For a state $\theta \in \Theta$ and action profile $a \in A$, we will write $X(\theta, a)$ to denote the vector of production across all agents so that $X(\theta, a) := (X_1(\theta_1, a_1), ..., X_n(\theta_n, a_n))$. The principal's payoff at state $\theta$ when using action profile $a$ is given by

$$V(\theta, a) := W(X(\theta, a)),$$

where $W : \mathbb{R}^n_{++} \to \mathbb{R}_{++}$ is any strictly increasing and strictly supermodular[4] function. The properties of $W$ ensure that the principal's payoff is increasing in the production of each agent and exhibits complementarities across agents.

We will refer to type $\theta_i$ of agent $i$ as a "high type" if $X_i(\theta_i, 1) > X_i(\theta_i, 0)$,and we refer to $\theta_i$ as a "low type" if $X_i(\theta_i, 1) < X_i(\theta_i, 0)$. Since production from action 0 is constant across types, it is easy to see that a high type of agent $i$ will always have weakly higher production than a low type of agent $i$, regardless of what actions are used. We let $\Theta_i^H$ and $\Theta_i^L$ denote the set of high and low types, respectively, for agent $i$.

We say the principal correctly responds to an agent's type when setting $a_i = 1$ if and only if agent $i$'s type is high. Incorrectly responding to an agent's type means setting $a_i = 1$ if and only if agent $i$'s type is low.

To avoid trivial cases, we assume that $\Theta_i^H$ and $\Theta_i^L$ are both non-empty for each agent. If either $\Theta_i^H$ or $\Theta_i^L$ is empty, then the principal knows ex ante what action she should take on agent $i$, and the problem could be rewritten for the remaining $n-1$ agents. For ease of exposition, we will also assume that $\Theta_i^L \cup \Theta_i^H = \Theta_i$ for each agent $i$, which only rules out types whose production does not depend on the action.

## The Principal's Design Problem

The principal commits to a mechanism, which assigns a lottery over $A$ to every state in $\Theta$. We are restricting attention to direct mechanisms, since the revelation principle applies in our setting. In other words, a mechanism is any function $g : \Theta \to \Delta A$, and we let $\mathcal{G}$ be

[4] Strict supermodularity can be defined as follows, for any $X, X' \in \mathbb{R}^n$: $W(X \vee X') + W(X \wedge X') \geq W(X) + W(X')$, with the inequality strict whenever $X \ngeq X'$ and $X' \ngeq X$.

the space of all mechanisms. We use $g(\theta)[a]$ to denote the probability mechanism $g$ assigns to action $a$ at state $\theta$.

For any mechanism $g$, state $\theta$ and agent $i$, we let $g_i(\theta) := \sum_{a \in A | a_i = 1} g(\theta)[a]$ be the probability that $g$ sets $a_i = 1$ conditional on state $\theta$. For any mechanism $g$, agent $i$, and $\theta_i \in \Theta_i$, define

$$p_i^g(\theta_i) := \sum_{\theta_{-i} \in \Theta_{-i}} g_i(\theta_i, \theta_{-i}) \mu_{-i}(\theta_{-i}),$$

as the probability that $g$ sets $a_i = 1$, conditional on agent $i$ reporting type $\theta_i$ and all other agents reporting truthfully. We can use $p_i^g(\theta_i)$ to provide a useful and tight characterization of incentive compatibility.

**Lemma 1.** *A mechanism $g$ is incentive-compatible if and only if $p_i^g(\theta_i) = p_i^g(\theta_i')$ for every agent $i$ and all $\theta_i, \theta_i' \in \Theta_i$.*

When deciding what report to make, each agent cares only about the probability of $a_i = 1$. Therefore, if $p_i^g(\theta_i) > p_i^g(\theta_i')$, then agent $i$ would never report type $\theta_i'$. Hence, truthful reporting requires that $p_i^g(\theta_i)$ give the same value for all $\theta_i \in \Theta_i$. And conversely, any agent is happy to report truthfully if he receives action 1 with the same probability regardless of his report. Stated another way, incentive compatibility holds if and only if all incentive constraints of every agent are binding.

Since there are no transfers, the worst case for each agent is always getting action 0. Hence, individual rationality constraints do not bind and can be safely disregarded. We can now write the principal's design problem with which the rest of the paper will be concerned:

$$\max_{g \in \mathcal{G}} \sum_{\theta \in \Theta, a \in A} \mu(\theta) g(\theta)[a] V(\theta, a),$$
$$\text{s.t.} \qquad p_i^g(\theta_i) = p_i^g(\theta_i') \text{ for all } i, \theta_i, \theta_i' \in \Theta_i.$$

We call a mechanism *optimal* if it solves the above maximization problem.

### Generalizations of Agent Preferences

We conclude our model discussion by providing two ways in which the preferences of the agents could be generalized without substantially altering the analysis that will follow. Lemma 1 would hold as stated if some agents always preferred action 1 and other agents always preferred action 0. Therefore, the set of optimal mechanisms is invariant to these changes in the preferences of the agents. We restricted ourselves to the case that all agents prefer action 1 only for ease of exposition.

A more substantial modification to the preference of the agents would be to allow some of the low types of each agent to prefer action 0; we refer to such types as passive. In terms of the applications, this may arise if, for example, an agent does not want to be placed on a project for which he is sufficiently unsuited. Passive agents would willingly reveal their unsuitably for action 1 to the principal, and hence any optimal mechanism will always take action 0 on them. This does not interfere with the incentives of the non-passive agents, since they prefer action 1 and thus have no incentive to pretend to be passive. Hence the design problem on the remaining non-passive agents resembles our baseline setting, and our main results could be generalized to this case.

Moreover, if the passive types publicly announce their type prior to non-passive agents making their reports, then the design problem exactly reduces to our baseline setting applied to the remaining agents once the passive types are removed. A public announcement could take the form of a passive agent not showing up to the meeting where a team is being selected or not completing a required self-evaluation.

## 4. Coordination

We begin our discussion of coordination by considering environments with fully symmetric agents who only have two possible types. The two-type symmetric environment provides a stark demonstration of coordination that highlights the underlying logic. We then provide a general result that extends the logic of coordination into the full model.

We first formally define the two-type-symmetric environment.

**Definition 1.** *We say the environment is* ***two-type symmetric*** *if there exists* $L, H \in \mathbb{R}_{++}$ *with* $H > L$*, and* $p \in (0,1)$ *such that:*

1. $\Theta_i^L = \{L\}$ *and* $\Theta_i^H = \{H\}$ *for all* $i$.
2. $\mu_i(H) = p$ *for all* $i$.
3. $X_i(\cdot,\cdot) = X_j(\cdot,\cdot)$ *for all* $i, j$.
4. $W(Y) = W(Y')$ *whenever* $Y$ *and* $Y'$ *are permutations of each other.*

We use $\overrightarrow{\mathbf{1}}$ and $\overrightarrow{\mathbf{0}}$ to indicate taking action 1 on every agent and taking action 0 on every agent, respectively.

**Theorem 1.** *In any two-type symmetric environment, there exists $m^H \geq m^L$ and $a^* \in \{\overrightarrow{\mathbf{1}}, \overrightarrow{\mathbf{0}}\}$ such that there is an optimal mechanism that:*

(1) *Correctly responds to every agent at states with strictly more than $m^H$ high types.*
(2) *Incorrectly responds to every agent at states with strictly less than $m^L$ high types.*
(3) *Plays $a^*$ at states with strictly between $m^L$ and $m^H$ high types.*

*Proof.* See Appendix. □

Theorem 1 provides a stark demonstration of how coordination works: by grouping correct responses on states with more high types, and grouping incorrect responses on states with more low types. The complementarities in the principal's payoff function rewards her for grouping high-output outcomes of the agents together, which coordination achieves by grouping correct responses on states with more high types. Sometimes responding incorrectly is necessary for truthful revelation: if a mechanism only responds correctly, then no agent would admit to being a low type. Hence, coordination accepts the failure necessary for incentive compatibility when there is little to lose, in order to succeed when there are strong complementarities to benefit from.

While omitted from the statement above, the proof of Theorem 1 also specifies what occurs at states with exactly $m^H$ or $m^L$ high types. At these cutoff states, the optimal mechanism mixes between the actions used at states just above and below the cutoff. For example, at states with exactly $m^H$ high types the optimal mechanism mixes between correctly responding to every agent and $a^*$.

Theorem 1 does not explicitly characterize the cutoffs $m^H$ and $m^L$. However, if the complementarities are strong enough, then the optimal mechanism will perform what we call "complete coordination" by setting $m^H = m^L$ and only using the "correctly respond to every agent" or "incorrectly respond to every agent" actions. In that case, the value of the cutoff $m^H = m^L$ can be calculated from the incentive constraints. In the appendix, we calculate the cutoff, and provide closed-form inequalities that characterize the requirement of enough complementarities for complete coordination to occur.

We now return to our full model with asymmetric agents who can have any number of types. To define our general notion of coordination, we introduce a partial order on the state space that gives a measure of the overall quality of the agents.

**Definition 2.** *Let $\succeq^*$ be the partial order on $\Theta$ such that $\theta' \succeq^* \theta$ if and only if for each agent $i$, either (i) $\theta'_i = \theta_i$ or (ii) $\theta_i$ is a low type and $\theta'_i$ is a high type. Moreover, we will write $\theta' \succ^* \theta$ to indicate $\theta' \succeq^* \theta$ and $\theta' \neq \theta$.*

In words, $\theta' \succ^* \theta$ indicates $\theta'$ and $\theta$ differ only in that some low type agents in state $\theta$ become high type agents in state $\theta'$. In the case that each agent has only two possible types, $\succeq^*$ is identical to the standard order on $\mathbb{R}^n$. When the agents have more than two possible types, $\succeq^*$ is coarser than the standard order in that $\theta' \succeq^* \theta$ implies $\theta' \geq \theta$, but not the other way around.

**Definition 3.** *We say $g$ is a **coordination mechanism** if for any $\theta, \theta' \in \Theta$ and $a, a' \in A$ such that $g(\theta)[a] > 0$ and $g(\theta')[a'] > 0$, we have:*

(1) *$\theta' \succ^* \theta$ implies $X(\theta', a') \geq X(\theta, a)$ (across-state coordination),*

(2) *$\theta' = \theta$ implies $X(\theta', a') \geq X(\theta, a)$ or $X(\theta', a') \leq X(\theta, a)$ (within-state coordination).*

Across states, coordination groups the correct responses to the agents' types to states higher in the $\succeq^*$-ranking, which leads to more production at those states. We can equivalently restate part 1 of Definition 3 in terms of correct and incorrect responses as follows. Suppose $\theta' \succ^* \theta$ and $\theta'_i = \theta_i$. If a coordination mechanism ever correctly responds to agent $i$'s type at $\theta$, then it must correctly respond to agent $i$'s type with probability one at $\theta'$. Conversely, if a coordination mechanism ever incorrectly responds to agent $i$'s type at $\theta'$, then it must incorrectly respond to agent $i$'s type with probability one at $\theta$.

Within a single state, coordination correlates the correct responses and incorrect responses together in such a way to create a most productive action, a second most productive action, and so forth. This is equivalent to saying that any two actions that are played with positive probability can be ordered in terms of their vectors of agent production. For a deterministic mechanism, within-state coordination is trivial because only one action is used at each state.

**Theorem 2.** *There always exists a coordination mechanism that is optimal.*

*Proof.* See section 4.1. □

Theorem 2 asserts the optimality of coordination. The underlying logic is the same as we discussed in the two-type symmetric case. The complementarities reward grouping high-production outcomes together. Both the state and the action influence each agent's production, and coordination across states uses both of these dimensions by grouping the correct responses to the agents' types to states higher in the $\succeq^*$-ranking. Incorrect responses, which are required for truthful revelation, are placed at states lower in the $\succeq^*$-ranking where they do the least damage. Within a single state, coordination leverages the complementarities by correlating the correct responses into a most productive action, and then a second most productive action, and so forth.

Notably, Theorem 2 only ensures that coordination occurs in *an* optimal mechanism and not for *every* optimal mechanism. Optimal non-coordination mechanisms arise because the principal is indifferent about the agent's type when taking action 0 on that agent. If we perturbed $X_i(\theta_i, 0)$ to be strictly increasing in $\theta_i$, then every optimal mechanism would be a coordination mechanism.

The proof of Theorem 2 is detailed in Section 4.1 below. Despite being discussed first, Theorem 1 is proved as a corollary to Theorem 2 using the fact that $\succeq^*$ *effectively* provides a total order of the states in the two-type symmetric environment. To see why, note that any state with $m$ high types will be lower, according to $\succeq^*$, than at least one state with $m+1$ high types. And the symmetry of the setting allows us to restrict attention to mechanisms that treat all states with the same number of high types symmetrically. Therefore, we can treat any state with $m$ high types as effectively lower, according to $\succeq^*$, than any state with $m+1$ high types.

## 4.1. Proof of Theorem 2

We first establish that there always exists an optimal mechanism that obeys part 1 of Definition 3 (across-state coordination). For any $\theta \in \Theta$, let $h(\theta)$ equal one plus the number of high-type agents at $\theta$. For any $\varepsilon \geq 0$, define a perturbed version of the principal's payoff $\tilde{V}_\varepsilon : \Theta \times A \to \mathbb{R}$ as

$$\tilde{V}_\varepsilon(\theta, a) := (h(\theta))^\varepsilon W(X((\theta, a))).$$

The $\varepsilon$-perturbed optimal design problem can be written as

$$\max_{g\in\mathcal{G}}\sum\mu\left(\theta\right)g\left(\theta\right)\left[a\right]\tilde{V}_{\varepsilon}\left(\theta,a\right),$$

such that for every $i$ and any $\theta_i,\theta_i'\in\Theta_i$, $p_i^g\left(\theta_i\right)=p_i^g\left(\theta_i'\right)$.

The constraint set of this perturbed maximization problem does not depend on $\varepsilon$, and the objective function is jointly continuous in $\varepsilon$ and $g$. Hence, the theorem of the maximum applies, and the optimal solution is upper hemicontinuous in $\varepsilon$. Therefore, it suffices to show that every optimal mechanism coordinates across states for any $\varepsilon>0$, because that would imply the existence of an optimal mechanism that coordinates across states when $\varepsilon=0$. And at $\varepsilon=0$ we have $\tilde{V}_{\varepsilon}=V$, which restores the original design problem with which we are concerned.

Now fix any $\varepsilon>0$, and let $g$ be an optimal mechanism of the perturbed problem. (The existence of an optimal mechanism follows from standard arguments.) Assume by way of contradiction that $g$ does not coordinate across states. We will construct a modified mechanism $\hat{g}$ that is both incentive-compatible and gives a strictly higher payoff than $g$, which will contradict the optimality of $g$. Since $g$ is assumed to fail across-state coordination, we know there exist two states $\theta,\theta'$ and two actions $a,a'$ such that $\theta'\succ^*\theta$, $g\left(\theta\right)\left[a\right]>0$, $g\left(\theta'\right)\left[a'\right]>0$ and $X\left(\theta,a\right)\nleq X\left(\theta',a'\right)$.

Let $I\subseteq\{1,...,n\}$ be such that $i\in I$ if and only if $\theta_i=\theta_i'$. Choose $\hat{a}\in A$ so that for every $i\notin I$ we have $\hat{a}_i=a_i$, and for every $i\in I$ we have $X\left(\theta_i,\hat{a}_i\right)=X\left(\theta_i,a_i\right)\wedge X\left(\theta_i',a_i'\right)$. Choose action $\hat{a}'$ so that for every $i\notin I$ we have $\hat{a}_i'=a_i'$, and for every $i\in I$ we have $X\left(\theta_i',\hat{a}_i'\right)=X\left(\theta_i,a_i\right)\vee X\left(\theta_i',a_i'\right)$. For all $i\notin I$ , $\theta'\succ^*\theta$ implies that $\theta_i'$ is a high type and $\theta_i$ is a low type, which means we know $X_i(\theta_i',a_i')\geq X_i(\theta_i,a_i)$ regardless of $a_i$ and $a_i'$. Therefore we know that

$$X\left(\theta,a\right)\vee X\left(\theta',a'\right)=X\left(\theta',\hat{a}'\right)\text{ and }X\left(\theta,a\right)\wedge X\left(\theta',a'\right)=X\left(\theta,\hat{a}\right).\tag{1}$$

Equation 1, along with the fact that $X\left(\theta,a\right)\nleq X\left(\theta',a'\right)$ implies $X\left(\theta',\hat{a}'\right)\geq X\left(\theta',a'\right)$ and $X\left(\theta',\hat{a}'\right)\neq X\left(\theta',a'\right)$. Therefore, by the strict monotonicity of $W$, we have

$$W\left(X\left(\theta',\hat{a}'\right)\right)>W\left(X\left(\theta',a'\right)\right).\tag{2}$$

Set $\eta > 0$ to be arbitrarily small and define a modified mechanism $\hat{g}$ by setting, for any $\theta^* \in \Theta$ and $a^* \in A$,

$$\hat{g}\left(\theta^*\right)\left[a^*\right] = \left\{ \begin{array}{l|l} g\left(\theta^*\right)\left[a^*\right] - \frac{\eta}{\mu(\theta^*)} & \text{if } \left(\theta^*, a^*\right) = \left(\theta, a\right) \text{ or } \left(\theta^*, a^*\right) = \left(\theta', a'\right) \\ g\left(\theta^*\right)\left[a^*\right] + \frac{\eta}{\mu(\theta^*)} & \text{if } \left(\theta^*, a^*\right) = \left(\theta, \hat{a}\right) \text{ or } \left(\theta^*, a^*\right) = \left(\theta', \hat{a}'\right) \\ g\left(\theta^*\right)\left[a^*\right] & \text{otherwise.} \end{array} \right.$$

Mechanism $\hat{g}$ differs from $g$ by replacing $a$ with $\hat{a}$ at $\theta$ and replacing $a'$ with $\hat{a}'$ at $\theta'$. In other words, $\hat{g}$ groups the correct responses from $a, a'$ at the higher state $\theta'$ and groups the incorrect responses at the lower state $\theta$.

Let $\tilde{V}_\varepsilon\left(\hat{g}\right)$ and $\tilde{V}_\varepsilon\left(g\right)$ be the principal's payoff from mechanisms $\hat{g}$ and $g$, respectively, in the perturbed setting. We then have that

$$\begin{aligned} \tilde{V}_\varepsilon\left(\hat{g}\right) - \tilde{V}_\varepsilon\left(g\right) &= \mu\left(\theta\right)\frac{\eta}{\mu\left(\theta\right)}\left(\tilde{V}_\varepsilon\left(\theta, \hat{a}\right) - \tilde{V}_\varepsilon\left(\theta, a\right)\right) + \mu\left(\theta'\right)\frac{\eta}{\mu\left(\theta'\right)}\left(\tilde{V}_\varepsilon\left(\theta', \hat{a}'\right) - \tilde{V}_\varepsilon\left(\theta', a'\right)\right) \\ &> \eta\left(h\left(\theta\right)\right)^\varepsilon\left\{W\left(X\left(\theta, \hat{a}\right)\right) - W\left(X\left(\theta, a\right)\right) + W\left(X\left(\theta', \hat{a}'\right)\right) - W\left(X\left(\theta', a'\right)\right)\right\} \\ &\geq 0. \end{aligned}$$

The strict inequality follows from the fact that $h\left(\theta'\right) > h\left(\theta\right)$ and Equation (2). The weak inequality follows from Equation (1) and the supermodularity of $W$. All that remains is to establish that $\hat{g}$ is incentive-compatible, for which, using the incentive compatibility of $g$, it suffices to prove for all agents $i$ and $\theta_i^* \in \Theta_i$ that $p_i^{\hat{g}}\left(\theta_i^*\right) = p_i^g\left(\theta_i^*\right)$.

