ASSORTATIVE MECHANISMS IN SCREENING WITHOUT TRANSFERS

NEMANJA ANTIĆ* AND KAI STEVERSON†

Abstract. A principal takes a binary action on each of n agents. Agents have type-independent preferences: every type of every agent prefers action 1 to action 0. Since no transfers are possible, a mechanism is incentive compatible only if the marginal probability of taking action 1 on an agent does not depend on his reported type. When the principal's payoff function exhibits complementarities across agents, she can benefit from an assortative mechanism. Relative to the default action 0, assortative mechanisms improve the principal's payoff in states with many high-quality agents (by frequently taking the action 1 on high types), but take a payoff loss necessary for incentive compatibility in states with many low-quality agents (by taking action 1 on low types). In some environments this can lead to optimal mechanisms where the marginal probability of taking action 1 is higher for an ex-ante inferior agent, a phenomenon we call strategic favoritism.

1. Introduction

We study a multi-agent screening problem without transfers. A principal has a project which requires n tasks to be completed. For each task, the principal decides whether to delegate that aspect of the project to an agent or not. The principal wants to delegate only to highly skilled agents, because her payoff is affected by the agents' types. Agents, however, have type-independent preferences: every type of every agent would like to be part of the project and cares about nothing else. This kind of problem arises when an executive assembles an internal team for a prestigious venture, when grant funding is being allocated, or when an organization is downsizing.

The literature on screening when agents have type-independent preferences typically has additional instruments or structure which the principal can exploit, e.g., repeated interactions or costly state verification. We ask what the principal can do without any additional tools, considering that

Date: November 2025.
Key words and phrases. Screening, Mechanism Design, Supermodularity.

* Nemanja Antic: Kellogg School of Management, Managerial Economics and Decision Sciences Department, Kellogg School of Management, Northwestern University, Evanston, Illinois; *nemanja.antic@kellogg.northwestern.edu*

† Kai Steverson: DCI Solutions, Aberdeen, MD.

Thanks to Eddie Dekel, Tim Fedderson, Faruk Gul, Peter Klibanoff, Gilat Levy, Stephen Morris, Wolfgang Pesendorfer, Nicola Persico, Ariel Rubinstein, Doron Ravid, Karl Schlag and Asher Wolinsky for helpful discussions. A special thank you to Meg Meyer for extensive discussions. Financial support from NIH grant R01DA038063 is gratefully acknowledged. This paper was previously titled 'Screening through Coordination'.

agents will report whatever type gives them the greatest probability of being delegated the task (included on the team, awarded the grant, or kept in the job).

In any incentive compatible mechanism, the marginal probability of an agent getting his preferred action must be independent of his reported type. The screening problem is trivial if the principal's payoff function is separable across agents; she cannot improve on taking an action that is not responsive to the agents' reports.

However, when the principal's payoff function is not separable across agents, the joint distribution of the agents' types, and not just the marginals, affect her payoff. As agents do not care about the joint distribution, the principal is able to offer them a set of alternatives which satisfies incentive compatibility and improves the principal's payoff.

We focus on supermodular objective functions for the principal, which arise when there are complementarities across agents, including in the applications above. Assortative mechanisms, which as much as possible delegate to agents of similar quality, be it high or low, are optimal in such environments. Rewarding low quality agents is necessary for incentive compatibility: if only high-quality agents get their preferred action, no agent would admit to being low quality. Assortative mechanisms accept the failure (ex post suboptimal decisions) necessary for incentive compatibility when complementarities are weak, in order to succeed when complementarities are strong.

Unlike in much of the comparative statics literature and related Bayesian persuasion models (see Section 1.1), the marginal probability with which an agent gets his preferred action is endogenously determined by the principal. This can give rise to a phenomenon we term strategic favoritism, where an agent with a worse type distribution gets the preferred action more often than an agent with a better type distribution (while the agents are symmetric in every other way).