For any $i \notin I$, the equality is immediate since, by construction, $\hat{g}_i\left(\cdot\right) = g_i\left(\cdot\right)$ for all such agents. For $i \in I$, $\hat{g}_i\left(\cdot\right)$ differs from $g_i\left(\cdot\right)$ at $\theta$ and $\theta'$ if and only if $a_i'$ was an incorrect response to $\theta_i'$ and $a_i$ was a correct response to $\theta_i$. So take that case and suppose $\theta_i$ is a high type; the case where $\theta_i$ is a low type works similarly. Recall $i \in I$ implies $\theta_i = \theta_i'$, and therefore we know that $a_i' = \hat{a}_i = 0$ and $a_i = \hat{a}_i' = 1$. Hence,

$$p_i^{\hat{g}}\left(\theta_i\right) - p_i^g\left(\theta_i\right) = \frac{\varepsilon}{\mu\left(\theta'\right)}\mu_{-i}\left(\theta_{-i}'\right) - \frac{\varepsilon}{\mu\left(\theta\right)}\mu_{-i}\left(\theta_{-i}\right) = 0.$$

This calculation uses the fact that $\theta_i = \theta_i'$. Hence, $\hat{g}$ is incentive-compatible. Hence we have established that there always exists an optimal mechanism that obeys across-state coordination.

The proof of Theorem 2 can be finished by applying the following lemma.

**Lemma 2.** *Every optimal mechanism obeys within-state coordination*

*Proof.* See Appendix. □

We defer the proof of Lemma 2 to the appendix, but the intuition is straight-forward and works as follows. Incentive compatibility depends only on the values of $g_i(\theta)$ for each $i$ and $\theta$. Therefore, any optimal mechanism $g$ must achieve the highest payoff for the principal among the class of mechanisms that exactly match $g$ on the values for $g_i(\theta)$. And that highest payoff is achieved by making maximum use of the supermodularity of $W$, by, at each state, grouping the correct responses together into a most productive action, then a second most productive action and so forth. And this is precisely what within-state coordination does. □

## 5. Strategic Favoritism and Meritocracy

In a variety of environments, the optimal mechanism will display **strategic favoritism** by taking action 1 more often on a less productive agent than on a more productive agent. By "more productive" we mean an unambiguously better distribution of type quality, while assuming the agents are symmetric in all ways other than their type distribution. Hence, favoritism involves rewarding an agent who is worse for the principal's own aims. In practice, favoritism is often viewed as inefficient and attributed to biases of the principal such as nepotism; a key insight of our model is that favoritism can arise from purely strategic considerations and can be optimal even for an unbiased principal. While favoritism occurs in a robust set of environments, it is not ubiquitous. The more natural outcome of **meritocracy**, where more productive agents receive action 1 more often, also occurs in a variety of environments.

The first result of this section, Theorem 3, establishes that, if the complementarities in the principal's payoff function exceed a fixed threshold, then strategic favoritism occurs in every optimal mechanism for an open set of prior beliefs. Moreover, for high-enough complementarities, the probability the less productive agent receives action 1 can be as much as fifty percentage points higher than the more productive agent. The second result of this section, Theorem 4, establishes that there always exists an open set of prior beliefs where meritocracy occurs in every optimal mechanism. Moreover, for any prior belief, when the complementarities are low enough, every optimal mechanism displays meritocracy and not favoritism. Taken together, these two theorems demonstrate that strategic favoritism and meritocracy are both robust phenomena, and the first is associated with high complementarities while the second is associated with low complementarities.

Throughout this section, we restrict attention to the case that agents are symmetric in all ways other than their type distribution. Specifically, we focus on **symmetric principal payoff functions** $W$, where $W(Y) = W(Y')$ whenever $Y$ is a permutation of $Y'$. Additionally, we will assume that the agents share the same production functions $X_i(\cdot,\cdot)$ and type spaces $\Theta_i$.

Given that agents only differ in their distribution of types, we can use stochastic dominance in those distributions as an unambiguous measure of when one agent is more productive than another. Specifically, for any $\mu_i \in \Delta\Theta_i$ and $\mu_j \in \Delta\Theta_j$, we say $\mu_i$ **strictly first-order stochastically dominates** $\mu_j$ (and write $\mu_i >_{FOSD} \mu_j$) if

$$\sum_{\theta_j \in \Theta_j | \theta_j \leq c} \mu_j(\theta_j) > \sum_{\theta_i \in \Theta_i | \theta_i \leq c} \mu_i(\theta_i),$$

for all $c \in \mathbb{R}$ such that the right-hand side is in $(0,1)$. This definition slightly strengthens the standard notion of first-order stochastic dominance; this strengthening is only needed to show that low enough complementarities always leads to meritocracy (part 2 of Theorem 4). For all other results, the standard definition would work as well.

To formally define our notions of strategic favoritism and meritocracy, we will find it use useful to set

$$P_i^g := \sum_{\theta \in \Theta} g_i(\theta)\, \mu(\theta),$$

which is the overall probability that mechanism $g$ takes action 1 on agent $i$ at prior $\mu$.

**Definition 4.** *We say a mechanism $g$ exhibits **strategic favoritism** at prior $\mu$ if there exist $i, j$ such that $\mu_i >_{FOSD} \mu_j$ but $P_i^g < P_j^g$. We say the strategic favoritism is of magnitude $\lambda$ if $P_i^g + \lambda < P_j^g$.*

We similarly define meritocracy as follows.

**Definition 5.** *We say a mechanism $g$ exhibits* ***meritocracy*** *at prior $\mu$ if $\mu_i >_{FOSD} \mu_j$ implies $P_i^g \geq P_j^g$, for all $i, j$. If, in addition, for some $i, j$ we have $\mu_i >_{FOSD} \mu_j$ and $P_i^g > P_j^g + \lambda$, then we say the meritocracy is of magnitude $\lambda$.*

Note that favoritism and meritocracy are mutually exclusive and every mechanism exhibits exactly one of them.

Whether strategic favoritism or meritocracy occurs will depend on the degree of complementarities found in the principal's payoff function, which we formally define as follows.

**Definition 6.** *For any $M \geq 1$ and $S \subseteq \Theta$, we say the principal's payoff function has* ***complementarities above degree M at S*** *if*

$$W\left(Y \vee Y'\right) - W\left(Y\right) \geq M\left(W\left(Y'\right) - W\left(Y \wedge Y'\right)\right), \tag{3}$$

*for any $Y$, $Y'$ such that $Y \not\leq Y'$, $Y \not\geq Y'$ and*

$$Y = X(\theta, a) \text{ and } Y' = X(\theta', a'),$$

*for some $\theta, \theta' \in S$ and $a, a' \in A$.*

For any $M \geq 1$, we will also define the principal's payoff function to have **complementarities below degree M at S** in the exact same way, except reversing the inequality in Equation (3).

When $M = 1$, Equation (3) holding for any $Y, Y'$ is exactly the definition of supermodularity, and therefore the principal's payoff function always has complementarities above degree 1. The degree of complementarities of the principal's payoff function is invariant to affine transformations on $W$, which is necessary for any result to work because affine transformations do not change the set of optimal mechanisms.[5]

We are now ready to present the first main result of this section.

**Theorem 3.** *Fix $(\Theta, \{X_i\}_{i=1}^n)$ such that $\Theta_i = \Theta_j$ and $X_i(\cdot, \cdot) = X_j(\cdot, \cdot)$ for all $i, j$. Let $\theta \in \Theta$ such that, for some $i, j$, $\theta_i$ is a high type and $\theta_j$ is a low type. Then:*

[5] Not all potential definitions would have this property. For example, the cross derivative of $W$ is not invariant to affine transformations.

(1) *For any symmetric principal payoff function $W$ with complementarities above degree 2 at $\{\theta\}$, there exists an open set of priors at which every optimal mechanism exhibits strategic favoritism.*

(2) *Choose any $\lambda \in \left(0, \frac{1}{2}\right)$. There exists an $M > 1$ such that, for any symmetric principal payoff function $W$ with complementarities above degree $M$ at $\{\theta\}$, there exists an open set of priors at which every optimal mechanism exhibits strategic favoritism of magnitude $\lambda$.*

*Proof.* See Appendix. □

Part 1 of Theorem 3 establishes that, in a variety of environments, **every** optimal mechanism exhibits strategic favoritism. Part 2 of Theorem 3 establishes that, when the complementarities are sufficiently high, the strategic favoritism can be quite significant and have magnitude up to one half.

While strategic favoritism is robust, it is not ubiquitous, and its limits can be understood through studying when meritocracy occurs.

**Theorem 4.** *Fix $(\Theta, \{X_i\}_{i=1}^n)$ such that $\Theta_i = \Theta_j$ and $X_i(\cdot,\cdot) = X_j(\cdot,\cdot)$ for all $i, j$. Then:*

(1) *For any $\lambda \in (0,1)$ and any symmetric principal payoff function $W$, there exists an open set of priors at which every optimal mechanism exhibits meritocracy of magnitude $\lambda$.*

(2) *For any prior $\mu$, there exists an $M > 1$ such that for any symmetric principal payoff function $W$ that has complementarities below degree $M$ at $\Theta$, every optimal mechanism exhibits meritocracy at $\mu$.*

*Proof.* See Appendix. □

Theorem 4 demonstrates that meritocracy also occurs in a variety of environments and always occurs when complementarities are sufficiently low. Since strategic favoritism and meritocracy are mutually exclusive, this provides a partial converse to Theorem 3 by demonstrating that strategic favoritism requires high-enough complementarities. Taken together, the previous two theorems demonstrate that strategic favoritism and meritocracy both occur in a variety of environments, and the first is associated with high complementarities while the second is associated with low complementarities.

To gain an intuition as to how strategic favoritism can be optimal, consider the case in which each agent has only two possible types: a high type ($H$) and a low type ($L$). For each agent $i$ and type $\theta_i \in \{H, L\}$, define $P_i^g(\theta_i) := p_i^g(\theta_i)\mu_i(\theta_i)$ to be the **joint** probability of agent $i$ being type $\theta_i$ and mechanism $g$ setting $a_i = 1$. In contrast, $p_i^g(\theta_i)$ gives the probability of $a_i = 1$ **conditional** on agent $i$ being type $\theta_i$. The joint probability allows for across-state comparisons in changes to $P_i^g$ (the overall probability of $a_i = 1$); for any $\theta_i$, an $\varepsilon$ change to $P_i^g(\theta_i)$ changes $P_i^g$ by $\varepsilon$.

We can now rewrite the incentive constraint in terms of joint probabilities as

$$P_i^g(L) = \frac{\mu_i(L)}{1 - \mu_i(L)} P_i^g(H), \text{ for all } i. \tag{4}$$

By inspecting Equation (4), we can see that raising the probability of action 1 on an $H$ type of agent $i$ requires raising the probability of action 1 on an $L$ type of agent $i$. Moreover, the required amount of additional action 1 on an $L$ type increases in $\mu_i(L)$. In other words, agents with worse type distributions receive more action 1 on $L$ types for the sake of balancing the incentive constraints. This creates a channel that advantages agents with worse type distributions, and can therefore lead to strategic favoritism.

In response to this channel, the optimal mechanism could simply take action 1 less overall, that is on both types, on agents with higher values for $\mu_i(L)$. However, the logic of coordination pushes back against this response. Consider the state in which every agent is type $H$, then the complementarities reward the principal for taking action 1 on all the agents together, including the agents with high values for $\mu_i(L)$. And when the complementarities are strong enough, this reward overcomes the fact that taking action 1 on an $H$-type agent with high $\mu_i(L)$ requires a lot of action 1 on an $L$ type elsewhere. In this way, strategic favoritism arises from how the logic of coordination interacts with the incentives of the agents.

To gain insight into the proof of Theorem 3, we further specialize to the case with two agents who each have only two possible types. Therefore, the state space can be written as $\Theta = \{L, H\} \times \{L, H\}$ ,where $H$ is a high type for both agents and $L$ is a low type for both agents.

In this setting, we will demonstrate how a complete form of coordination can lead to either strategic favoritism or meritocracy. By a "complete form of coordination", we mean a mechanism that always either correctly responds to the types of both agents or incorrectly responds to the types of both agents. We will also require that a complete coordination mechanism correctly (incorrectly) responds with probability one to the types of both agents when both agents are type $H$ ($L$). Let $g$ be a complete coordination mechanism, then we know that

$$g_i(HH) = 1,\ g_i\left(LL\right) = 1, \text{ for } i \in \{1,2\},$$
$$g_1\left(HL\right) + g_2\left(HL\right) = 1 \text{ and } g_1\left(LH\right) + g_2\left(LH\right) = 1.$$

Substituting the above restrictions into the incentive constraint of agents 1 and 2 yields the following two equations:

$$\begin{aligned} g_1\left(HL\right) &= 1 - \frac{\mu_2(H)}{\mu_2\left(L\right)} g_2\left(LH\right) \\ g_2\left(LH\right) &= 1 - \frac{\mu_1\left(H\right)}{\mu_1\left(L\right)} g_1\left(HL\right). \end{aligned}$$

This is a system of two equations with two unknowns, which we can solve for $g_1\left(HL\right)$ and $g_2\left(LH\right)$. Thus we have $g_i(\theta)$ for all $\theta$, which allows us to compute

$$\begin{aligned} P_1^g - P_2^g &= \sum_{\theta\in\Theta} \mu(\theta)\left(g_1(\theta) - g_2(\theta)\right) \\ &= \left(\mu\left(LH\right) - \mu\left(HL\right)\right)\left(\frac{\mu\left(HH\right) + \mu\left(LL\right)}{\mu\left(HH\right) - \mu\left(LL\right)}\right). \end{aligned}$$

From this equation we can see that, when $\mu_1\left(H\right) > \mu_2\left(H\right) > \frac{1}{2}$, the value of $P_1^g - P_2^g$ will be negative, and therefore any complete coordination mechanism exhibits strategic favoritism. Conversely, if $\frac{1}{2} > \mu_1\left(H\right) > \mu_2\left(H\right)$, then $P_1^g - P_2^g$ will be positive and so any complete coordination mechanism exhibits meritocracy.

The first step of the detailed proof of Theorem 3 establishes strategic favoritism in the case with two types and $n = 2$ using a similar argument to the above. We use the choice of the open set of priors and the fact that complementarities are above degree 2 to show that the optimal mechanism belongs to a small handful of cases, which includes the complete coordination. We then show that each of those cases exhibits strategic favoritism, much as we did above for the complete-coordination case. To extend this result to the general case with any number of types and agents, we focus on priors in which all but two agents are a low type with probability close to one, and the remaining two agents are one of two types with a probability very close to one. These priors are effectively very close to the two-agent two-type case, which allows us to finish the proof of Theorem 3 by establishing that the set of optimal mechanisms moves upper hemicontinuously and applying a continuity argument.

# 6. Extensions

We now consider two extensions to our baseline model. In section 6.1, we examine the robustness of our results when relaxing the assumption that the agents draw their types independently. In section 6.2, we show how the principal can construct a virtually optimal and coalition-proof mechanism.

## 6.1. Robustness of Coordination Mechanisms

In this section we discuss the robustness of our main results to a small amount of correlation in the agents' types. For the purposes of this section, we limit ourselves to the case where each agent has only two possible types: a high type ($H$) and a low type ($L$). We establish continuity results when **positive dependence** among the agents' types is introduced. Positive dependence requires that, when agent $i$ is a high (low) type, the added correlation makes higher (lower) type profiles more likely for the other $n-1$ agents. Under a coordination mechanism, an agent reporting a high (low) type is more likely to receive action 1 when the other agents have a higher (lower) type profile. Hence, positive dependence reinforces each agent's incentive to tell the truth in a coordination mechanism, which is what allows our continuity results to work.

A notable feature of our baseline setting is that in any incentive-compatible mechanism all incentive constraints are binding due to the fact that each agent has the same incentives regardless of his type. Relaxing the independence assumption leads to different types of the same agent having different incentives due to differing beliefs about the type distributions of the other agents. Therefore, once we move away from independence, incentive compatibility no longer leads to everywhere-binding incentive constraints. By showing robustness to positive dependence, we establish that our results do not depend in a knife-edge way on the large degree of indifference found in the baseline setting. We provide an additional result, which goes beyond this and shows that the optimal mechanism with independent types is strictly incentive-compatible and almost optimal on an open set of non-independent priors.

Throughout this section, we suppose each agent has only two types: one high type ($H$) and one low type ($L$), so that the state space is $\Theta = \{L, H\}^n$. The general space of prior beliefs on the state space is given by $\Upsilon = \Delta\Theta$. We will let $\Upsilon_I \subset \Upsilon$ be the space of priors for which types are independent across agents. For any $\mu \in \Upsilon$ and $\theta_{-i} \in \Theta_{-i}$, we will write $\mu(\theta_{-i})$ and $\mu(\theta_{-i}|t)$ to be the overall probability of $\theta_{-i}$ and the conditional probability of $\theta_{-i}$ given that agent $i$ has type $t$. For any $\mu \in \Upsilon$, let $I^\mu \in \Upsilon_I$ be the unique independent prior that puts the same marginal probability as $\mu$ on each individual type of each agent.

**Definition 7.** *We say $\mu \in \Upsilon$ has **positive dependence among the agents' types** if for every agent $i$ and lower set $S \subseteq \Theta_{-i}$*

$$\sum_{\theta_{-i}\in S} \mu(\theta_{-i}|H) \leq \sum_{\theta_{-i}\in S} I^\mu(\theta_{-i}) \leq \sum_{\theta_{-i}\in S} \mu(\theta_{-i}|L).$$

*Let $\Upsilon_p$ be the space of all priors with positive dependence among the agents' types.*

The above definition is equivalent to requiring that $\mu(\theta_{-i}|H)$ first-order stochastically dominates $I^\mu(\theta_{-i})$, which in turn first-order stochastically dominates $\mu(\theta_{-i}|L)$. This uses a multivariate notion of first-order stochastic dominance, which we draw from Shaked and Shanthikumar (2007).[6] Therefore, positive dependence requires that adding correlation makes higher $\theta_{-i}$ more likely when agent $i$ is a high type and less likely when agent $i$ is a low type.

**Theorem 5.** *Fix any $\mu \in \Upsilon_I$ and any sequence $\{\mu^m\}_{m=1}^\infty \subset \Upsilon_p$ such that $\mu^m \to \mu$. Then there exists a coordination mechanism $g$ and a sequence of mechanism $\{g^m\}_{m=1}^\infty$, such that $g$ is optimal at $\mu$, $g^m$ is optimal at $\mu^m$ and $g^m \to g$.*

*Proof.* See Appendix. □

While not implied by the statement of Theorem 5, a similar proof could show the optimal mechanism moves upper hemicontinuously at any independent prior in the space $\Upsilon_P$. However upper hemicontinuity by itself would not guarantee that the optimal mechanisms with small amounts of positive dependence would be close to a coordination mechanism. This gap occurs because Theorem 2 only established coordination in an optimal mechanism and not for every optimal mechanism. Theorem 5 explicitly addresses this gap by showing that,

[6] They call this property the "usual stochastic order", and it is defined on page 266.

with small a amount of positive dependence, there exists an optimal mechanism close to a coordination mechanism.