To see how strategic favoritism can arise, consider two agents, each of whom can either be a high or a low type. Agent 1 has an even chance of being high or low, while agent 2 is very likely to be the high type. For many supermodular objective functions, the principal has large gains from delegating to both agents if each type is high. Since it is very likely that agent 2 is the high type, if agent 1 reports "high" he will get his preferred action with a high probability. To satisfy agent 1's incentive compatibility, the principal must delegate to him with an equally high probability when he reports "low". The same argument for agent 2 does not guarantee that he is often assigned the project, because agent 1 is the high type with probability one-half.[1]

[1]Of course, incentive compatibility for agent 2 is satisfied if he is delegated the task with probability 1. While this is indeed sometimes optimal, for supermodular objective functions where delegating to a low type is sufficiently harmful,

We address two robustness concerns in our model. In our baseline setting with independent priors, incentive compatibility requires all incentive constraints to bind. By introducing correlation in agent types, we show that our main results do not depend in a knife-edge way on everywhere-binding incentive constraints. We also show that assortative mechanisms can be made coalition-proof. Since these mechanisms delegate to agents when they all report a high type, the agents could benefit from collusion. 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. In Section 1.1 we discuss the related literature. Section 2 presents the model and a simple example. In Section 3 we give our main results on assortative mechanisms. Section 4 shows the possibility of strategic favoritism in the optimal mechanism. In Section 5 we address questions of robustness. Section 6 concludes.

1.1. Literature Review

Our model builds on the literature on screening when agents have type-independent preferences. This literature differs from our paper, since it gives the principal some way to check agents' reports. This can take the form of (1) direct verification, where either the principal, or a third party, has the (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 different agents by exploiting correlation between agents' types. The papers in each of these three categories do not benefit from assortative mechanisms 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.

A related paper by Ben-Porath, Dekel, and Lipman (2014) falls in the first category of direct verification. In their paper, the principal allocates an indivisible good among n agents, who all want the good regardless of their type. The principal's payoff depends only on the type of agent who receives the good. 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 truth-telling.

The use of repeated interactions in such screening problems goes back to the literature on "linking decisions" (Radner, 1981; Rubinstein and Yaari, 1983; Fang and Norman, 2006; Jackson and Sonnenschein, 2007). For example, in Jackson and Sonnenschein (2007), how often each agent can

---

the principal is better off not delegating to agent 2 when they are a low type and the other player is a high type. This leads to strategic favoritism.

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.[2] While certain applications of type-independent screening have dynamic components, our exercise adds value in two ways: (1) some applications, such as downsizing, are clearly best modeled as static and (2) the static problem is an interesting benchmark regarding what is possible without resorting to dynamic incentives.

In our final category of using correlation to check agents' reports, we are unaware of any application to kind of screening problems we study. 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). These ideas could be applied to screening problems with type-independent preferences.

Technically, our work builds on the literature on comparative statics (Tchen, 1980; Müller and Scarsini, 2000; Meyer and Strulovici, 2012, 2017), which considers ranking joint distributions according to a supermodular order (which is what a principal with a supermodular objective would like). This literature takes marginal distributions as given, whereas in our paper the marginals are endogenous: the probability of taking each agent's preferred action is part of the solution to the screening problem.

We also relate to the literature on persuasion where there are either multiple attributes and correlation neglect[3] or multiple agents. Levy, Moreno de Barreda, and Razin (2021) consider a model in which a sender tries to persuade an audience which neglects correlation across data sources, but is aware of the marginal distributions of each source. Arieli and Babichenko (2019) study a related model of persuasion, in which a principal is able to send private signals to n agents. In these models, the sender can design the joint distribution, but must respect fixed marginals due to Bayes plausibility.