We now move to the second main result of this section, which establishes that a small amount of positive dependence has the potential to provide strict incentives. For this result, we will need to exclude trivial mechanisms. Recall that $P_i^g$ is the overall probability that $g$ sets $a_i = 1$.

**Definition 8.** *A mechanism $g$ is trivial if there exists an $i$ such that either $P_i^g = 1$ or $P_i^g = 0$.*

A mechanism is trivial if there exists an agent who either always or never receives action 1. A trivial mechanism could never provide strict incentives.

**Theorem 6.** *Take any $\mu \in \Upsilon_I$ and $\varepsilon > 0$; let $g$ be any optimal coordination mechanism at $\mu$. If $g$ is nontrivial, then there exists an open set of priors $\Upsilon_O$, with $\mu$ on the boundary of $\Upsilon_O$, such that for any $\nu \in \Upsilon_O$, $g$ is strictly incentive-compatible and furthermore*

$$\max_{g' \in \mathcal{G}} V\left(g'|\nu\right) < V\left(g|\nu\right) + \varepsilon.$$

*Proof.* See Appendix. □

Theorem 6 shows that a coordination mechanism optimal at independent prior $\mu$ becomes strictly incentive-compatible and almost optimal for some open set of nearby priors $\Upsilon_O$. Strict incentives makes the mechanism robustly incentive-compatible even if we allow agents to make small mistakes in their reports. The fact that $\mu$ is not in the set $\Upsilon_O$ follows from the fact that incentive-compatible mechanisms can never provide strict incentives at an independent prior. The best we could hope for is that $\mu$ is on the boundary of $\Upsilon_O$, which is what Theorem 6 shows.

## 6.2. Coalition-Proof Mechanisms

One concern with the optimal direct mechanism is that the agents may be able to collude to improve their outcomes.[7] For example, in the coordination mechanism discussed in section 2, the agents could always get action 1 if they both report a high type. Surprisingly, it turns out that we can amend the optimal direct mechanism to make it immune to collusion, i.e., coalition-proof, with virtually the same payoff to the principal.

[7] This is a concern in related literature, e.g., Jackson and Sonnenschein (2007).

The coalition-proof mechanism we propose is inspired by techniques found in the virtual implementation literature (e.g., Abreu and Matsushima (1992)), where with arbitrarily high probability, the optimal direct mechanism is implemented. However, given our assumption that the agents' preferences are independent of their type, our environment fails to satisfy Abreu-Matsushima measurability,[8], and the results in that work cannot be directly applied to our setting. Interestingly, we are able to add ideas from the classic implementation literature, notably integer games (Maskin (1999)), to get around this problem. Our results also differ from the aforementioned implementation literatures since we focus on immunity from coalition deviations instead of equilibrium uniqueness.

Our notion of collusion is a coalition of agents performing a joint deviation that strictly benefits everyone in the coalition. We require that the coalition not be vulnerable to internal defections; in other words, coalition members must be best-responding to the deviation the coalition is performing. Members outside the coalition are unaware of the deviation and report truthfully.

Our coalition-proof mechanism will not be a direct mechanism, but instead requires agents to report their own type, a guess for each other agent's type, and an integer. Therefore, a strategy of agent $i$ is a function $\sigma_i : \Theta_i \to \Delta(\Theta \times \mathbb{Z})$, where $\sigma_i(\theta_i)[t, z_i]$ gives the probability that agent $i$ of type $\theta_i$ will report profile $t \in \Theta$ and integer $z_i \in \mathbb{Z}$. We will say an agent uses a truthful reporting strategy if he reports his own type truthfully. Truthful reporting makes no restriction on $z_i$ or the guess made about the other agents' types.

**Definition 9.** *For a fixed mechanism, a* ***valid coalition*** *is a pair $(I, \{\sigma_i\}_{i=1}^n)$ that consists of a non-empty subset of the agents $I \subseteq \{1, ..., n\}$ and a strategy for each agent i $\sigma_i$ such that the following three conditions hold:*

1. *For every $i \notin I$, $\sigma_i$ is a truthful reporting strategy.*
2. *For all $i \in I$, $\sigma_i$ is a best response for agent i to $\{\sigma_i\}_{i=1}^n$.*
3. *For all $i \in I$, agent i receives a strictly higher payoff than if all agents reported truthfully.*

We say a mechanism is **coalition-proof** if there does not exist any valid coalitions.

[8] Since each type of every player assesses all Anscombe-Aumann acts in the same manner, the limit partition of the set of types for each player is the entire set of types (i.e., in the notation of Abreu and Matsushima (1992) $\Psi_i^0 = \Theta_i = \Psi_i^*$). This implies that, in our setting, only constant social-choice functions are Abreu-Matsushima measurable, and so the designer can simply choose all players with some exogenous probability.

For each $\varepsilon > 0$, we will define a coalition-proof mechanism that we denote $g^\varepsilon$. As $\varepsilon$ goes to zero, $g^\varepsilon$ will converge to the optimal direction mechanism in both payoffs and outcomes. Therefore, we can generate a coalition-proof mechanism arbitrarily close to optimality. We will only provide an informal description of $g^\varepsilon$, which will suffice to convey the key technique. A formal description of $g^\varepsilon$ can be found in the appendix.

With probability $1 - \varepsilon$, $g^\varepsilon$ implements the optimal direct mechanism, which only relies on each agent's report about his own type. With probability $\varepsilon$, $g^\varepsilon$ plays a betting game that rewards agents for correctly guessing the other agents' reports.[9] However, only the bet of the agent who reports the $z_i$ will count; all other agents will receive no reward. Agents can opt out of the betting by reporting $z_i = 0$, in which case they receive a fixed reward. The bets will be calibrated so that the agents break even on every possible bet when everyone is reporting truthfully. When collusion occurs, the agents in the deviating coalition can make a strict gain from betting, and hence all agents will desire to report the $z_i$, which destroys any possible coalition equilibrium.

**Theorem 7.** *For any $\varepsilon \in (0, 1)$, mechanism $g^\varepsilon$ is coalition proof.*

*Proof.* See Appendix. □

As $\varepsilon$ goes to zero, mechanism $g^\varepsilon$ implements the optimal direct mechanism with probability approaching 1, and, because the principal's payoff is bounded, the payoff of $g^\varepsilon$ converges to the optimal payoff. And Theorem 7 establishes that $g^\varepsilon$ is immune to coalitions for any $\varepsilon \in (0, 1)$. Hence, we have shown how the principal can achieve a virtually optimal payoff in a coalition-proof mechanism.

## 7. Concluding Remarks

In this paper, we studied a principal facing agents with type-independent preferences. Type-independent preferences, while natural in a number of economic environments, pose a challenge to the typical screening technique that tailors the principal's action on each agent to the preference associated with that agent's reported type. Type-independent preferences makes this tailoring impossible. When faced with a single agent, the challenge of type

[9] Since we have no transfers, the rewards the betting uses simply entails taking action 1 on the agent with some probability. Higher rewards correspond to higher probability.

independent preferences is insurmountable and the principal gains nothing from the ability to commit to a mechanism.

In contrast, a principal who faces multiple agents can leverage complementarities in her own payoff function by coordinating her actions. Coordination groups the correct responses to the types of the agents at states with more high types. Conversely incorrect responses to the types of the agents are grouped at states with more low types. Under complementarities, the principal's payoff has greater potential when there are more high type agents. Hence, coordination allows the principal to accept the failure necessary for incentive compatibility when the consequence is small in order to ensure success when the potential gain is high. In the two-type symmetric environment, coordination takes the form of correctly responding to every agent when the number of high types exceeds a fixed threshold and incorrectly responding to every agent when the number of high types falls below a second fixed threshold.

We also showed how the optimal mechanism can sometimes display strategic favoritism by rewarding an agent who is inferior for the principal's own aims. Behavior such as favoritism is sometimes attributed to biases such as nepotism, but we show it can arise from purely strategic considerations and be optimal for an unbiased principal.

The logic of coordination applies beyond the complementarities assumption: whenever the principal's utility function is not additively separable, a similar analysis could have been done. We limited ourselves to the complementarities case to focus our discussion. But generalizing our results to a more general non-separable environment presents an interesting direction for future work.

Our focus on type-independent preferences served to highlight our key results by making standard screening techniques impossible. However, the logic behind coordination would still operate even if the preferences of agents did depend on their type. Determining how coordination would combine with standard screening techniques presents another interesting direction for future study. Allowing transfers or repeated interactions represent two natural ways to approach this question.

## References

Abreu, D. and H. Matsushima (1992). Virtual Implementation in Iteratively Undominated Strategies: Complete Information. *Econometrica 60*(5), 993–1008.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2014). Optimal Allocation with Costly Verification. *American Economic Review 104*(12), 3779–3813.

Chassang, S. and G. Padro i Miquel (2014). Corruption, Intimidation, and Whistle-blowing: a Theory of Inference from Unverifiable Reports. *Working Paper*.

Cremer, J. and R. P. Mclean (1988). Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions. *Econometrica 56*(6), 1247–1257.

Fang, H. and P. Norman (2006). To Bundle or Not to Bundle. *The RAND Journal of Economics 37*(4), 946–963.

Guo, Y. and J. Hörner (2015). Dynamic Mechanisms without Money. *Working Paper*.

Jackson, M. O. and H. F. Sonnenschein (2007). Overcoming Incentive Constraints by Linking Decisions. *Econometrica 75*(1), 241–257.

Li, J., N. Matouschek, and M. Powell (2015). The Burden of Past Promises. *Working Paper*.

Lipnowski, E. and J. Ramos (2015). Repeated Delegation. *Working Paper*, 1–50.

Maskin, E. (1999). Nash Equilibrium and Welfare Optimality. *Review of Economic Studies 66*(1), 23–38.

Rahman, D. (2012). But Who Will Monitor the Monitor. *American Economic Review 102*(6), 2767–2797.

Rubinstein, A. and M. E. Yaari (1983). Repeated Insurance Contracts and Moral Hazard. *Journal of Economic Theory 30*(1), 74–97.

Shaked, M. and J. G. Shanthikumar (2007). *Stochastic Orders.* New York: Springer Science and Business Media.

# A. Proofs From Section 4 (For Online Publication)

## A.1. Proof of Lemma 2

Let $g^*$ be an optimal mechanism, and suppose for contradiction that it does not obey within-state coordination. Then there exists $\theta \in \Theta$ and $a, a' \in A$ such that $g^*(\theta)[a] > 0, g^*(\theta)[a'] > 0$ and $X(\theta, a) \ngeq X(\theta, a')$ and $X(\theta, a) \nleq X(\theta, a')$. Now construct $\hat{a}, \hat{a}' \in A$ as follows

$$\hat{a}_i = \left\{ \begin{array}{c|l} a_i & \text{if } X_i(\theta_i, a_i) \geq X_i(\theta_i, a'_i) \\ a'_i & \text{otherwise} \end{array} \right.$$

and

$$\hat{a}'_i = \left\{ \begin{array}{c|l} a_i & \text{if } X_i(\theta_i, a_i) < X_i(\theta_i, a'_i) \\ a'_i & \text{otherwise} \end{array} \right. .$$

By construction we have

$$X(\theta, \hat{a}) = X(\theta, a) \vee X(\theta, a') \text{ and } X(\theta, \hat{a}') = X(\theta, a) \wedge X(\theta, a').$$

Pick $\varepsilon > 0$ to be smaller than both $g^*(\theta)[a]$ and $g^*(\theta)[a']$. Construct a mechanism $g$ defined as

$$g\left(\tilde{\theta}\right)[\tilde{a}] = \left\{ \begin{array}{l|l} g^*\left(\tilde{\theta}\right)[\tilde{a}] - \varepsilon & \text{if } \tilde{\theta} = \theta \text{ and } \tilde{a} \in \{a, a'\} \\ g^*\left(\tilde{\theta}\right)[\tilde{a}] + \varepsilon & \text{if } \tilde{\theta} = \theta \text{ and } \tilde{a} \in \{\hat{a}, \hat{a}'\} \\ g^*\left(\tilde{\theta}\right)[\tilde{a}] & \text{otherwise} \end{array} \right. .$$

It is easy to check that, for all $\tilde{\theta} \in \Theta$ and for each $i$, $g_i(\tilde{\theta}) = g_i^*(\tilde{\theta})$, which implies $p_i^g(\tilde{\theta}) = p_i^{g^*}(\tilde{\theta})$. Therefore by Lemma 1, $g$ is incentive compatible.

Let $V(g)$ and $V(g^*)$ be the payoff to the principal from mechanism $g$ and $g^*$ respectively. Then:

$$V(g^*) - V(g) \quad = \quad \varepsilon\left(W(X(\theta, a)) + W(X(\theta, a')) - W(X(\theta, \hat{a})) - W(X(\theta, \hat{a}'))\right) < 0$$

The strict inequality follows from the strict supermodularity of $W$. And this contradicts the optimality of $g^*$, and we have established that $g^*$ obeys within-state coordination, as desired.

□

## A.2. Proof of Theorem 1

We start by formally defining what it means for an a mechanism to be anonymous. For any bijection $\sigma : \{1,...,n\} \to \{1,...,n\}$ we use $\sigma(\theta) \in \Theta$ and $\sigma(a) \in A$ to mean:

$$\begin{aligned} \sigma(\theta) &= \left(\theta_{\sigma(1)}, \theta_{\sigma(2)}, ..., \theta_{\sigma(n)}\right) \\ \sigma(a) &= \left(a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)}\right). \end{aligned}$$

**Definition 10.** *We say a mechanism $g$ is anonymous if for any bijection $\sigma : \{1,...,n\} \to \{1,...,n\}$, and for all $\theta \in \Theta$, $a \in A$:*

$$g(\theta)[a] = g(\sigma(\theta))[\sigma(a)].$$

We will first show there always exists an anonymous coordination mechanism that is optimal. To prove this claim perform an $\varepsilon$ perturbation of the principal's payoff function precisely as was done in the proof of Theorem 2. Importantly, this perturbation maintains the symmetry of the principal's payoff function. Let $g$ be an optimal mechanism in this perturbed setting. Let $\Sigma$ be the set of all bijections from $\{1,...,n\} \to \{1,...,n\}$. For each $\sigma \in \Sigma$, define mechanism $g^\sigma$ by setting for each $\theta, a$:

$$g^\sigma(\theta)[a] = g(\sigma(\theta))[\sigma(a)].$$

By the symmetry of the setting, $g^\sigma$ is optimal for every $\sigma \in \Sigma$ because it has the same incentive compatibility properties and principal payoff as $g$. Now define $g^*_\varepsilon$ as

$$g^*_\varepsilon(\theta)[a] = \frac{1}{|\Sigma|} \sum_{\sigma \in \Sigma} g^\sigma(\theta)[a].$$

The mechanism $g^*_\varepsilon$ is a mixture of optimal mechanisms and is therefore optimal itself. And by construction $g^*_\varepsilon$ is also anonymous. Using the same proof as in Theorem 2, we can show every optimal mechanisms in the $\varepsilon$-perturbed setting is a coordination mechanism, including $g^*_\varepsilon$. We can then take $\varepsilon$ to zero while constructing a sequence of optimal and anonymous mechanisms. Since the space of mechanisms is compact, this sequence must have a convergent sub-sequence, and the same upper hemicontinuity argument used in the proof of Theorem 2 will imply that the limit of this sequence is optimal in the original setting. And that limit mechanism will be both anonymous and a coordination mechanism.

We now know there exists an anonymous coordination mechanism that is optimal, which we call $g$. For the remainder of this proof involves showing that $g$ has all the properties stated in Theorem 1.

Part 2 of Definition 3 (within-state coordination) combined with anonymity requires that if $\theta_i = \theta_j$ and $a_i \neq a_j$, then $g(\theta)[a] = 0$. Hence, at any state $\theta$, the mechanism $g$ has only four possible actions: (1) incorrectly respond to every agent's type ($\overrightarrow{\mathbf{1}}_{\theta_i=L}$), (2) correctly respond to every agent's type ($\overrightarrow{\mathbf{1}}_{\theta_i=H}$), (3) take action 1 on everyone ($\overrightarrow{\mathbf{1}}$) and (4) take action 0 on everyone ($\overrightarrow{\mathbf{0}}$).

Define $h'(\theta)$ to be the number of high types at $\theta$, (i.e. $h'(\theta) = h(\theta) - 1$, where $h(\theta)$ was defined in the proof of Theorem 2). Using what we established in the previous paragraph along with the definition of an anonymous coordination mechanism, it is straight-forward to show the following four facts:

Fact 1: If $h'(\theta) \geq 1$ and $g(\theta)\left[\overrightarrow{\mathbf{1}}_{\theta_i=H}\right] > 0$, then $g(\theta')\left[\overrightarrow{\mathbf{1}}_{\theta'_i=H}\right] = 1$ whenever $h'(\theta') > h'(\theta)$.
Fact 2: If $h'(\theta) \leq n-1$ and $g(\theta)\left[\overrightarrow{\mathbf{1}}_{\theta_i=L}\right] > 0$, then $g(\theta')\left[\overrightarrow{\mathbf{1}}_{\theta'_i=L}\right] = 1$ whenever $h'(\theta') < h'(\theta)$.
Fact 3: If $h'(\theta) \leq n-1$ and $g(\theta)[\overrightarrow{\mathbf{0}}] > 0$, then $g(\theta')[\overrightarrow{\mathbf{1}}] = 0$ whenever $h'(\theta') \leq n-1$.
Fact 4: If $h'(\theta) \geq 1$ and $g(\theta)[\overrightarrow{\mathbf{1}}] > 0$, then $g(\theta')[\overrightarrow{\mathbf{0}}] = 0$ whenever $h'(\theta') \geq 1$.

Let $m^H$ be the smallest number weakly greater than 1 such that there exists $\theta \in \Theta$ with $h'(\theta) = m^H$ and $g(\theta)[\overrightarrow{\mathbf{1}}_{\theta_i=H}] > 0$. If such a number does not exist, then set $m^H = n$. Let $m^L$ be the largest number weakly less than $n-1$ such that there exists $\theta \in \Theta$ with $h(\theta) = m^L$ and where $g(\theta)[\overrightarrow{\mathbf{1}}_{\theta_i=L}] > 0$. If such a number does not exist, then set $m^L = 0$.

Using facts 1-2 from above, $m^H$ and $m^L$ satisfy parts (1) and (2) of Theorem 1. Additionally, by how $m^L$, $m^H$ were chosen, at any $\theta \in \Theta$ with $m^L < h'(\theta) < m^H$, $g$ can only use actions $\overrightarrow{\mathbf{0}}$ and $\overrightarrow{\mathbf{1}}$. And for any $\theta$ with $m^L < h'(\theta) < m^H$, we know that $1 \leq h'(\theta) \leq n-1$. Therefore, by facts 3-4, there exists $a^* \in \{\overrightarrow{\mathbf{1}}, \overrightarrow{\mathbf{0}}\}$, such that $g(\theta)[a^*] = 1$ whenever $m^L < h'(\theta) < m^H$.

This finishes the proof of Theorem 1, but we now consider what happens at states with exactly $m^L$ or $m^H$ high types.

First take the case that $m^H > m^L$. At states where $h'(\theta) = m^H$, we know that $\overrightarrow{\mathbf{1}}_{\theta'_i=L}$ cannot be used since that would contradict how $m^L$ was chosen. And combining this with facts 3-4 from above, we get that only $a^*$ and $\overrightarrow{\mathbf{1}}_{\theta'_i=H}$ can be used. (Note that if $h'(\theta) = n$, then fact 3 above does not apply, but then $\overrightarrow{\mathbf{1}}_{\theta'_i=H} = \overrightarrow{\mathbf{1}}$ anyway). And by analogous logic, at states with $h'(\theta) = m^L$, only actions $a^*$ and $\overrightarrow{\mathbf{1}}_{\theta'_i=L}$ can be used.

Now take the case that $m^H = m^L$. Actions $\overrightarrow{\mathbf{1}}$ and $\overrightarrow{\mathbf{0}}$ can never be used at the same state since that would violate within-state coordination. Therefore, we can choose $a^* \in \{\overrightarrow{\mathbf{1}}, \overrightarrow{\mathbf{0}}\}$, such that only $a^*$, $\overrightarrow{\mathbf{1}}_{\theta'_i=H}$ and $\overrightarrow{\mathbf{1}}_{\theta'_i=L}$ are used at states where $h'(\theta) = m^H = m^L$. And we are free to define $a^*$ in this way since, in the $m^H = m^L$ case, $a^*$ is never used anywhere else.

## A.3. Complete Coordination in the Two-Type Symmetric Environment

As promised in the text of Section 4, we now characterize the complete coordination mechanisms within the Two-Type Symmetric Environment, where $m^H = m^L$ and only the "correctly respond to every agent" and "incorrectly respond to every agent" actions are used. Any complete coordination mechanisms can be fully described by two values, as given in the following definition.

**Definition 11.** *For any $m \in \{0, 1, ..., n\}$ and $q \in [0, 1)$, the **complete coordination mechanism** $g^{m,q}$*

1. *Correctly responds to all agents at states that have strictly more than $m$ high types.*
2. *Incorrectly responds to all agents at states that have strictly less than $m$ high types.*
3. *Correctly and incorrectly responds to all agents with with probability $1 - q$ and $q$ respectively, at states with exactly $m$ high types.*

From Lemma 1, we know $g^{m,q}$ is incentive-compatible if and only if each agent has the same probability of receiving action 1 regardless of what report he makes. This constraint can be written as

$$\sum_{k=m}^{n-1} \mathcal{H}(k) + (1-q)\mathcal{H}(m-1) = \sum_{k=0}^{m-1} \mathcal{H}(k) + q\mathcal{H}(m), ^{10} \tag{5}$$

where for any $k \in 0, ..., n$, we define

$$\mathcal{H}(k) := \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

Fixing any agent, $\mathcal{H}(k)$ is the probability that exactly $k$ of the $n-1$ other agents are type $H$.

[10] As a convention, we set $\sum_{k=n}^{n-1} \mathcal{H}(k)$ and $\sum_{k=0}^{-1} \mathcal{H}(k)$ as equal to zero.

The left-hand side and right-hand side of Equation (5) give the probability an agent receives action 1 conditional on reporting $H$ and $L$, respectively. The two terms on the left-hand side correspond to the agent receiving action 1 with certainty if at least $m$ other agents report $H$ and with probability $1-q$ if exactly $m-1$ other agents report $H$. The right-hand side works analogously.

One can show that, for any fixed values of $p$ and $n$, there exists a unique pair $(m,q)$ that satisfies Equation (5). In other words, for a fixed two-type-symmetric environment, there is exactly one complete coordination mechanism that is incentive-compatible. The following proposition characterizes when that complete coordination mechanism is optimal. We let the indicator functions $\overrightarrow{\mathbf{1}}_{\theta_i=H}$ and $\overrightarrow{\mathbf{1}}_{\theta_i=L}$, respectively, denote correctly responding and incorrectly responding to all agents at state $\theta$.

**Proposition 1.** *Assume a two-type symmetric environment, and let $(m^*,q^*)$ be the unique solution to Equation (5). Define*

$$\lambda := \frac{(n-m^*)\,p}{m^*\,(1-p)+(n-m^*)\,p}.$$

*Then $g^{m^*,q^*}$ is optimal if and only if at any $\theta\in\Theta$ with exactly $m^*$ high types,*

$$(1-\lambda)\,V\left(\theta,\overrightarrow{\mathbf{1}}_{\theta_i=H}\right)+\lambda V\left(\theta,\overrightarrow{\mathbf{1}}_{\theta_i=L}\right)\geq V\left(\theta,\overrightarrow{\mathbf{1}}\right)$$

*and*

$$\lambda V\left(\theta,\overrightarrow{\mathbf{1}}_{\theta_i=H}\right)+(1-\lambda)\,V\left(\theta,\overrightarrow{\mathbf{1}}_{\theta_i=L}\right)\geq V\left(\theta,\overrightarrow{\mathbf{0}}\right).$$

The full proof of this proposition appeared in an earlier working-paper version of this work and is proved as a corollary to Theorem 1.

## B. Proofs from Section 5 (For Online Publication)

Throughout this section we relax our assumption that $\mu_i$ is full support for each agent $i$. For the results in this section, this relaxation is without loss of generality. For Theorem 3 and part 1 of Theorem 4 this is because any open set of priors contains an open subset in which every prior is full support. And if part 2 of Theorem 4 holds for all priors, it clearly holds for all full-support priors.

Given a state space $\Theta$ and prior $\mu$, we will use $supp(\mu_i)\subseteq\Theta_i$ to denote the support of $\mu_i$.

## B.1. Proof of Theorem 3

Choose $(\Theta, \{X_i\}_{i=1}^n)$ such that $\Theta_i = \Theta_j$ and $X_i = X_j$ for all $i, j$. Let $\hat{\theta}$ be any state with at least one high type and one low type. Without loss of generality, we will assume $\hat{\theta}_1$ is a low type and $\hat{\theta}_2$ is a high type. Set $L = \hat{\theta}_1$ and $H = \hat{\theta}_2$

Fix any symmetric principal payoff function with complementarities above degree 2 at $\{\hat{\theta}\}$. We will show there always exists an open set of priors at which every optimal mechanism exhibits strategic favoritism, which will prove the first part of Theorem 3. To do so we first state a useful proposition.

**Proposition 2.** *Let $n = 2$ and for any $L, H \in \mathbb{R}$ with $L < H$, let $\Theta = \{L, H\} \times \{L, H\}$ and let $X_i(\cdot, \cdot) = X_j(\cdot, \cdot)$ for all $i, j$. Suppose the principal's payoff function is symmetric and has complementarities above degree 2 at state $(H, L)$. Then there exists $\mu \in \Delta\{L, H\} \times \Delta\{L, H\}$ with $1 > \mu_1(H) > \mu_2(H) > 0$ such that for every optimal mechanism $g$, $P_2^g > P_1^g$.*

*Proof.* See Section B.4. □

To make use of the above proposition, suppose for contradiction that there exists no open set of priors where strategic favoritism occurs in every optimal mechanism. By assumption, we know that $L, H \in \Theta_i$ for all $i$. Define $U$ to be the set containing all priors $\mu$ that obey the following three conditions:

- $i > 2 \Rightarrow \mu_i(L) = 1$
- $i \leq 2 \Rightarrow \mu_i(H) + \mu_i(L) = 1$
- $1 > \mu_1(H) > \mu_2(H) > 0$

Fix an arbitrary $\hat{\mu} \in U$. Since we can think of each prior as a finite dimensional vector, we can endow $\prod \Delta\Theta_i$ with the Euclidian metric. For any $\varepsilon > 0$, let $B_\varepsilon(\hat{\mu})$ be the epsilon ball around $\hat{\mu}$ in the space $\prod \Delta\Theta_i$. Let $D \subseteq \prod \Delta\Theta_i$ be the set of all priors such that $\mu_1 >_{FOSD} \mu_2$. For each natural number $m$, define $E_m(\hat{\mu})$ as

$$E_m(\hat{\mu}) = B_{\frac{1}{m}}(\hat{\mu}) \cap D.$$

For each $m$, $E_m(\hat{\mu})$ contains an open subset, and hence by our contradiction assumption there must exist $\mu^m \in E_m(\hat{\mu})$ such that there exists a mechanism $g^m$ that is optimal at $\mu^m$ and $P_1^{g^m} \geq P_2^{g^m}$. Since $\mathcal{G}$ is compact, we know the sequence $\{g^m\}_{m=1}^{\infty}$ has a convergent subsequence with limit which we denote $g$. And $g$ will have the property that $P_1^g \geq P_2^g$.

Additionally, since $\mu^m \in B_{\frac{1}{m}}(\hat{\mu})$ for each $m$, we know that $\mu^m \to \hat{\mu}$. We show in section B.3 below that the set of optimal mechanisms is upper hemicontinuous in the domain $\prod \Delta\Theta_i$. Hence, $g$ must be optimal at $\hat{\mu}$. So we have shown that for every $\hat{\mu} \in U$ there exists at least one optimal mechanism with the property that $P_1^g \geq P_2^g$.

Since every $\hat{\mu} \in U$ assigns probability one to every agent $i > 2$ being a low type, any optimal mechanism will always take action 0 on every agent $i > 2$. Therefore, we can effectively reduce the design problem to just agents 1 and 2. And since agents 1 and 2 can only be type $H$ or type $L$ at $\hat{\mu}$, we can effectively reduce the type space to just include those two types. By doing so, we have reduced our setting to the two-type two-agent environment described in the statement of Proposition 2. And the fact that there exists at least one optimal mechanism with the property that $P_1^g \geq P_2^g$ for every $\hat{\mu} \in U$, will therefore contradict Proposition 2. And hence we have derived our contradiction, and proved part one of Theorem 3.

To prove the second part of Theorem 3 we can employ an analogous argument as we just used along with the following proposition.

**Proposition 3.** *Let $n = 2$ and for some $L, H \in \mathbb{R}$ with $L < H$, let $\Theta = \{L, H\} \times \{L, H\}$ and let $X_i(\cdot,\cdot) = X_j(\cdot,\cdot)$ Then for any $\lambda \in \left(0, \frac{1}{2}\right)$ there exists a natural number $M$ such that if the principal's payoff function is symmetric and has complementarities above degree $M$ at $(H, L)$, then there exists $\mu \in \Delta\{L, H\} \times \Delta\{L, H\}$ with $1 > \mu_1(H) > \mu_2(H) > 0$ such that in every optimal mechanism $g$ has the property that $P_2^g \geq P_1^g + \lambda$.*

*Proof.* See Section B.5. □

## B.2. Proof of Theorem 4

To prove this theorem we make use of the following two propositions, the proofs of which can be found in sections B.6 and B.7, respectively.

**Proposition 4.** *Fix $(\Theta, \{X_i\}_{i=1}^n)$ and any principal payoff function. Choose an agent $i$ and let $t^L \in \Theta_i^L$. Then there exists a $M_L \in (0, 1)$ such that for any prior with $\mu\left(t^L\right) > M_L$, in every optimal mechanism action 1 is taken on agent $i$ with probability 0. Moreover for any agent $i$ and $t^H \in \Theta_i^H$ there exists $M_H \in (0, 1)$ such that for any prior with $\mu\left(t^H\right) > M_H$, in every optimal mechanism action 1 is taken on agent $i$ with probability 1.*

**Proposition 5.** *Choose* $(\Theta, \{X_i\}_{i=1}^n)$ *with* $\Theta_i = \Theta_j$ *and* $X_i(\cdot,\cdot) = X_j(\cdot,\cdot)$ *for all* $i, j$ *and any prior* $\mu$. *Let* $g$ *be an incentive compatible mechanism that does not exhibit meritocracy. Then there exists* $M > 1$ *such that, given any symmetric principal payoff function with complementarities below degree* $M$ *at* $\Theta$, $g$ *is not optimal.*

Choose any $(\Theta, \{X_i\}_{i=1}^b)$ with $\Theta_i = \Theta_j$ and $X_i(\cdot,\cdot) = X_j(\cdot,\cdot)$ for all $i, j$. Fix any symmetric principal payoff function. Let $t^H$,$t^L$ be the highest and lowest possible types in each $\Theta_i$. By Proposition 4, for each agent $i$ there exists a $M_H^i \in (0,1)$ such that $\mu_i\left(t^H\right) > M_H^i$ implies every optimal mechanism will take action 1 on agent $i$ with probability 1. Similarly there exists a $M_L^i \in (0,1)$ such that $\mu_i\left(t^L\right) > M_L^i$ implies every optimal mechanism will take action 1 on agent $i$ with probability 0. Let $M_H = \max_i M_H^i$ and $M_L = \max_i M_L^i$. Now it is clearly possible to construct an open set of priors such that each prior $\mu$ in the set has $\mu_1\left(t^H\right) > M_H$ and $\mu_i\left(t^L\right) > M_L$ for each $i \geq 2$, as well as $\mu_1 >_{FOSD} \mu_2 >_{FOSD} ... >_{FOSD} \mu_n$. Therefore, every optimal mechanism $g$ at $\mu$ has $P_1^g = 1$ and $P_i^g = 0$ for all $i > 1$ . Hence $g$ displays meritocracy of magnitude $\lambda$ for any $\lambda \in (0,1)$, as desired.

To prove the second part of Theorem 4, fix any prior $\mu$. Any optimal mechanism solves a finite dimensional linear program whose constraint set is entirely determined by $\mu$ and $\Theta$, and does not depend on the principal's payoff function. Therefore, we can identify a finite number of "extreme" mechanisms $g_1, ..., g_m$, such that, for any principal payoff function, any optimal mechanism is a convex combination of optimal extreme mechanisms. Re-order the extreme mechanisms so that $g_1, ...g_k$ is the subset of these extreme mechanism that do not display meritocracy. (Note that whether a mechanism exhibits meritocracy does not depend on the principal's payoff function). For each $i = 1, ..., k$ let $M_i > 1$ be the value specified in Proposition 5. Let $M = \min\{M_1, ..., M_k\}$. Then for any symmetric principal payoff function with complementarities below $M$ we know that $g_i$ is not optimal for all $i \leq k$. Hence any optimal mechanism is a convex combination of two "extreme" mechanisms that display meritocracy. And meritocracy is preserved through convex combinations, and hence all optimal mechanisms exhibit meritocracy. And this proves the second part of Theorem 4.

### B.3. Proof that the optimal mechanism is upper hemicontinuous in the domain of independent priors.

Fix $\Theta, V$ and $n$. Let $\Gamma : \prod_{i=1}^{n} \Delta\Theta_i \rightrightarrows \mathcal{G}$ be the correspondence where $\Gamma(\mu)$ gives the set of optimal mechanisms at prior $\mu$. Standard arguments will show $\Gamma$ is compact-valued and non-empty valued. We want to show in addition that $\Gamma$ is upper hemicontinuous. By the theorem of the maximum it will suffice to show that the set of incentive compatible mechanisms moves continuously in the domain $\prod_{i=1}^{n} \Delta\Theta_i$.

The set of equalities that characterize incentive compatibility move continuously with $\mu$. And the set of solutions to a continuously changing set of equalities always moves upper hemicontinuously as long as the set of solutions is always non-empty. And the set of incentive compatible mechanisms is non-empty since always taking action 1 on all the agents is always incentive compatible. Hence the set of incentive compatible mechanisms moves upper hemicontinuously.

Now we only need to establish lower hemicontinuity of the set of incentive compatible mechanisms. Let $\mu^m \in \prod_{i=1}^{n} \Delta\Theta_i$ with $\mu^m \to \mu$. Then for every mechanism $g$ that is incentive compatible at $\mu$, we want to show there exists a sequence of mechanisms $g^m$ such that $g^m \to g$ and $g^m$ is incentive compatible at $\mu^m$ for every $m$.

Define $p_i^g(\theta_i, \mu)$ using the same expression as we used to define $p_i^g(\theta_i)$, except we now include $\mu$ as an explicit argument instead of a fixed parameter. For $i = 0, ...n$, we will define $\{g^{m,i}\}_{m=1}^{\infty}$ so that for each $i$, $g^{m,i}$ is incentive compatible at $\mu^m$ for agents $1, .., i$ and $g^{m,i} \to g$ as $m \to \infty$. Hence $g^{m,n}$ will produce the desired sequence. Without loss of generality, we will assume there is no agent $i$ for which $g$ always takes action 1 on. If any such an agent existed, we could simply set our sequences $\{g^{m,i}\}_{m=1}^{\infty}$ to always take action 1 on that agent and reformulate the problem on the remaining $n-1$ agents.

We proceed by induction on $i$. As a base case set $g^{m,0} = g$ for all $m$. Now assume $\{g^{m,i-1}\}_{m=1}^{\infty}$ has been appropriately defined and we will show how to define $\{g^{m,i}\}_{m=1}^{\infty}$. For each $t \in supp(\mu_i)$, choose $\theta^t \in \Theta, a^t \in A$ such that $a_i^t = 0, \theta_i^t = t$ and $\mu(\theta^t) > 0, g(\theta^t)[a^t] > 0$. If no such $\theta^t, a^t$ exists that must mean that $p_i^g(t, \mu) = 1$ , and since $g$ is incentive compatible at $\mu$ that must mean that $p_i^g(t', \mu) = 1$ for all $t' \in supp(\mu_i)$, implying $g$ always takes action 1 on agent $i$, which we ruled out earlier. Hence, an appropriate $\theta^t, a^t$ must always exist.