2. Model

There is one principal and n agents, denoted $i \in \{1, ..., n\}$. For each agent i, there is finite set $\Theta_i \subset \mathbb{R}$ of types, and the state space is $\Theta := \prod_{i=1}^{n} \Theta_i$. Agents draw their types independently according to distribution $\mu_i \in \Delta\Theta_i$, 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 $\theta \in \Theta$. Without loss of generality assume that μ has full support.

[2]Work by Guo and Hörner (2015), Li et al. (2015) and Lipnowski and Ramos (2015) falls into the second category as well. The optimal mechanisms in these papers use dynamic budgets or tokens that limit how often the agent can receive his preferred action.

[3]See Ortoleva and Snowberg (2015) and Levy and Razin (2015).

For each agent i, the principal chooses a binary action, $a_i \in \{0,1\}$. Therefore, the principal's action space is $A = \{0,1\}^n$. 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 $a_i = 1$.

We can think of the principal as having a project requiring n tasks be completed. The "quality" or "value" of each completed task i is $a_i\theta_i$ and we write $a \circ \theta$ for the Hadamard product of vectors a and θ, i.e., $a \circ \theta = (a_1\theta_1, ..., a_n\theta_n)$. The principal's payoff at state θ when using action profile a is given by $W(a \circ \theta)$ where $W : \mathbb{R}^n \to \mathbb{R}_+$ is any strictly increasing and (weakly) supermodular function. The properties of W ensure that the principal's payoff is increasing in the quality of each agent and exhibits complementarities across agents.

We refer to θ_i as a "high type" if $\theta_i > 0$, and call it a "low type" otherwise. 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 $\Theta_i^H, \Theta_i^L \subset \Theta_i$ 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 his type is high. Incorrectly responding to an agent's type means setting $a_i = 1$ if his type is low.

To avoid trivial cases, we assume that Θ_i^H and Θ_i^L are both non-empty for each agent. For ease of exposition, we assume that $0 \notin \Theta_i$ for all i.[4]

The Principal's Problem

The principal commits to a mechanism, which assigns a lottery over A to every state in Θ. By the revelation principle, we restrict attention to direct mechanisms. A mechanism is any function $g : \Theta \to \Delta A$; let $\mathcal{G}$ denote the space of all mechanisms. We use $g(\theta)[a]$ to denote the probability mechanism g assigns to action a at state θ.

Let $g_i(\theta) := \sum_{a \in A | a_i = 1} g(\theta)[a]$ be the probability that in state θ the mechanism sets $a_i = 1$, $G_i(\theta_i) := \sum_{\theta_{-i} \in \Theta_{-i}} g_i(\theta_i, \theta_{-i}) \mu_{-i}(\theta_{-i})$ be the probability that mechanism g sets $a_i = 1$, conditional on agent i reporting type θ_i and all other agents reporting truthfully.

Fact 1. A mechanism g is incentive-compatible if and only if $G_i(\theta_i) = G_i\left(\theta_i'\right)$ for every agent i and all $\theta_i, \theta_i' \in \Theta_i$.

Since each agent cares only about the probability of $a_i = 1$, if $G_i(\theta_i) > G_i\left(\theta_i'\right)$ then agent i would never report type θ_i'. Hence, truthful reporting requires that $G_i(\theta_i)$ be the same for all $\theta_i \in \Theta_i$.

[4] If Θ_i^H (or Θ_i^L) was empty for agent i, an optimal mechanism would be to take action 0 (or action 1) on agent i and then append the solution for the remaining $n-1$ agents.

Conversely, any agent is happy to report truthfully if he receives action 1 with some probability that does not depend on his report.

The principal's design problem is:

$$\begin{aligned} &\max_{g\in\mathcal{G}} \sum_{\theta\in\Theta, a\in A} \mu(\theta)\, g(\theta)\, [a]\, W(a\circ\theta), \\ \text{s.t.} \qquad &G_i(\theta_i) = G_i\left(\theta_i'\right) \text{ for all } i, \theta_i, \theta_i' \in \Theta_i. \end{aligned}$$

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

Since, agents' preferences are type-independent, exploiting variations in preferences between types, which much of the mechanism design literature focuses on, is not possible. However, agents have "thick" indifference curves relative to the principal and we show how the principal can use this fact for a particular class of principal 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. Fact 1 remains true if some agents always prefer action 1, while others always prefer action 0, so the set of optimal mechanisms is invariant to these changes. We can also allow some low types of each agent to prefer action 0. These 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 other agents.

2.1. Simple Example

Consider a principal that is deciding whether to delegate (take action 1) two aspects of a project to two agents, agent 1 and agent 2. If the principal doesn't delegate an aspect of the project to an agent, she performs that aspect of the task herself (takes action 0). Each agent has a private type, θ_i, either low ($L = -1$) or high ($H = 1$), that determines his suitability for the project. Types are drawn independently and distributed uniformly, i.e., each agent has probability $1/2$ of being the high or the low type.

Agents always want to be delegated the project regardless of their type, while the principal would like to delegate to high types and not to low types. Consider the following payoff function for the principal

$$W(a\circ\theta) = (a_1\theta_1 + 2)(a_2\theta_2 + 2).$$

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 delegating to agent 1, but taking action 0 for agent 2.

Since agents have type-independent preferences, it is not clear if the principal can improve on committing to an action for each agent ex-ante, i.e., by implementing a constant allocation rule for each agent. Without asking the agents to make reports, the principal can either delegate to both agents in all states of the world, never delegate to either agent, or always delegate to one of the agents, but never the other. In all of these instances the principal's expected utility is 4.

State	HH	LH	HL	LL
Action	11	00	00	11
Payoff	9	4	4	1

Table 1: Assortative mechanism

Table 1 presents an example of a mechanism which improves upon this. The principal asks both agents to report their types: if the types match (even if they are low) both agents are delegated the project, while if the types do not match, neither agent is. We call this mechanism **assortative** because agents of high and low ability are grouped together. Truthful reporting is incentive compatible, since from the perspective of each agent there is an even chance of the types matching regardless of what they say, given that the other person is truthfully reporting. This mechanism yields an expected payoff of 4.5 and is optimal.

Note that if the principal's payoff is separable across agents, e.g., $W(a \circ \theta) = (a_1\theta_1 + 2) + (a_2\theta_2 + 2)$, the principal only cares about the marginal distribution of actions taken on each agent and not the joint distribution. Thus, the principal can do no better than treat this as two independent screening problems, one for each agent.[5] In such a screening...

3. Assortative Mechanisms

We begin our discussion of assortative mechanisms by considering environments with fully symmetric agents who only have two possible types. This generalizes the preceding example and highlights the underlying logic. We then provide a general result.

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

[5]The agents already treat these as independent, since each agent only cares about the action he gets.

(1) $\Theta_i = \{L, H\}$ *for all* i,

(2) $\mu_i(H) = p$ *for all* i, *and*

(3) $W(Y) = W(Y')$ *whenever* Y' *is a permutation of* Y.

We use $\overrightarrow{\mathbf{1}}$ and $\overrightarrow{\mathbf{0}}$ to denote a vector of ones and zeros, 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 assortative mechanisms work: they group correct responses on states with many high types, and group incorrect responses on states with many low types. Because of the symmetry of the environment, there is an optimal anonymous mechanism, where two states with the same number of high types take the exact same actions on the high and low types (after taking into account that the identities of the agents who are a high type may have changed). The complementarities in the principal's payoff function reward her for grouping high-output outcomes of the agents together. Sometimes responding incorrectly is necessary for incentive compatibility. Assortative mechanisms accept 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 or takes a default action a^*. In the example in Section 2.1 $m^H = m^L = 1$ and $a^* = \overrightarrow{\mathbf{0}}$.