For any natural number $m$, define $\varepsilon^m : supp(\mu_i) \to \mathbb{R}$ as:

$$\varepsilon^m(t) := \max_{t' \in \Theta_i} p_i^{g^{m,i-1}}(t', \mu^m) - p_i^{g^{m,i-1}}(t, \mu^m).$$

By construction $\varepsilon^m(\cdot) \geq 0$ for all $m$. The value $\varepsilon^m$ measures how far $g^{m,i-1}$ is from being incentive compatible for agent $i$ at prior $\mu^m$. In the limit, $g^{m,i-1} \to g$ which is fully incentive compatible at $\mu$, hence $\varepsilon^m(t) \to 0$ as $m \to \infty$ for any $t$. Since $\mu(\theta^t) > 0$ and $g(\theta^t)[a^t] > 0$, we know $\mu_{-i}^m(\theta_{-i}^t)$ and $g^{m,i-1}(\theta^t)[a^t]$ converge to values strictly above zero as $m \to \infty$. Hence, we can choose $M$ large enough so that for all $m \geq M$ we have

$$\frac{\varepsilon^m(t)}{\mu_{-i}^m(\theta_{-i}^t)} < g^{m,i-1}(\theta^t)[a^t] \text{ for all } t \in \Theta_i.$$

For each $m \geq M$ define $g^{m,i}$ as follows:

$$g^{m,i}(\theta)[a] = \begin{cases} g^{m,i-1}(\theta)[a] - \frac{\varepsilon^m(\theta_i)}{\mu_{-i}^m(\theta_{-i})} & \text{if } (\theta, a) = (\theta^t, a^t) \text{ for some } t \in supp(\mu_i) \\ g^{m,i-1}(\theta)[a] + \frac{\varepsilon^m(\theta_i)}{\mu_{-i}^m(\theta_{-i})} & \text{if } (\theta, a) = \left(\theta^t, \left(1, a_{-i}^t\right)\right) \text{ for some } t \in supp(\mu_i) \\ g^{m,i-1}(\theta)[a] & \text{otherwise.} \end{cases}$$

And since $\varepsilon^m \to 0$, $g^{m,i}$and $g^{m,i-1}$ converge together as $m \to \infty$, hence $g^{m,i} \to g$ must hold. We next show that for all $m \geq M$, $g^{m,i}$ is incentive compatible at $\mu^m$ for each agent $j$ where $j \leq i$. Well, for $j < i$, mechanism $g^{m,i}$ has the same probability of taking $a_j$ at each state as $g^{m,i-1}$. Therefore $g^{m,i}$ is incentive compatible for agent $j$ since $g^{m,i-i}$ is incentive compatible for agent $j$. Now for agent $i$ we have that for any $t \in supp(\mu_i)$:

$$\begin{aligned} p_i^{g^{m,i}}(t, \mu^m) &= p_i^{g^{m,i-1}}(t, \mu^m) + \frac{\varepsilon^m(\theta_i^t)}{\mu_{-i}^m(\theta_{-i}^t)} \mu_{-i}^m(\theta_{-i}^t) \\ &= p_i^{g^{m,i-1}}(t, \mu^m) + \varepsilon^m(\theta_i^t) \\ &= \max_{t' \in \Theta_i} p_i^{g^{m-1}}(t', \mu^m). \end{aligned}$$

From which it follows that $p_i^{g^{m,i}}(t, \mu^m) = p_i^{g^{m,i}}(t', \mu^m)$ for all $t, t' \in supp(\mu_i)$ and hence $g^{m,i}$ is incentive compatible for agent $i$ at $\mu^m$ for all $m \geq M$. Therefore, we have constructed a sequence $g^{m,i}$ from $M$ to $\infty$ with the desired properties, and now we simply renumber the sequence to go from 1 to $\infty$, and we are done.□

## B.4. Proof of Proposition 2

We have two agents and a state space of

$$\Theta = \{L, H\} \times \{L, H\},$$

where $L < H$ and $L \in \Theta_i^L$ and $H \in \Theta_i^H$ for each $i$. For any prior $\mu$, we write $\mu_i$ to mean $\mu_i(H)$, and $1 - \mu_i$ for $\mu_i(L)$. Without loss of generality, we can set for each $i$:

$$X_i(\theta_i, a_i) = \begin{cases} 1 & \text{if } \theta_i = H \text{ and } a_i = 1 \\ 0 & \text{if } \theta_i = L \text{ and } a_i = 1 \\ \frac{1}{2} & \text{if } a_i = 0 \end{cases}.$$

For compactness, given production vector $X$ we will write $W(X_1 X_2)$ to mean $W((X_1, X_2))$. So for example if the mechanism takes action 0 on both agents, then the payoff to the principal is $W\left(\frac{1}{2}\frac{1}{2}\right)$. The assumption of complementarities above degree 2 at $(H, L)$ implies that

$$W\left(1\frac{1}{2}\right) - W\left(\frac{1}{2}\frac{1}{2}\right) \geq 2\left(W(10) - W\left(\frac{1}{2}0\right)\right),$$

which be rearranged to yield

$$W\left(1\frac{1}{2}\right) + 2W\left(\frac{1}{2}0\right) - W(\frac{1}{2}\frac{1}{2}) \geq 2W(10).$$

By the strict monotonicity of $W$, $W\left(\frac{1}{2}0\right) < W\left(\frac{1}{2}\frac{1}{2}\right)$, and therefore

$$W\left(1\frac{1}{2}\right) + W\left(\frac{1}{2}0\right) > 2W(10). \tag{6}$$

**Lemma 3.** *Fix a prior $\mu$ such that $\mu_1, \mu_2 \in \left(\frac{1}{2}, 1\right)$ and*

$$\frac{W(1\frac{1}{2}) - W(10)}{W\left(1\frac{1}{2}\right) - W\left(\frac{1}{2}\frac{1}{2}\right)} < \frac{\mu_i}{1 - \mu_i} \text{ for } i = 1, 2.$$

*Then in any optimal mechanism $g(HH)[11] = 1$.*

*Proof.* Using the supermodularity and symmetry of $W$, inequality assumed in the statement of the lemma implies that

$$\frac{W\left(1\frac{1}{2}\right) - W(10)}{W(11) - W\left(1\frac{1}{2}\right)} < \frac{\mu_i}{1 - \mu_i} \text{ for } i = 1, 2. \tag{7}$$

Let $g$ be an optimal mechanism. We will first establish that $g\left(HH\right)\left[10\right] = 0$ and $g\left(HH\right)\left[01\right] = 0$. Suppose for contradiction that $g\left(HH\right)\left[10\right] > 0$, which implies $P_2^g < 1$. Incentive compatibility will require there to exist $\hat{\theta}, \hat{a}$ with $\hat{\theta}_2 = L$, $\hat{a}_2 = 0$ and $g(\hat{\theta})\left[\hat{a}\right] > 0$. Define $\tilde{a} = (\hat{a}_1, 1)$. For an arbitrarily small $\varepsilon > 0$, construct the following alternate mechanism $\tilde{g}$.

$$\tilde{g}\left(\theta\right)\left[a\right] = \begin{cases} g\left(\theta\right)\left[a\right] + \frac{\varepsilon}{\mu_1} & \text{if } \left(\theta, a\right) = \left(\left(HH\right), \left(11\right)\right) \\ g\left(\theta\right)\left[a\right] - \frac{\varepsilon}{\mu_1} & \text{if } \left(\theta, a\right) = \left(\left(HH\right), \left(10\right)\right) \\ g\left(\theta\right)\left[a\right] + \frac{\varepsilon}{\mu_1(\theta_1)} & \text{if } \left(\theta, a\right) = \left(\hat{\theta}, \tilde{a}\right) \\ g\left(\theta\right)\left[a\right] - \frac{\varepsilon}{\mu_1(\theta_1)} & \text{if } \left(\theta, a\right) = \left(\hat{\theta}, \hat{a}\right) \\ g\left(\theta\right)\left[a\right] & \text{otherwise} \end{cases}$$

When agent 1 receives action 1 has not changed, hence $\tilde{g}$ is incentive compatible for agent 1. To check incentive compatibility for agent 2 notice that

$$\begin{aligned} p_2^{\tilde{g}}\left(H\right) &= p_2^g\left(H\right) + \mu_1 \frac{\varepsilon}{\mu_1} = p_2^g\left(H\right) + \varepsilon \\ p_2^{\tilde{g}}\left(L\right) &= p_2^g\left(L\right) + \mu_1\left(\hat{\theta}_1\right) \frac{\varepsilon}{\mu_1\left(\hat{\theta}_1\right)} = p_2^g\left(L\right) + \varepsilon. \end{aligned}$$

Since $g$ is incentive compatible, $p_2^g\left(H\right) = p_2^g\left(L\right)$ and hence $p_2^{\tilde{g}}\left(H\right) = p_2^{\tilde{g}}\left(L\right)$ implying that $\tilde{g}$ is incentive compatible.

The difference in the principal's payoff between $\tilde{g}$ and $g$ is given by:

$$\begin{aligned} V\left(\tilde{g}\right) - V\left(g\right) &= \frac{\varepsilon}{\mu_1}\mu\left(HH\right)\left(W\left(11\right) - W(1\frac{1}{2})\right) - \mu\left(\hat{\theta}\right)\frac{\varepsilon}{\mu_1\left(\hat{\theta}_1\right)}\left(V\left(\hat{\theta}, \hat{a}\right) - V\left(\hat{\theta}, \tilde{a}\right)\right) \\ &= \varepsilon\mu_2\left(W\left(11\right) - W(1\frac{1}{2})\right) - \left(1 - \mu_2\right)\varepsilon\left(W\left(X_1\left(\hat{\theta}_1, \hat{a}_1\right)\frac{1}{2}\right) - W\left(X_1\left(\hat{\theta}_1, \tilde{a}_1\right)0\right)\right) \\ &\geq \varepsilon\mu_2\left(W\left(11\right) - W(1\frac{1}{2})\right) - \left(1 - \mu_2\right)\varepsilon\left(W(1\frac{1}{2}) - W\left(10\right)\right) \\ &> 0. \end{aligned}$$

The first inequality holds using supermodularity of $W$ and the fact that $\hat{a}_1 = \tilde{a}_1$. The second inequality holds due to Equation (7). And we have violated the optimality of mechanism $g$, which provides our desired contradiction and establishes that $g\left(HH\right)\left[10\right] = 0$. A similar argument will show $g\left(HH\right)\left[01\right] = 0$.

Now suppose for contradiction that $g\left(HH\right)\left[11\right] < 1$. By what we just showed it must be that $g\left(HH\right)\left[00\right] > 0$. Therefore, incentive compatibility requires that for $i = 1, 2$ there exists $\theta^i, a^i$ with $\theta^i_i = L$, $a^i_i = 0$ and $g\left(\theta^i\right)\left[a^i\right] > 0$. Define $\tilde{a}^1 = (1, a^1_2)$ and $\tilde{a}^2 = (a^2_1, 1)$. For an arbitrarily small $\varepsilon > 0$ construct the following modified mechanism $\tilde{g}$

$$\tilde{g}\left(\theta\right)\left[a\right] = \left\{\begin{array}{l|l} g\left(\theta\right)\left[a\right] + \varepsilon & \text{if } \left(\theta, a\right) = \left(\left(HH\right), \left(11\right)\right) \\ g\left(\theta\right)\left[a\right] - \varepsilon & \text{if } \left(\theta, a\right) = \left(\left(HH\right), \left(00\right)\right) \\ g\left(\theta\right)\left[a\right] + \varepsilon\frac{\mu_2}{\mu_2\left(\theta^1_2\right)} & \text{if } \left(\theta, a\right) = \left(\theta^1, \tilde{a}^1\right) \\ g\left(\theta\right)\left[a\right] + \varepsilon\frac{\mu_1}{\mu_1\left(\theta^2_1\right)} & \text{if } \left(\theta, a\right) = \left(\theta^2, \tilde{a}^2\right) \\ g\left(\theta\right)\left[a\right] - \varepsilon\frac{\mu_2}{\mu_2\left(\theta^1_2\right)} & \text{if } \left(\theta, a\right) = \left(\theta^1, a^1\right) \\ g\left(\theta\right)\left[a\right] - \varepsilon\frac{\mu_1}{\mu_1\left(\theta^2_1\right)} & \text{if } \left(\theta, a\right) = \left(\theta^2, a^2\right) \\ g\left(\theta\right)\left[a\right] & \text{otherwise.} \end{array}\right.$$

To check that $\tilde{g}$ is incentive compatible for agent 1 notice that

$$\begin{aligned} p^{\tilde{g}}_1\left(H\right) &= p^g_1\left(H\right) + \varepsilon\mu_2 \\ p^{\tilde{g}}_1\left(L\right) &= p^g_1\left(L\right) + \varepsilon\frac{\mu_2}{\mu_2\left(\theta^1_2\right)}\mu_2\left(\theta^1_2\right) = p^g_1\left(L\right) + \varepsilon\mu_2 \end{aligned}$$

and hence $p^{\tilde{g}}_1\left(H\right) = p^{\tilde{g}}_1\left(L\right)$ since $g$ is incentive compatible. A similar argument can be made for agent 2.

The payoff difference to the principal between $g$ and $\tilde{g}$ is given by:

$$\begin{aligned} V\left(\tilde{g}\right) - V\left(g\right) &= \varepsilon\mu\left(HH\right)\left(W\left(11\right) - W(\frac{1}{2}\frac{1}{2})\right) + \varepsilon\frac{\mu_2}{\mu_2\left(\theta^1_2\right)}\mu\left(\theta^1\right)\left(V\left(\theta^1, \tilde{a}^1\right) - V\left(\theta^1, a^1\right)\right) \\ &\quad + \varepsilon\frac{\mu_1}{\mu_1\left(\theta^2_1\right)}\mu\left(\theta^2\right)\left(V\left(\theta^2, \tilde{a}^2\right) - V\left(\theta^2, a^2\right)\right) \\ &\geq \mu\left(HH\right)\varepsilon\left(W\left(11\right) - W(\frac{1}{2}\frac{1}{2})\right) - \varepsilon\left(\mu\left(HL\right) + \mu\left(LH\right)\right)\left(W(1\frac{1}{2}) - W\left(10\right)\right) \\ &\geq \mu\left(HH\right)\varepsilon\left(W(1\frac{1}{2}) - W(\frac{1}{2}\frac{1}{2}) - \frac{1-\mu_2}{\mu_2}\left(W(1\frac{1}{2}) - W\left(10\right)\right)\right) \\ &\quad + \mu\left(HH\right)\varepsilon\left(W(1\frac{1}{2}) - W(\frac{1}{2}\frac{1}{2}) - \frac{1-\mu_1}{\mu_1}\left(W(1\frac{1}{2}) - W\left(10\right)\right)\right) \\ &> 0 \end{aligned}$$

The first inequality uses the supermodularity and symmetry of $W$, while the third inequality follows from the inequality assumed in the statement of the lemma. Proving the second inequality makes use of the following relationship which follows from the supermodularity of

$W$:

$$\begin{aligned} W(11) - W(\frac{1}{2}\frac{1}{2}) &= W(11) - W(1\frac{1}{2}) + W(1\frac{1}{2}) - W(\frac{1}{2}\frac{1}{2}) \\ &\geq 2\left(W(1\frac{1}{2}) - W(\frac{1}{2}\frac{1}{2})\right). \end{aligned}$$

Hence we have

$$V(\tilde{g}) - V(g) > 0,$$

which contradicts the optimality of $g$, and we have established that $g(HH)[11] = 1$. □

**Lemma 4.** *Fix any prior $\mu$ such that $\mu_1, \mu_2 \in \left(\frac{1}{2}, 1\right)$. If $g$ is an optimal mechanism with $g(HH)[11] = 1$, then $g(LL)[11] = 1$.*

*Proof.* Let $g$ be any optimal mechanism with $g(HH)[11] = 1$. By the fact that $\mu_1, \mu_2 > \frac{1}{2}$, incentive compatibility will require $g_1(LH) > 0$ and $g_2(HL) > 0$, which means for $i = 1, 2$ there exists $a^i$ with $a^i_i = 1$ and $g(LH)[a^1] > 0, g(HL)[a^2] > 0$. Define $\tilde{a}^1 = (0, a^1_2)$ and $\tilde{a}^2 = (a^2_1, 0)$.

We will first show that $g(LL)[00] = 0$. Suppose not for contradiction. For arbitrary small $\varepsilon > 0$, construct alternate mechanism $\tilde{g}$ as follows

$$\tilde{g}(\theta)[a] = \left\{ \begin{array}{l|l} g(\theta)[a] + \varepsilon & \text{if } (\theta, a) = ((LL), (11)) \\ g(\theta)[a] + \varepsilon\frac{1-\mu_2}{\mu_2} & \text{if } (\theta, a) = ((LH), \tilde{a}^1) \\ g(\theta)[a] + \varepsilon\frac{1-\mu_1}{\mu_1} & \text{if } (\theta, a) = (HL, \tilde{a}^2) \\ g(\theta)[a] - \varepsilon & \text{if } (\theta, a) = ((LL), (00)) \\ g(\theta)[a] - \varepsilon\frac{1-\mu_2}{\mu_2} & \text{if } (\theta, a) = ((LH), a^1) \\ g(\theta)[a] - \varepsilon\frac{1-\mu_1}{\mu_1} & \text{if } (\theta, a) = (HL, a^2) \\ g(\theta[a]) & \text{otherwise} \end{array} \right.$$

To check incentive compatibility of $\tilde{g}$ notice that

$$\begin{aligned} p^{\tilde{g}}_1(H) &= p^g_1(H) \\ p^{\tilde{g}}_1(L) &= p^g_1(L) + \varepsilon(1-\mu_2) - \varepsilon\frac{1-\mu_2}{\mu_2}\mu_2 = p^g_1(L). \end{aligned}$$

A similar argument can be made of agent 2. Hence $\tilde{g}$ is incentive compatible because $g$ is incentive compatible. The change in payoff between mechanism $g$ and $\tilde{g}$ is given by:

$$\begin{aligned}
V(\tilde{g}) - V(g) &= \varepsilon\mu(LH)\frac{1-\mu_2}{\mu_2}\left(V\left(\theta,\tilde{a}^1\right) - V\left(\theta,a^1\right)\right) + \varepsilon\mu(HL)\frac{1-\mu_1}{\mu_1}\left(V\left(\theta,\tilde{a}^2\right) - V\left(\theta,a^2\right)\right) \\
&\quad + \varepsilon\mu(LL)\left(W(00) - W(\frac{1}{2}\frac{1}{2})\right) \\
&= \varepsilon\mu(LL)\left(\left(V\left(\theta,\tilde{a}^1\right) - V\left(\theta,a^1\right)\right) + V\left(\theta,\tilde{a}^2\right) - V\left(\theta,a^2\right) - \left(W(\frac{1}{2}\frac{1}{2}) - W(00)\right)\right) \\
&\geq \varepsilon\mu(LL)\left(2\left(W(\frac{1}{2}\frac{1}{2}) - W(\frac{1}{2}0)\right) - \left(W(\frac{1}{2}\frac{1}{2}) - W(00)\right)\right) \\
&> \varepsilon\mu(LL)\left(W(\frac{1}{2}\frac{1}{2}) - W(\frac{1}{2}0) + W(0\frac{1}{2}) - W(00) - \left(W(\frac{1}{2}\frac{1}{2}) - W(00)\right)\right) \\
&= 0
\end{aligned}$$

The first and second inequality follow from the strict supermodularity of $W$. And we have contradicted the optimality of $g$ and hence $g(LL)[00] = 0$.