We now return to our full model with asymmetric agents who can have any number of types. To define our general notion of an assortative mechanism, 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 a partial order on* Θ *such that* $\theta' \succeq^* \theta$ *if for each agent* i, *either (i)* $\theta'_i = \theta_i$ *or (ii)* θ_i *is a low type and* θ'_i *is a high type. We write* $\theta' \succ^* \theta$ *to indicate* $\theta' \succeq^* \theta$ *and* $\theta' \neq \theta$.

In words, $\theta' \succ^* \theta$ indicates θ' and θ differ only in that some low type agents in state θ become high type agents in state θ'. 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 an **assortative 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 $a' \circ \theta' \geq a \circ \theta$ (across-state assortative),*

(2) *$a' \circ \theta \geq a \circ \theta$, or $a' \circ \theta \leq a \circ \theta$, if $g(\theta)[a'] > 0$ (within-state assortative).*

Across states, an assortative mechanism groups the correct responses to the agents' types to states higher in the $\succeq^*$-order, which leads to a higher payoff for the principal in 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 an assortative mechanism ever correctly responds to agent i's type at θ, then it must correctly respond to agent i's type with probability one in the higher state θ'. Conversely, if an assortative mechanism ever incorrectly responds to agent i's type at θ', then it must incorrectly respond to agent i's type with probability one in the lower state θ.

Within a specific state of the world, assortative mechanisms correlates the correct responses and incorrect responses together in such a way to create a most productive action, a second most productive action, etc., if multiple actions are used (i.e., if the mechanism is stochastic at this state). This is equivalent to saying that any two actions that are played with positive probability can be ordered in terms of the vectors of agent production.

Theorem 2. *There always exists an assortative mechanism that is optimal.*

Proof. See section 3.1. □

Theorem 2 asserts the optimality of assortative mechanisms. The underlying logic is the same as in the two-type symmetric case. Complementarities reward grouping high-production outcomes together. Both the state and the action influence each agent's production, and the mechanism is assortative across states since it groups the correct responses to the agents' types to states higher in the $\succeq^*$-order. Incorrect responses, which are required for truthful revelation, are placed at states lower in the $\succeq^*$-order where they do the least damage. Within a single state, assortative mechanisms leverage 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 *an* optimal mechanism is assortative, but not that *every* optimal mechanism is. The proof of Theorem 2 is detailed in Section 3.1 below. Despite being discussed first, Theorem 1 is proved later 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.

If the marginal distributions with which the principal took action 1 for each agent were fixed, our results would closely relate to the theory of comparative statics. Following the statistics literature, Meyer and Strulovici (2012) define a supermodular order on distributions according to their expectation given a supermodular objective function. The elementary transformations which they use to characterize this order (Meyer and Strulovici 2012, theorem 1) are also exploited by assortative mechanisms. These elementary transformations leave the marginal distributions unchanged. In our screening environment, the principal then has a secondary problem of also choosing the marginals, which can lead to counter-intuitive results (see Section 3).

The class of coordination mechanisms is potentially large and one may wonder whether tighter results are possible. For example, Theorem 1 does not explicitly characterize the cutoffs m^H and m^L, which leaves room for a number of possibilities. Without making further assumptions on the principal's payoff function, we cannot say more. However, if complementarities are strong enough, then the optimal mechanism will be "completely assortative" 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 a completely assortative mechanism. More generally, a complete characterization of the optimal mechanism in this environment requires further assumptions on the principal's payoff function. In particular, we would need to know how a principal compares certain lotteries above and beyond the comparisons implied by supermodularity.

3.1. Proof of Theorem 2

We first establish that there exists an optimal mechanism that obeys part 1 of Definition 3 (across-state assortative). Optimal non-assortative mechanisms can arise because the principal is indifferent about the agent's type when taking action 0 on that agent. If we perturb the principal's payoff function to be strictly increasing in the number of high types and W is strictly supermodular, then every optimal mechanism would be an assortative mechanism. Proposition 1 proves this result, which is also a key step in the proof of Theorem 2.