We next show that $g(LL)[10] = 0$. Suppose $g(LL)[10] > 0$ for contradiction. For an arbitrarily small $\varepsilon > 0$, construct the following alternative mechanism $\tilde{g}$.

$$\tilde{g}(\theta)[a] = \left\{\begin{array}{l|l}
g(\theta)[a] + \varepsilon & \text{if } (\theta,a) = ((LL),(11)) \\
g(\theta)[a] + \varepsilon\frac{1-\mu_1}{\mu_1} & \text{if } (\theta,a) = ((HL),\tilde{a}^2) \\
g(\theta)[a] - \varepsilon & \text{if } (\theta,a) = ((LL),(10)) \\
g(\theta)[a] - \varepsilon\frac{1-\mu_1}{\mu_1} & \text{if } (\theta,a) = ((HL),a^2) \\
g(\theta)[a] & \text{otherwise}
\end{array}\right.$$

Incentive compatibility of $\tilde{g}$ for agent 1 is immediate because when $a_1 = 1$ is played has not changed. Incentive compatibility for agent 2 follows from the fact that

$$\begin{aligned}
p_2^{\tilde{g}}(H) &= p_2^g(H) \\
p_2^{\tilde{g}}(L) &= p_2^g(L) + \varepsilon(1-\mu_1) - \varepsilon\frac{1-\mu_1}{\mu_1}\mu_1 = p_2^g(L).
\end{aligned}$$

The payoff difference between $\tilde{g}$ and $g$ is given by:

$$\begin{aligned} V(\tilde{g}) - V(g) &= \varepsilon\mu(HL)\frac{1-\mu_1}{\mu_1}\left(V\left(\theta, \tilde{a}^2\right) - V\left(\theta, a^2\right)\right) + \varepsilon\mu(LL)\left(W(00) - W(0\frac{1}{2})\right) \\ &\geq \varepsilon\mu(LL)\left(W(\frac{1}{2}\frac{1}{2}) - W(\frac{1}{2}0) + W(00) - W\left(0\frac{1}{2}\right)\right) \\ &> \varepsilon\mu(LL)\left(W(0\frac{1}{2}) - W(00) + W(00) - W(0\frac{1}{2})\right) \\ &= 0 \end{aligned}$$

The two inequalities follow from the strict supermodularity and symmetry of $W$. And this contradicts the optimality of $g$, and hence we have shown that $g(LL)[10] = 0$. A similar proof will show that $g(LL)[01] = 0$. Hence it must be that $g(LL)[11] = 1$, which finishes the proof of this lemma. □

Now let $g$ be any optimal mechanism with the property that

$$g(HH)[11] = g(LL)[11] = 1.$$

By Lemma 2, $g$ must obey within-state coordination (part 2 of Definition 3). At states $HL$ and $LH$, within-state coordination implies that if 11 is used, then action 00 is not. Therefore, if $g_1(HL) \geq 1 - g_2(HL)$, then $g(HL)[00] = 0$, and if $g_1(HL) \leq 1 - g_2(HL)$ then $g(HL)[11] = 0$. Similarly, $g_1(LH) \geq 1 - g_2(LH)$ implies $g(LH)[00] = 0$, and $g_1(LH) \leq 1 - g_2(LH)$ implies $g(LH)[11] = 0$. These facts allows us to fully characterize the mechanism as a function of the following four values: $g_1(HL)$, $g_1(LH)$, $g_2(LH)$ and $g_2(HL)$. For example, if we are in the case that $g(HL)[00] = 0$, then we know $g(HL)[10] = 1 - g_2(HL)$ and $g(HL)[01] = 1 - g_1(HL)$. And $g(HL)[11]$ takes the remaining probability. And the remaining cases can be similarly analyzed.

Incentive compatibility requires that $p_i^g(\theta_i) = P_i^g$ for each agent $i$ and type $\theta_i$, which allows us to define $g_i(\theta_i)$ in terms of $P_1^g$ and $P_2^g$ as follows.

$$\begin{aligned} g_1(HL) &= \frac{P_1^g - \mu_2}{1-\mu_2} \text{ and } g_1(LH) = 1 - \frac{1-P_1^g}{\mu_2} \\ g_2(HL) &= 1 - \frac{1-P_2^g}{\mu_1} \text{ and } g_2(LH) = \frac{P_2^g - \mu_1}{1-\mu_1}. \end{aligned}$$

Since $g_i(\theta)$ must be between 0 and 1, we know that $P_1^g \in [\mu_2, 1]$ and $P_2^g \in [\mu_1, 1]$ must hold.

We have shown that the values of $g_i(\theta)$ fully characterize the mechanism, and that those values are in turn characterized by $P_i^g$. Therefore, we can rewrite the principal's maximization problem in terms of $P_1^g$ and $P_2^g$, constrained by $P_1^g \in [\mu_2, 1]$ and $P_2^g \in [\mu_1, 1]$. (The incentive constraints are implicitly accounted for by how we substituted them in to achieve the previous display equations). The objective function of this maximization problem is a piece-wise linear function[11], and therefore all optimal points have to either be extreme points of the feasible set, kinks in the objective function or convex combinations of such points. Hence, it will suffice to show that there is no optimal extreme or kink point with the property that $P_2^g \leq P_1^g$. The kinks in the linear optimization occur when $g_1(HL) = 1 - g_2(HL)$ or $g_1(LH) = 1 - g_2(LH)$, which, as we discussed above, is when the mechanism switches between using action 11 and action 00. Therefore we we can enumerate the seven feasible extreme or kink points as follows:

(1) $\frac{P_1^g-\mu_2}{1-\mu_2} = \frac{1-P_2^g}{\mu_1}, P_2^g = \mu_1$
(2) $P_1^g = 1, P_2^g = \mu_1$
(3) $P_1^g = 1, P_2^g = 1$
(4) $\frac{P_1^g-\mu_2}{1-\mu_2} = \frac{1-P_2^g}{\mu_1}, \frac{P_2^g-\mu_1}{1-\mu_1} = \frac{1-P_1^g}{\mu_2}$
(5) $P_1^g = \mu_2, P_2^g = \mu_1$
(6) $P_1^g = \mu_2, P_2^g = 1$
(7) $P_1^g = \mu_2, \frac{P_2^g-\mu_1}{1-\mu_1} = \frac{1-P_1^g}{\mu_2}$

Some potential points were excluded from this list for being redundant. For example the point

$$\frac{P_1^g - \mu_2}{1-\mu_2} = \frac{1-P_2^g}{\mu_1}, P_2^g = 1$$

is equivalent to point (6) above. Whenever $1 > \mu_1 > \mu_2 > \frac{1}{2}$, one can verify that only points 1, 2 and 3 have the property that $P_2 \leq P_1$. Define $U_i$ to be the sum of the principal's payoff on states $(HL)$ and $(LH)$ from point $i$ in the above list. (We exclude the payoff from states $(HH)$ and $(LL)$ since, under the assumption $g(HH)[11] = g(LL)[11] = 1$, payoff from those states are fixed.) If $\max\{U_1, U_2, U_3\} < \max\{U_4, U_5\}$, then any optimal extreme point or kink must have $P_1^g < P_2^g$. Using the sufficient conditions in Lemmas 3 and 4 to ensure

[11] The domain of the payoff function is $[\mu_1, 1] \times [\mu_2, 1]$. By piece-wise linear we mean there are a finite number of closed sets that cover this domain and such that the payoff function is linear within each of the closed sets.

that $g\left(HH\right)[11] = g\left(LL\right)[11] = 1$ in any optimal mechanism, we have proved the following result.

**Lemma 5.** *Suppose that* $1 > \mu_1 > \mu_2 > \frac{1}{2}$, $\max\left\{U_1, U_2, U_3\right\} < \max\left\{U_4, U_5\right\}$ *and for* $i \in \{1, 2\}$

$$\frac{W(1\frac{1}{2}) - W\left(10\right)}{W\left(1\frac{1}{2}\right) - W\left(\frac{1}{2}\frac{1}{2}\right)} < \frac{\mu_i}{1-\mu_i}.$$

*Then in any optimal mechanism* $g$, $P_1^g < P_2^g$.

We now seek to find a $\mu$ that satisfies all the conditions in Lemma 5, for which it will be useful to prove a few more lemmas.

**Lemma 6.** *Take any* $\mu$ *with* $\mu_1 = \mu_2 > \frac{1}{2}$ *and such that*

$$W(1\frac{1}{2})\left(1-\mu_1\right) + W(\frac{1}{2}0)\mu_1 > W\left(10\right),$$

*then* $\max\left\{U_2, U_3\right\} < U_4$.

*Proof.* Using $\mu_1 = \mu_2$ we can calculate that following equations

$$\begin{aligned}
\frac{U_2}{\mu\left(HL\right)} &= W(1\frac{1}{2})\frac{1-\mu_1}{\mu_1} + W\left(10\right)\frac{2\mu_1 - 1}{\mu_1} + W(\frac{1}{2}0) \\
\frac{U_3}{\mu\left(HL\right)} &= 2W\left(10\right) \\
\frac{U_4}{\mu\left(HL\right)} &= W(1\frac{1}{2})2\left(1-\mu_1\right) + W(\frac{1}{2}0)2\mu_1
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{U_4 - U_2}{\mu\left(HL\right)} &= W(1\frac{1}{2})\left(2\left(1-\mu_1\right) - \frac{1-\mu_1}{\mu_1}\right) - W\left(10\right)\left(\frac{2\mu_1 - 1}{\mu_1}\right) + W(\frac{1}{2}0)\left(2\mu_1 - 1\right) \\
&= \frac{\left(2\mu_1 - 1\right)}{\mu_1}\left(W(1\frac{1}{2})\left(1-\mu_1\right) + W(\frac{1}{2}0)\mu_1 - W\left(10\right)\right) > 0
\end{aligned}$$

The inequality follows from assumptions in the statement of the lemma and the fact that $2\mu_1 - 1 > 0$. Next we calculate that

$$\begin{aligned}
\frac{U_4 - U_3}{\mu\left(HL\right)} &= W(1\frac{1}{2})2\left(1-\mu_1\right) - 2W\left(10\right) + W(\frac{1}{2}0)2\mu_1 \\
&= 2\left(W(1\frac{1}{2})\left(1-\mu_1\right) + W(\frac{1}{2}0)\mu_1 - W\left(10\right)\right) > 0
\end{aligned}$$

as desired. □

**Lemma 7.** *Take any $\mu$ with $\mu_1 = \mu_2 > \frac{1}{2}$ and such that $W\left(1\frac{1}{2}\right)\mu_1 + W\left(\frac{1}{2}0\right)(1-\mu_1) \neq W\left(\frac{1}{2}\frac{1}{2}\right)$, then $U_1 < \max\{U_4, U_5\}$.*

*Proof.* For contradiction suppose that

$$U_1 \geq \max\{U_4, U_5\}.$$

Using $\mu_1 = \mu_2$ we can calculate that:

$$\begin{aligned}
\frac{U_1}{\mu(HL)} &= W\left(1\frac{1}{2}\right)\frac{1-\mu_1}{\mu_1} + W\left(\frac{1}{2}0\right)\frac{(5\mu_1^2 - 4\mu_1 + 1)}{\mu_1^2} + W\left(\frac{1}{2}\frac{1}{2}\right)\frac{(1-\mu_1)(2\mu_1 - 1)}{\mu_1^2} \\
\frac{U_4}{\mu(HL)} &= W\left(1\frac{1}{2}\right)2(1-\mu_1) + W\left(\frac{1}{2}0\right)2\mu_1 \\
\frac{U_5}{\mu(HL)} &= W\left(\frac{1}{2}0\right)2\left(\frac{2\mu_1 - 1}{\mu_1}\right) + W\left(\frac{1}{2}\frac{1}{2}\right)2\frac{1-\mu_1}{\mu_1}
\end{aligned}$$

Therefore we can calculate that:

$$\begin{aligned}
\frac{U_1 - U_4}{\mu(HL)} &= W(1\frac{1}{2})\left(\frac{1-\mu_1}{\mu_1} - 2(1-\mu_1)\right) + W(\frac{1}{2}0)\left(\frac{(5\mu_1^2 - 4\mu_1 + 1)}{\mu_1^2} - 2\mu_1\right) \\
&\quad + W(\frac{1}{2}\frac{1}{2})\left(\frac{(1-\mu_1)(2\mu_1 - 1)}{\mu_1^2}\right) \\
&= -W\left(1\frac{1}{2}\right)\frac{(2\mu_1 - 1)(1-\mu_1)}{\mu_1} - W\left(\frac{1}{2}0\right)\left(\frac{(1-\mu_1)^2(2\mu_1 - 1)}{\mu_1^2}\right) \\
&\quad + W\left(\frac{1}{2}\frac{1}{2}\right)\frac{(1-\mu_1)(2\mu_1 - 1)}{\mu_1^2}
\end{aligned}$$

Since $U_1 \geq U_4$ we must have

$$W\left(1\frac{1}{2}\right)\mu_1 + W\left(\frac{1}{2}0\right)(1-\mu_1) \leq W\left(\frac{1}{2}\frac{1}{2}\right)$$

And we can also calculate that

$$\begin{aligned}
\frac{U_1 - U_5}{\mu(HL)} &= W\left(1\frac{1}{2}\right)\frac{1-\mu_1}{\mu_1} + W\left(\frac{1}{2}0\right)\left(\frac{5\mu_1^2 - 4\mu_1 + 1}{\mu_1^2} - 2\frac{2\mu_1 - 1}{\mu_1}\right) \\
&\quad + W\left(\frac{1}{2}\frac{1}{2}\right)\left(\frac{(1-\mu_1)(2\mu_1 - 1)}{\mu_1^2} - 2\frac{1-\mu_1}{\mu_1}\right) \\
&= W\left(1\frac{1}{2}\right)\frac{1-\mu_1}{\mu_1} + W\left(\frac{1}{2}0\right)\frac{(1-\mu_1)^2}{\mu_1^2} - W\left(\frac{1}{2}\frac{1}{2}\right)\frac{1-\mu_1}{\mu_1^2}
\end{aligned}$$

Since $U_1 \geq U_5$, we must have

$$W\left(1\frac{1}{2}\right)\mu_1 + W\left(\frac{1}{2}0\right)(1-\mu_1) \geq W\left(\frac{1}{2}\frac{1}{2}\right)$$

Combining with what we had before gives

$$W\left(1\frac{1}{2}\right)\mu_1 + W\left(\frac{1}{2}0\right)(1-\mu_1) = W\left(\frac{1}{2}\frac{1}{2}\right)$$

which contradicts the assumptions in the statement of this lemma. And it follows that

$$U_1 < \max\{U_4, U_5\}.$$

□

**Lemma 8.** *Let $\alpha \in \left(\frac{1}{2}, 1\right)$ such that $W\left(1\frac{1}{2}\right)(1-\alpha) + W(\frac{1}{2}0)\alpha > W(10)$ and $W\left(1\frac{1}{2}\right)\alpha + W\left(\frac{1}{2}0\right)(1-\alpha) \neq W\left(\frac{1}{2}\frac{1}{2}\right)$. Then there exists a $\varepsilon > 0$ such that for any prior $\mu$ with $\mu_1 = \alpha, \mu_2 \in (\alpha - \varepsilon, \alpha)$ we have*

$$\max\{U_1, U_2, U_3\} < \max\{U_4, U_5\}.$$

*Proof.* For this proof we will treat each $U_i$ as a function $\Delta\Theta \to \mathbb{R}$ where $U_i(\mu)$ is the value of solution $i$ when the common prior is $\mu$. It is clear that $U_i(\cdot)$ is a continuous function for each $i$. Suppose for contradiction no such $\varepsilon$ exists then we can take a strictly positive sequence $\varepsilon_n$ such that $\varepsilon_n \to 0$ and define $\mu^n$ as $\mu_1^n = \alpha, \mu_2^n = \alpha - \varepsilon_n$ and at each $n$ we would have

$$\max\{U_1(\mu^n), U_2(\mu^n), U_3(\mu^n)\} \geq \max\{U_4(\mu^n), U_5(\mu^n)\}.$$

Clearly $\mu^n \to \mu$ where $\mu_1 = \mu_2 = \alpha$. By continuity:

$$\max\{U_1(\mu), U_2(\mu), U_3(\mu)\} \geq \max\{U_4(\mu), U_5(\mu)\}.$$

But $\mu$ obeys the assumptions in the statements of Lemmas 6 and 7, from which it follows that

$$\max\{U_1(\mu), U_2(\mu), U_3(\mu)\} < \max\{U_4(\mu), U_5(\mu)\}.$$

So we have derived our desired contradiction. □

We now complete the proof of Proposition 2 by constructing a $\mu$ that satisfies the conditions in Lemma 5. By supermodularity it is clear that

$$\frac{W\left(1\frac{1}{2}\right)-W\left(10\right)}{W\left(1\frac{1}{2}\right)-W\left(\frac{1}{2}\frac{1}{2}\right)}<\frac{W\left(1\frac{1}{2}\right)-W\left(10\right)}{W\left(10\right)-W\left(\frac{1}{2}0\right)}$$

and Equation (6) guarantees that

$$\frac{W\left(1\frac{1}{2}\right)-W\left(10\right)}{W\left(10\right)-W\left(\frac{1}{2}0\right)}>1,$$

hence we can find $\frac{1}{2}<\alpha_L<\alpha_H<1$ such that for all $\alpha\in(\alpha_L,\alpha_H)$

$$\frac{W\left(1\frac{1}{2}\right)-W\left(10\right)}{W\left(1\frac{1}{2}\right)-W\left(\frac{1}{2}\frac{1}{2}\right)}<\frac{\alpha}{1-\alpha}<\frac{W\left(1\frac{1}{2}\right)-W\left(10\right)}{W\left(10\right)-W\left(\frac{1}{2}0\right)}. \tag{8}$$

Moreover, since $W\left(1\frac{1}{2}\right)>W\left(\frac{1}{2}0\right)$, we know there exists $\alpha^*\in(\alpha_L,\alpha_H)$ such that

$$W\left(1\frac{1}{2}\right)\alpha^*+W\left(\frac{1}{2}0\right)(1-\alpha^*)\neq W\left(\frac{1}{2}\frac{1}{2}\right).$$

Define $\mu$ so that $\mu_1=\alpha^*$. By Lemma 8, we know that there exists an $\varepsilon>0$ small enough so that

$$S:=(\alpha^*-\varepsilon,\alpha^*)\cap(\alpha_L,\alpha_H),$$

is non-empty, and if $\mu_1=\alpha^*$ and $\mu_2\in S$, then

$$\max\left\{U_1,U_2,U_3\right\}<\max\left\{U_4,U_5\right\}.$$

How we constructed $\mu_1,\mu_2$, and Equation (8) together imply $\mu$ obeys the conditions in the statement of Lemma 3. Hence, by Lemma 5 we can conclude that $P_1^g<P_2^g$ for every optimal mechanism $g$ at $\mu$, as desired.

## B.5. Proof of Proposition 3

For this proof we will use the same notation and several of the lemmas that were introduced in section B.4.

**Lemma 9.** *Fix any $\mu$. Let $a,b,c,d\in\mathbb{R}$, such that $a>0$ and $a+b+c+d=0$. Then there exists a natural number $M$ such that for any symmetric principal payoff functions that has* *complementarities above degree $M$ at $(H, L)$*

$$aW\left(1\frac{1}{2}\right) + bW\left(\frac{1}{2}\frac{1}{2}\right) + cW\left(10\right) + dW\left(\frac{1}{2}0\right) > 0.$$

*Proof.* Set $M$ high enough so that

$$\frac{|b| + |c|}{M} < a.$$

For any symmetric principal payoff function with complementarities above degree $M$ at $(H, L)$ we have that

$$\begin{aligned} a\left(W\left(1\frac{1}{2}\right) - W\left(\frac{1}{2}0\right)\right) &> \frac{|b| + |c|}{M}\left(W\left(1\frac{1}{2}\right) - W\left(\frac{1}{2}0\right)\right) \\ &> \frac{|b|}{M}\left(W\left(1\frac{1}{2}\right) - W\left(10\right)\right) + \frac{|c|}{M}\left(W\left(1\frac{1}{2}\right) - W\left(\frac{1}{2}\frac{1}{2}\right)\right) \\ &> |b|\left(W\left(\frac{1}{2}\frac{1}{2}\right) - W\left(\frac{1}{2}0\right)\right) + |c|\left(W\left(10\right) - W\left(\frac{1}{2}0\right)\right). \end{aligned}$$

The first inequality uses how $M$ was defined. The second inequality uses the monotonicity of $W$. The third inequality uses the complementarities being above degree $M$ at $(H, L)$. We then have that

$$aW\left(1\frac{1}{2}\right) - |b|W\left(\frac{1}{2}\frac{1}{2}\right) - |c|W\left(10\right) + \left(-a - |b| - |c|\right)W\left(\frac{1}{2}0\right) > 0.$$

It is clear that $b \geq -|b|$, $c \geq -|c|$ and $-a - |b| - |c| \leq -a - b - c = d$. And these inequalities, along with the previous display equation and the fact that $W(\cdot)$ is defined to be always positive, imply that

$$aW\left(1\frac{1}{2}\right) + bW\left(\frac{1}{2}\frac{1}{2}\right) + cW\left(10\right) + dW\left(\frac{1}{2}0\right) > 0,$$

as desired. □

Define $U_i$ for $i = 1, ..., 7$ as was done in the proof of Proposition 2 using the list of cases enumerated there. Now for any $\varepsilon \in \left(0, \frac{1}{4}\right)$ define prior $\mu^\varepsilon$ as

$$\mu_1^\varepsilon = \frac{3}{4} + \varepsilon \text{ and } \mu_2^\varepsilon = \frac{3}{4} - \varepsilon.$$

In each of the mechanism in cases $4 - 7$, we have that $P_2^g - P_1^g$ at $\mu^\varepsilon$ converges to $\frac{1}{2}$ as $\varepsilon \to \frac{1}{4}$. Therefore we can find $\varepsilon^* \in \left(0, \frac{1}{4}\right)$ such that $P_2^g - P_1^g > \lambda$ at $\mu^{\varepsilon^*}$ in all those cases. Set $\mu = \mu^{\varepsilon^*}$.

We now use Lemma 9 to show that at $\mu$ there exists a $M^*$ such that any symmetric principal payoff function that has complementarities above degree $M$ at $(H,L)$ will have the property:

$$\max\{U_1,U_2,U_3\}<U_4.$$

To see why note that for each $i=1,2,3$ one can write

$$U_4-U_i=aW\left(1\frac{1}{2}\right)+bW\left(\frac{1}{2}\frac{1}{2}\right)+cW\left(10\right)+dW\left(\frac{1}{2}0\right)$$

for the appropriate choice of $a,b,c,d\in\mathbb{R}$ with $a+b+c+d=0$. One can check that $a>0$ for each $i=1,2,3$ at $\mu$. Therefore, for each $i=1,2,3$, Lemma 9 implies there exists an $M_i^*$ so that the $U_4-U_i>0$ for all principal payoff functions that have complementarities above degree $M_i^*$ at $(H,L)$. We will also make use of the following lemma.

**Lemma 10.** *For any $M>1$, if the principal's payoff function has complementarities above degree $M$ at $(H,L)$, then*

$$\frac{W\left(1\frac{1}{2}\right)-W\left(10\right)}{W\left(1\frac{1}{2}\right)-W\left(\frac{1}{2}\frac{1}{2}\right)}\le\frac{M}{M-1}.$$

*Proof.* First suppose that $W\left(\frac{1}{2}\frac{1}{2}\right)\le W\left(10\right)$, then we have

$$\frac{W\left(1\frac{1}{2}\right)-W\left(10\right)}{W\left(1\frac{1}{2}\right)-W\left(\frac{1}{2}\frac{1}{2}\right)}\le\frac{W\left(1\frac{1}{2}\right)-W\left(\frac{1}{2}\frac{1}{2}\right)}{W\left(1\frac{1}{2}\right)-W\left(\frac{1}{2}\frac{1}{2}\right)}=1\le\frac{M}{M-1},$$

as desired.

Now suppose that $W\left(\frac{1}{2}\frac{1}{2}\right)>W\left(10\right)$. By definition of complementarities above degree $M$ we have that

$$W\left(1\frac{1}{2}\right)-W\left(10\right)\ge M\left(W\left(\frac{1}{2}\frac{1}{2}\right)-W\left(\frac{1}{2}0\right)\right).$$

By monotonicity, $W\left(10\right)>W\left(\frac{1}{2}0\right)$ so that

$$W\left(1\frac{1}{2}\right)-W\left(10\right)\ge M\left(W\left(\frac{1}{2}\frac{1}{2}\right)-W\left(10\right)\right).$$

We can rearrange this inequality to get

$$W\left(1\frac{1}{2}\right)\ge MW\left(\frac{1}{2}\frac{1}{2}\right)-W\left(10\right)\left(M-1\right).$$

Since we assumed $W\left(\frac{1}{2}\frac{1}{2}\right) > W(10)$, it follows that

$$\frac{x - W(10)}{x - W\left(\frac{1}{2}\frac{1}{2}\right)}$$

is decreasing in $x$ as long as the numerator and denominator are both positive. Therefore we have that

$$\begin{aligned}\frac{W\left(1\frac{1}{2}\right) - W(10)}{W\left(1\frac{1}{2}\right) - W\left(\frac{1}{2}\frac{1}{2}\right)} &\leq \frac{MW\left(\frac{1}{2}\frac{1}{2}\right) - W(10)(M-1) - W(10)}{MW\left(\frac{1}{2}\frac{1}{2}\right) - W(10)(M-1) - W\left(\frac{1}{2}\frac{1}{2}\right)} \\ &= \frac{MW\left(\frac{1}{2}\frac{1}{2}\right) - MW(10)}{(M-1)W\left(\frac{1}{2}\frac{1}{2}\right) - (M-1)W(10)} \\ &= \left(\frac{M}{(M-1)}\right)\frac{W\left(\frac{1}{2}\frac{1}{2}\right) - W(10)}{W\left(\frac{1}{2}\frac{1}{2}\right) - W(10)} = \frac{M}{M-1},\end{aligned}$$

as desired. □

Since $\mu_i > \frac{1}{2}$ for each $i$, it follows that

$$\frac{\mu_i}{1-\mu_i} > 1.$$

Therefore we can find $M_4^*$ large enough so that for each $i$

$$\frac{M_4^*}{M_4^* - 1} < \frac{\mu_i}{1-\mu_i}.$$

Now set $M^* = \max\{M_1^*, M_2^*, M_3^*, M_4^*\}$. Fix any symmetric principal payoff function that has complementarities above degree $M^*$ at $(H, L)$. And all the conditions of Lemma 5 are satisfied, which implies that every optimal mechanism has to be in cases 4-7. And we choose $\mu$ in such a way that in each of those cases $P_2^g - P_1^g > \lambda$, as desired.

## B.6. Proof of Proposition 4

Fix an agent $i$ and choose any $t^L \in \Theta_i^L$. Define $W_{\max}$ and $W_{\min}$ as

$$\begin{aligned} W_{\max} &= \max_{\theta, a_{-i} | \theta_i \in \Theta_i^H} \{V(\theta, (1, a_{-i})) - V(\theta, (0, a_{-i}))\} \\ W_{\min} &= \min_{\theta, a_{-i} | \theta_i \in \Theta_i^L} \{V(\theta, (0, a_{-i})) - V(\theta, (1, a_{-i}))\}\end{aligned}$$

By definition of $\Theta_i^L$, $\Theta_i^H$ and fact that $W$ is strictly increasing we have that $W_{\max}$ and $W_{\min}$ are both strictly positive.

Define $M_L$ as

$$M_L = \frac{W_{\max}}{W_{\min} + W_{\max}}.$$

Clearly $M_L \in (0,1)$. Now consider any prior $\mu$ with $\mu_i\left(t^L\right) > M_L$, which is equivalent to

$$0 < -\left(1 - \mu_i\left(t^L\right)\right) W_{\max} + \mu_i\left(t^L\right) W_{\min}.$$

Suppose for contradiction that $g$ is an optimal mechanism with $P_i^g > 0$ at $\mu$. Incentive compatibility requires that $p_i^g\left(t\right) = P_i^g$ for all $t \in supp(\mu_i)$. Therefore, for each $t \in supp(\mu_i)$, we can find $\theta^t, a^t$ such that $\theta_i^t = t$ and $a_i^t = 1$, $\mu(\theta^t) > 0$ and $g\left(\theta^t\right)\left[a^t\right] > 0$. Choose $\varepsilon > 0$ such that

$$\varepsilon < \min_{t \in \Theta_i} \frac{g\left(\theta^t\right)\left[a^t\right]}{\mu_{-i}\left(\theta_{-i}^t\right)}.$$

Now consider the following alternative mechanism $g'$:

$$g'\left(\theta\right)\left[a\right] = \begin{cases} g\left(\theta\right)\left[a\right] - \frac{\varepsilon}{\mu_{-i}(\theta_{-i})} & \text{if } \left(\theta, a\right) = \left(\theta^t, a^t\right) \text{ for some } t \in supp(\mu_i) \\ g\left(\theta\right)\left[a\right] + \frac{\varepsilon}{\mu_{-i}(\theta_{-i})} & \text{if } \left(\theta, a\right) = \left(\theta^t, \left(0, a_{-i}^t\right)\right) \text{ for some } t \in supp(\mu_i) \\ g\left(\theta\right)\left[a\right] & \text{otherwise} \end{cases}$$

To verify that $g'$ is incentive compatible note that for each $t \in supp(\mu_i)$

$$p_i^{g'}\left(t\right) = p_i^g\left(t\right) - \varepsilon = P_i^g - \varepsilon.$$

The payoff difference to the principal between $g'$ and $g$ is given by:

$$\begin{aligned} V\left(g'\right) - V\left(g\right) &= \sum_{t \in supp(\mu_i)} \mu\left(\theta^t\right) \varepsilon \frac{1}{\mu_{-i}\left(\theta_{-i}^t\right)} \left(V\left(\theta^t, \left(0, a_{-i}^t\right)\right) - V\left(\theta^t, a^t\right)\right). \\ &= -\sum_{t \in \Theta_i^H \cap supp(\mu_i)} \varepsilon\left(\mu_i\left(t\right) V\left(\theta^t, a^t\right) - V\left(\theta^t, \left(0, a_{-i}^t\right)\right)\right) \\ &+ \sum_{t \in \Theta_i^L \cap supp(\mu_i)} \varepsilon \mu_i\left(t\right)\left(V\left(\theta^t, \left(0, a_{-i}^t\right)\right) - V\left(\theta^t, a^t\right)\right) \\ &\geq \varepsilon\left(-\sum_{t \in \Theta_i^H \cap supp(\mu_i)} \mu_i\left(t\right) W_{\max} + \sum_{t \in \Theta_i^L \cap supp(\mu_i)} \mu_i\left(t\right) W_{\min}\right) \\ &\geq \varepsilon\left(-\left(1 - \mu_i\left(t^L\right)\right) W_{\max} + \mu_i\left(t^L\right) W_{\min}\right) \\ &> 0. \end{aligned}$$

And we have violated the optimality of $g$ which gives us our desired contradiction, which proves that $P_i^g = 0$ as desired.

We can use a similar argument to prove the second part of the proposition and find an appropriate $M_H$. Choose any $t^H \in \Theta_i^H$ and define

$$\begin{aligned} W_{\max} &= \max_{\theta, a_{-i} | \theta_i \in \Theta_i^L} \left\{ V\left(\theta, (0, a_{-i})\right) - V\left(\theta, (1, a_{-i})\right) \right\} \\ W_{\min} &= \min_{\theta, a_{-i} | \theta_i \in \Theta_i^L} \left\{ V\left(\theta, (1, a_{-i})\right) - V\left(\theta, (0, a_{-i})\right) \right\}. \end{aligned}$$

Define

$$M_H = \frac{W_{\max}}{W_{\min} + W_{\max}}.$$

It is easy to verify that $M_H \in (0,1)$, and assume that $\mu\left(t^H\right) > M_H$ which is equivalent to

$$0 < -\left(1 - \mu\left(t^H\right)\right) W_{\max} + \mu\left(t^H\right) W_{\min}.$$

Now suppose for contradiction that an optimal mechanism takes action 1 on agent $i$ with probability less than 1. Similar to above, we then construct an alternative mechanism $g'$ which slightly raises the probability of action 1 on agent $i$, and show that it delivers a higher payoff than the original mechanism. Hence we contradict the optimality of $g$ and are done.

## B.7. Proof of Proposition 5

Let $\theta^L$ and $\theta^H$ be the shared smallest and highest type of all the agents. Fix any prior $\mu$, and suppose $g$ does not exhibit meritocracy at $\mu$, which means there exists $j, k$ such that $\mu_j >_{FOSD} \mu_k$ and $P_j^g < P_k^g$. For any $\varepsilon \in \mathbb{R}$ and any agent $i$, define $\mu_i^\varepsilon \in \Delta\Theta_i$ as

$$\mu_i^\varepsilon\left(\theta_i\right) := \begin{cases} \mu_i\left(\theta_i\right) - \varepsilon & \text{if } \theta_i = \theta^H \\ \mu_i\left(\theta_i\right) + \varepsilon & \text{if } \theta_i = \theta^L \\ \mu_i\left(\theta_i\right) & \text{otherwise} \end{cases}.$$

By the fact that we have $\mu_j >_{FOSD} \mu_k$, we can fix an $\varepsilon > 0$ small enough so that $\mu_j >_{FOSD} \mu_j^\varepsilon >_{FOSD} \mu_k^{-\varepsilon} >_{FOSD} \mu_k$, and moreover $\mu_j^\varepsilon$ and $\mu_k^{-\varepsilon}$ are both well defined. That we can find such an $\varepsilon$ uses the fact that we defined a strict notion of first-order stochastic dominance instead of the standard notion. Define

$$M := 1 + \frac{\varepsilon\left(1 - P_j^g\right) P_k^g}{4n},$$

and it is clear that $M > 1$. Fix any symmetric principal payoff function with complementarities below degree $M$ at $\Theta$. We want to show $g$ is not optimal. For conciseness we set $L = X_i(\theta^L, 1)$ and $H = X_i(\theta^H, 1)$ for every agent $i$. Also we set $X_0 = X_i(\theta_i, 0)$ for any agent $i$ and any $\theta_i \in \Theta_i$. Let $\overrightarrow{L} \in \mathbb{R}^n$ be the vector assigning production $L$ to every agent and let $\overrightarrow{L}_{-i}$ indicate assigning production $L$ to every agent except $i$. Define

$$\mathcal{X} := \{X \in \mathbb{R}^n | \exists \theta, a \text{ such that } X = X(\theta, a)\}.$$

For any agent $i$ and $\mu_i \in \Delta\Theta_i$ define

$$U_i(\mu_i) := \sum_{\theta_i \in \Theta_i} \mu_i(\theta_i) \Big( W(X_i(\theta_i, 1), \overrightarrow{L}_{-i}) - W(X_0, \overrightarrow{L}_{-i}) \Big).$$

Our symmetry assumptions ensure that $U_i(\cdot) = U_j(\cdot)$ for all $i, j$. Therefore, if $\mu_i >_{\text{FOSD}} \mu_j$, then $U_i(\mu_i) > U_j(\mu_j)$.

For any $X, X' \in \mathcal{X}$ the definition of complementarities below degree $M$ imply

$$W(X \vee X') + W(X \wedge X') \leq W(X) + MW(X') - (M-1) W(X \wedge X').$$

For any $X \in \mathcal{X}$, repeatedly applying the previous inequality yields

$$\begin{aligned} W(X) + (n-1) W(\overrightarrow{L}) &\leq -(M-1)(n-1) W(\overrightarrow{L}) + W(X_n, \overrightarrow{L}_{-n}) + M \sum_{i=1}^{n-1} W(X_i, \overrightarrow{L}_{-i}) \\ &\leq -(M-1) n W(\overrightarrow{L}) + M \sum_{i=1}^{n} W(X_i, \overrightarrow{L}_{-i}). \end{aligned}$$

The second inequality uses the monotonicity of $W$ and the fact that $X_n \geq L$. And rearranging the above inequality yields that for any $X \in \mathcal{X}$:

$$W(X) \leq -(Mn-1) W(\overrightarrow{L}) + M \sum_{i=1}^{n} W(X_i, \overrightarrow{L}_{-i}).$$

Using the supermodularity of $W$ and a similar sequence of steps we can derive that for any $X \in \mathcal{X}$ :

$$W(X) \geq -(n-1) W(\overrightarrow{L}) + \sum_{i=1}^{n} W(X_i, \overrightarrow{L}_{-i}).$$

Let $g'$ be any incentive compatible mechanism at $\mu$. Let $V(g')$ be the payoff to the principal from $g'$. The inequalities above can be used to derive that

$$V(g') \leq -(Mn-1)\,W(\overrightarrow{L}) + MnW(X_0, \overrightarrow{L}_{-1}) + M\sum_{i=1}^{n} P_i^{g'} U_i(\mu_i).$$

We can also derive that

$$V(g') \geq -(n-1)\,W(\overrightarrow{L}) + nW(X_0, \overrightarrow{L}_{-1}) + \sum_{i=1}^{n} P_i^{g'} U_i(\mu_i).$$

Deriving these bounds makes use of the symmetry of $W$ as well as the the fact that $P_i^{g'} = p_i^{g'}(\theta_i)$ for all $\theta_i$, which follows from the incentive compatibility of $g'$.

Now take the case that $U_j(\mu_j^{\varepsilon}) \geq 0$, and we have

$$\begin{aligned} U_j(\mu_j) &= U_j(\mu_j^{\varepsilon}) + \varepsilon\left(W(H, \overrightarrow{L}_{-j}) - W(\overrightarrow{L})\right) \\ &\geq \varepsilon\left(W(H, \overrightarrow{L}_{-j}) - W(\overrightarrow{L})\right). \end{aligned}$$

And by definition, for $i$ any we have

$$U_i(\mu_i) \leq W\left(H, \overrightarrow{L}_{-i}\right) - W(\overrightarrow{L}).$$

Let $g'$ be any incentive compatible mechanism such that $P_i^{g'} = P_i^g$ for all $i \neq j$ and $P_j^{g'} = 1$. Using the bounds we derived earlier and our definition of $M$ we have

$$\begin{aligned} V(g') - V(g) &\geq -n(M-1)\left(W(X_0, \overrightarrow{L}_{-1}) - W(\overrightarrow{L})\right) - (M-1)\sum_{i\neq j} P_i^g U_i(\mu_i) + \left(1-P_j^g\right) U_j(\mu_j) \\ &\geq -\frac{\varepsilon\left(1-P_j^g\right)P_k^g}{4}\left(W(X_0, \overrightarrow{L}_{-1}) - W(\overrightarrow{L})\right) - \frac{\varepsilon\left(1-P_j^g\right)P_k^g}{4}\left(W\left(H, \overrightarrow{L}_{-i}\right) - W(\overrightarrow{L})\right) \\ &\quad + \left(1-P_j^g\right)\varepsilon\left(W\left(H, \overrightarrow{L}_{-j}\right) - W(\overrightarrow{L})\right) \\ &\geq \frac{\left(1-P_j^g\right)}{2}\varepsilon\left(W\left(H, \overrightarrow{L}_{-j}\right) - W(\overrightarrow{L})\right) \\ &> 0. \end{aligned}$$

Therefore, $g'$ has a strictly higher payoff than $g$, and $g$ cannot be optimal.

Now take the case where $U_j\left(\mu_j^{\varepsilon}\right) < 0$ which implies $U_k\left(\mu_k^{-\varepsilon}\right) < 0$ and therefore that

$$\begin{aligned} U_k\left(\mu_k\right) &= U_k\left(\mu_k^{-\varepsilon}\right) - \varepsilon\left(W(H, \overrightarrow{L}_{-k}) - W(\overrightarrow{L})\right) \\ &\leq -\varepsilon\left(W(H, \overrightarrow{L}_{-j}) - W(\overrightarrow{L})\right). \end{aligned}$$

Let $g'$ be any incentive compatible mechanism such that $P_i^{g'} = P_i^{g}$ for all $i \neq k$ and $P_k^{g'} = 0$. And a similar sequence of inequalities to what we used above will show $g'$ gives a strictly higher payoff than $g$, and hence $g$ cannot be optimal. And that finishes the proof. □

# C. Proofs from Section 6 (For Online Publication)

## C.1. Proof of Theorem 5

We make use of the following lemma the proof of which can be found in section C.1.1.

**Lemma 11.** *Let $\mu \in \Upsilon_I$ and $\mu' \in \Upsilon_p$ such that $\mu = I^{\mu'}$. Then any coordination mechanism that is incentive compatible at $\mu$ is incentive compatible at $\mu'$.*

Fix any $\mu \in \Upsilon_I$ and let $\mu^m \to \mu$ where $\mu^m \in \Upsilon_p$ for each $m$. For every $\varepsilon > 0$, define a perturbed version of the the principal's payoff function, $V^{\varepsilon}$, precisely as was done in the proof of Theorem 2. For each $\varepsilon > 0$, let $\{g^{m,\varepsilon}\}_{m=1}^{\infty}$ be a sequence of mechanisms such that $g^{m,\varepsilon}$ is optimal optimal at $\mu^m$ under payoff function $V^{\varepsilon}$. Since $\mathcal{G}$ is compact, we can assume without loss of generality that this sequence converges and set $g^{\varepsilon} = \lim_{m\to\infty} g^{m,\varepsilon}$. Since the incentive constraints are a finite set of weak inequalities that move continuously with the prior, the set of incentive compatible mechanisms must move upper hemicontinuously with the prior. Hence it follows that $g^{\varepsilon}$ is incentive compatible at $\mu$. Now let $U^{m,\varepsilon}$ and $U^{\varepsilon}$ be the optimal principal payoff at $\mu^m$ and $\mu$ respectively under $V^{\varepsilon}$. Note that the payoff from $g^{\varepsilon}$ equals $\lim_{m\to\infty} U^{m,\varepsilon}$. We will show $g^{\varepsilon}$ is optimal at $\mu$ under $V^{\varepsilon}$, in other words that

$$\lim_{m\to\infty} U^{m,\varepsilon} = U^{\varepsilon}. \tag{9}$$

Since $g^{\varepsilon}$ is incentive compatible at $\mu$, it is immediate that $\lim_{m\to\infty} U^{m,\varepsilon} \leq U^{\varepsilon}$. For each $m$ and $\varepsilon$, let $h^{m,\varepsilon}$ be an optimal mechanism at prior $I^{\mu^m}$ under $V^{\varepsilon}$ with associated payoff $U^{m,\varepsilon,I}$. We know that $h^{m,\varepsilon}$ is a coordination mechanism since in the proof of Theorem 2 we established that all optimal mechanisms are coordination mechanisms under $V^{\varepsilon}$ at any independent prior. Therefore by Lemma 11, $h^{m,\varepsilon}$ is incentive compatible at $\mu^m$, which implies that for each $m : U^{m,\varepsilon} \geq U^{m,\varepsilon,I}$, which further implies $\lim_{m\to\infty} U^{m,\varepsilon} \geq \lim_{m\to\infty} U^{m,\varepsilon,I}$. Without loss of generality we can suppose $h^{m,\varepsilon}$ converges since $\mathcal{G}$ is compact, and let $h^{\varepsilon} = \lim_{m\to\infty} h^{m,\varepsilon}$ . And the fact that $\mu^m \to \mu$ and $\mu \in \Upsilon_I$ implies $I^{\mu^m} \to \mu$. In section B.3, we proved the optimal mechanism moves upper hemicontinuously in the domain $\Upsilon_I$, which implies $h^{\varepsilon}$ is optimal at $\mu$ under $V^{\varepsilon}$, and therefore

$$\lim_{m\to\infty} U^{m,\varepsilon} \geq \lim_{m\to\infty} U^{m,\varepsilon,I} = U^{\varepsilon}.$$

We have now shown Equation (9), which implies that $g^{\varepsilon}$ is optimal at $\mu$ under $V^{\varepsilon}$ and we know that $g^{\varepsilon}$ is a coordination mechanism because in the proof of Theorem 2 we showed that all optimal mechanism under $V^{\varepsilon}$ at a prior in $\Upsilon_I$ are coordination mechanisms . Now define $g := \lim_{\varepsilon\to 0} g^{\varepsilon}$ and $g^m := \lim_{\varepsilon\to 0} g^{m,\varepsilon}$. By the theorem of the maximum, the set of optimal mechanisms moves upper hemicontinuously in $\varepsilon$, and, hence, $g$ is an optimal coordination mechanism under $V$ at $\mu$ and $g^m$ is an optimal mechanism under $V$ at $\mu^m$ for each $m$. Moreover, by exchanging the order of the limits we get $\lim_{m\to\infty} g^m = g$, as desired.

### C.1.1. Proof of Lemma 11

Let $\mu \in \Upsilon_I$ and $\mu' \in \Upsilon_p$ such that $\mu = I^{\mu'}$. Let $g$ be any incentive compatible coordination mechanism at $\mu$. Fix a single agent $i$ and we will show $\mu$ is incentive compatible at $\mu'$ for agent $i$. We will assume nothing specific about agent $i$, so that incentive compatibility for the rest of the agents can be proved identically. Mechanism $g$ is incentive compatible for agent $i$ at $\mu'$ if and only if for $(t,t') = (L,H)$ or $(t,t') = (H,L)$ the following inequality holds:

$$\sum_{\theta_{-i}\in\Theta_{-i}} g_i\left(t,\theta_{-i}\right)\mu'\left(\theta_{-i}|\theta_i = t\right) \geq \sum_{\theta_{-i}\in\Theta_{-i}} g_i\left(t',\theta_{-i}\right)\mu'\left(\theta_{-i}|\theta_i = t\right).$$

We know that $g$ is incentive compatible at $\mu$ which implies that

$$\sum_{\theta_{-i}\in\Theta_{-i}} g_i\left(H,\theta_{-i}\right)\mu\left(\theta_{-i}\right) = \sum_{\theta_{-i}\in\Theta_{-i}} g_i\left(L,\theta_{-i}\right)\mu\left(\theta_{-i}\right).$$

Hence it suffices for us to show that for $(t,t') = (L,H)$ or $(t,t') = (H,L)$

$$\begin{aligned}\sum_{\theta_{-i}\in\Theta_{-i}} g_i\left(t,\theta_{-i}\right)\mu'\left(\theta_{-i}|\theta_i = t\right) &\geq \sum_{\theta_{-i}\in\Theta_{-i}} g_i\left(t,\theta_{-i}\right)\mu\left(\theta_{-i}\right), \qquad \text{and} \\ \sum_{\theta_{-i}\in\Theta_{-i}} g_i\left(t',\theta_{-i}\right)\mu\left(\theta_{-i}\right) &\geq \sum_{\theta_{-i}\in\Theta_{-i}} g_i\left(t',\theta_{-i}\right)\mu'\left(\theta_{-i}|\theta_i = t\right).\end{aligned}$$

We will only treat the case where $(t, t') = (H, L)$; the other case follows mutatis mutandis. Since we only have two types per agent, the orders $\succeq^*$ and $\geq$ coincide. Therefore, by definition of a coordination mechanism, $g_i(L, \theta_{-i})$ is decreasing in $\theta_{-i}$ and $g_i(H, \theta_{-i})$ is increasing in $\theta_{-i}$. And since $\mu' \in \Upsilon_P$, we know that $\mu_{-i}(\theta_i|\theta_i = H)$ first-order stochastically dominates $\mu(\theta_{-i})$, and $\mu(\theta_{-i})$ first-order stochastically dominates $\mu_{-i}(\theta_i|\theta_i = L)$. Standard properties of stochastic dominances then imply the required inequalities. (See page 266 of Shaked and Shanthikumar (2007) for details).

## C.2. Proof of Theorem 6

Fix any $\mu \in \Upsilon_I$ and let $g$ to be an optimal coordination mechanism at $\mu$. Let $\theta^H$ be the state where every agent is type $H$, and let $\theta^L$ be the state where every agent is type $L$. Let $\Theta^{n-1}$ be the set of $n$ states in which exactly $n-1$ agents are type $H$. Let $\Theta^1$ be the set of $n$ states in which exactly 1 agent is type $H$ type.

For any $\varepsilon > 0$ define $v^\varepsilon \in \Upsilon$ as

$$\nu^\varepsilon(\theta) = \begin{cases} \mu(\theta) + \varepsilon & \text{if } \theta = \theta^H \text{ or } \theta = \theta^L \\ \mu(\theta) - \frac{\varepsilon}{n} & \text{if } \theta \in \Theta^{n-1} \text{ or } \theta \in \Theta^1 \\ \mu(\theta) & \text{otherwise} \end{cases}.$$

Because $\mu$ is assumed to be full support, there exists $\bar{\varepsilon} > 0$ such that the $v^\varepsilon(\theta) > 0$ for all $\theta \in \Theta$ and $\varepsilon \in (0, \bar{\varepsilon})$. Define the set $\Upsilon_O \subseteq \Upsilon$ as the set of $\nu^\varepsilon$ for all $\varepsilon \in (0, \bar{\varepsilon})$. Clearly $\mu$ is on the boundary of $\Upsilon_O$ and $I^v = \mu$ for any $v \in \Upsilon$.

Since $g$ is incentive compatible at $\mu$, for any agent $i$ and any $\theta_i, \theta_i' \in \Theta_i$

$$\sum_{\theta_{-i}} g_i(\theta_i, \theta_{-i}) \mu(\theta_i, \theta_{-i}|\theta_i) \geq \sum_{\theta_{-i}} g_i(\theta_i', \theta_{-i}) \mu(\theta_i, \theta_{-i}|\theta_i).$$

Let $v \in \Upsilon_O$. We want to show $g$ is strictly incentive compatible at $v$, for which it suffices to show that for each agent $i$ and $\theta_i' \neq \theta_i$:

$$\sum_{\theta_{-i}} g_i(\theta_i, \theta_{-i}) \nu(\theta_i, \theta_{-i}|\theta_i) \geq \sum_{\theta_{-i}} g_i(\theta_i, \theta_{-i}) \mu(\theta_i, \theta_{-i}|\theta_i),$$

$$\sum_{\theta_{-i}} g_i(\theta_i', \theta_{-i}) \mu(\theta_i, \theta_{-i}|\theta_i) \geq \sum_{\theta_{-i}} g_i(\theta_i', \theta_{-i}) \nu(\theta_i, \theta_{-i}|\theta_i),$$

with at least one of the above being strict.

We will only treat the case that $\theta_i = H$; the case where $\theta_i = L$ would be similar. Then the first of the desired inequalities holds strictly if and only if:

$$\sum_{\theta_{-i}} g_i\left(H, \theta_{-i}\right)\left(\nu\left(H, \theta_{-i}|\theta_i\right) - \mu\left(H, \theta_{-i}|\theta_i\right)\right) > 0.$$

Using the fact that $I^v = \mu$ we can rewrite the above inequality as

$$\sum_{\theta_{-i}} g_i\left(H, \theta_{-i}\right)\left(\nu\left(H, \theta_{-i}\right) - \mu\left(H, \theta_{-i}\right)\right) > 0,$$

$$\Leftrightarrow \quad g_i\left(\theta^H\right)\varepsilon - g_i\left(H, \theta_{-i}^L\right)\frac{\varepsilon}{n} - \sum_{\theta \in \Theta^{HL}|\theta_i = H} g_i\left(\theta\right)\frac{\varepsilon}{n} > 0.$$

And this inequality holds due to the fact that $g$ is a non-trivial coordination mechanism. In particular that $g$ is a coordination mechanism implies

$$g_i\left(\theta^H\right) \geq g_i\left(\theta\right),$$

for all $\theta \in \Theta^{n-1}$ such that $\theta_i = H$. And that $g$ is non-trivial implies

$$g_i\left(\theta^H\right) > g_i\left(H, \theta_{-i}^L\right).$$

The second of the desired inequalities holds weakly if and only if:

$$\sum_{\theta_{-i}} g_i\left(L, \theta_{-i}\right)\left(\nu\left(H, \theta_{-i}|\theta_i\right) - \mu\left(H, \theta_{-i}|\theta_i\right)\right) \geq 0,$$

$$\Leftrightarrow \quad - g_i\left(L, \theta_{-i}^H\right)\varepsilon + g_i\left(L, \theta_{-i}^L\right)\frac{\varepsilon}{n} + \sum_{\theta \in \Theta^1|\theta_i = L} g_i\left(\theta\right)\frac{\varepsilon}{n} \geq 0.$$

By the same logic as before, this inequality holds since $g$ is a coordination mechanism. By definition, $\Upsilon_O \subseteq \Upsilon_p$, which implies, using an argument similar to the proof of Theorem 5, that there exists $\Upsilon_O' \subset \Upsilon_O$ such that for any $v \in \Upsilon_O'$ :

$$\max_{g' \in \mathcal{G}} V\left(g'|\nu\right) < V\left(g|\nu\right) + \varepsilon.$$

## C.3. Proof of Theorem 7

We first formally define our coalition-proof mechanism. For any $\varepsilon > 0$, we define $g^\varepsilon$ : $\Theta^n \times \mathbb{Z}_+^n \to \Delta A$. The combined reports of the agents are given by $(T, z) \in \Theta^n \times \mathbb{Z}_+^n$, where $T$ gives all the agents' report about the total type profile and $z$ is the vector of integers used in the betting game. We let $T_i \in \Theta$ and $T_{ij} \in \Theta_j$ denote agent $i$'s report on the total type profile and agent $j$'s type, respectively. We also let $z_i$ denote the integer reported by agent $i$. Let $diag\left(T\right) = \left(T_{11}, T_{22}, ..., T_{nn}\right)$ be the diagonal vector of each agent's report about his own type.

With probability $1 - \varepsilon$, $g^{\varepsilon}$ plays the optimal direct mechanism using reports $diag\left(T\right)$, and with probability $\varepsilon$, $g^{\varepsilon}$ plays a betting mechanism instead. The betting mechanism is described by $g^b : \Theta^n \times \mathbb{Z}_+^n \to \Delta A$. We let $g_i^b\left(T, z\right)$ denote the marginal probability that the betting mechanism takes action 1 on agent $i$ following report $\left(T, z\right)$. We define $g^b$ so that

$$g_i^b\left(T, z\right) = \begin{cases} \arg\min_{\theta_{-i} \in \Theta_{-i}} \mu_{-i}\left(\theta_{-i}\right) & \text{if } z_i = 0 \\ \left(\mu_{-i}\left(T_{i,-i}\right)\right)^{-1} \arg\min_{\theta_{-i} \in \Theta_{-i}} \mu_{-i}\left(\theta_{-i}\right) & \text{if } z_i > \max_{j \neq i} z_j, \text{ and } diag\left(T\right) = T_i \\ 0 & \text{otherwise} \end{cases}.$$

The above definition only pins down the marginal probability of acting on each agent at each state. Since the betting game exists only to incentivize the agents, the marginal probabilities are all that matter. But to be concrete, we could fully define $g^b$ by specifying that the actions are taken independently across agents within every state

We now show that $g^{\varepsilon}$ is coalition-proof for all $\varepsilon > 0$. Fix $\varepsilon \in (0, 1)$, and suppose for contradiction that $\left(\left\{\sigma_i\right\}_{i=1}^n, I\right)$ is a valid coalition deviation for $g^{\varepsilon}$. For any $t_i \in \Theta_i$, let $\rho_i\left(t_i\right)$ denote the probability that agent $i$ reports his own type to be $t$ under strategy $\sigma_i$. For any $t_{-i} \in \Theta_{-i}$, we will define $\rho_{-i}\left(t_{-i}\right) := \prod_{j \neq i} \rho_j\left(t_j\right)$. For each agent $i$ define

$$D_i := \max_{\theta_i \in \Theta_i} \left(\rho_i\left(\theta_i\right) - \mu_i\left(\theta_i\right)\right).$$

Since $\rho_i(\theta_i)$ and $\mu_i(\theta_i)$ are both non-negative and sum to 1 over $\theta_i \in \Theta_i$, we know that $D_i \geq 0$ for all $i$. Let $S \subseteq I$ be the set of all agents with $D_i > 0$.

First suppose that $|S| \leq 1$. If $S$ is non-empty, then let $i \in S$. Otherwise let $i \in I$. In either case, for any $j \neq i$ we have $D_j = 0$ which implies $\rho_j\left(\theta_j\right) = \mu_j\left(\theta_j\right)$ for all $\theta_j \in \Theta_j$. Since types are drawn independently, for any possible strategy $i$ could employ he receives the same payoff as if each other agent was reporting their own type truthfully. And since the optimal direct mechanism is incentive compatible, truthfully reporting his own type is a best response for agent $i$. Moreover, if agent $i$ gives any report $z_i > 0$, his payoff from $g^b$ is bounded from above by what he would get if he always had the highest $z_i$. At that upper-bound his expected payoff from $g^b$ when reporting $t \in \Theta_{-i}$ is

$$\mu_{-i}(t)\left(\mu_{-i}(t)\right)^{-1} \arg\min_{\theta_{-i}\in\Theta_{-i}} \mu_{-i}(\theta_{-i}) = \arg\min_{\theta_{-i}\in\Theta_{-i}} \mu_{-i}(\theta_{-i}),$$

which is the same payoff from $g^b$ for setting $z_i = 0$. Hence $z_i = 0$ is a best response for agent $i$. So we have established that truthful reporting and setting $z_i = 0$ is a best response for agent $i$, which means agent $i$ receives the same payoff as he would in the truthful equilibrium. But that contradicts condition (3) of the definition of a valid coalition deviation.

Hence it must be that $|S| \geq 2$. For each $i \in S$ there exists $\theta_i^* \in \Theta_i$ such that $\rho_i(\theta_i^*) > \mu_i(\theta_i^*)$. Now fix any agent $i \in I$ . Choose any $t \in \Theta_{-i}$ such that $t_j = \theta_j^*$ for all $j \in S$ which implies $\rho_j(t_j) > \mu_j(t_j)$. And for all $j \notin S$ we have $D_j = 0$ and hence $\rho_j(t_j) = \mu_j(t_j)$. And since $S\setminus\{i\}$ is non-empty we have that $\rho_{-i}(t) > \mu_{-i}(t)$. Therefore:

$$\rho_{-i}(t)\left(\mu_{-i}(t)\right)^{-1} \arg\min_{\theta_{-i}\in\Theta_{-i}} \mu_{-i}(\theta_{-i}) > \arg\min_{\theta_{-i}\in\Theta_{-i}} \mu_{-i}(\theta_{-i}) .$$

The left-hand side of the above inequality is the payoff agent $i$ gets from the $g^b$ when reporting $t$ for the other agent's types and when reporting the highest integer. And the right-hand side of the inequality gives agent $j$'s payoff from reporting $z_j = 0$. Therefore each agent in $I$ can profit by having the highest $z_i$ and making an appropriate bet $t \in \Theta_{-i}$. Therefore each agent in $I$'s best response to $\{\sigma_i\}_{i=1}^n$ involves increasing $z_i$ until their probability of having the strictly highest $z_i$ is one. However, we know there are at least two agents in $I$, and it is impossible for two agents to have the strictly highest $z_i$ with probability 1. Hence any such $\left(\{\sigma_i\}_{i=1}^n, I\right)$ will violate condition (2) of the definition of coalition proof and we have ruled out coalition deviations with $|S| \geq 2$. And that covers all the possibilities and shows that $g^\varepsilon$ is coalition proof for all $\varepsilon \in (0,1)$.