For any $\theta \in \Theta$, let h_θ equal the number of high-type agents at θ. For any $\varepsilon \geq 0$, define a perturbed version of the principal's payoff $\widetilde{W}_\varepsilon : \Theta \times A \to \mathbb{R}$ as $\widetilde{W}_\varepsilon(\theta, a) := (h_\theta)^\varepsilon W(a \circ \theta)$. At $\varepsilon = 0$ we have $\widetilde{W}_\varepsilon = W$.[6] The ε-perturbed principal's design problem can be written as

$$\max_{g \in \mathcal{G}} \sum \mu(\theta) g(\theta)[a] \widetilde{W}_\varepsilon(\theta, a),$$

such that for every i and any $\theta_i, \theta_i' \in \Theta_i$, $G_i(\theta_i) = G_i(\theta_i')$.

The constraint set of this perturbed maximization problem does not depend on ε, and the objective function is jointly continuous in ε and g. Hence, the theorem of the maximum applies, and the optimal solution is upper hemicontinuous in ε. Therefore, it suffices to show that every optimal mechanism is across-state assortative for any $\varepsilon > 0$, because that would imply the existence of an optimal mechanism is across-state assortative when $\varepsilon = 0$.

Proposition 1. *For any $\varepsilon > 0$, every optimal mechanism is assortative for the ε-perturbed problem, if W is strictly supermodular.*

Proof. 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 is not across-state assortative (Definition 3, part 1). We will construct a modified mechanism $\hat{g}$ that gives a strictly higher payoff than g and is incentive compatible.

Since g is not across-state assortative, there exist two states θ, θ' and two actions a, a' such that $\theta' \succ^* \theta$, $g(\theta)[a] > 0$, $g(\theta')[a'] > 0$ and $a \circ \theta \not\leq a' \circ \theta'$. Let $I \subseteq \{1, ..., n\}$ be such that $i \in I$ if and only if $\theta_i = \theta_i'$. Take $\hat{a} \in A$ so that for every $i \notin I$ we have $\hat{a}_i = a_i$, and for every $i \in I$ we have $\hat{a}_i = \min\{a_i, a_i'\}$. 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 $\hat{a}_i' = \max\{a_i, a_i'\}$. For all $i \notin I$, $\theta' \succ^* \theta$ implies that θ_i' is a high type and θ_i is a low type, which means we know $a_i'\theta_i' \geq a_i\theta_i$ regardless of a_i and a_i'. Therefore we know that

$$a \circ \theta \vee a' \circ \theta' = \hat{a}' \circ \theta' \text{ and } a \circ \theta \wedge a' \circ \theta' = \hat{a} \circ \theta. \tag{1}$$

This implies $\hat{a}' \circ \theta' \geq a' \circ \theta'$ and $\hat{a}' \circ \theta' \neq a' \circ \theta'$. Therefore, by the strict monotonicity of W, we have

$$W(\hat{a}' \circ \theta') > W(a' \circ \theta'). \tag{2}$$

[6] We use the convention that $0^0 = 1$.

Let $\eta > 0$ be small and for any $\theta^* \in \Theta$ and $a^* \in A$, define

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

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

Comparing the principal's expected payoffs from mechanisms $\hat{g}$ and g, we have

$$\begin{aligned} \widetilde{W}_\varepsilon(\hat{g}) - \widetilde{W}_\varepsilon(g) &= \mu(\theta)\frac{\eta}{\mu(\theta)}\left(\widetilde{W}_\varepsilon(\theta, \hat{a}) - \widetilde{W}_\varepsilon(\theta, a)\right) + \mu(\theta')\frac{\eta}{\mu(\theta')}\left(\widetilde{W}_\varepsilon(\theta', \hat{a}') - \widetilde{W}_\varepsilon(\theta', a')\right) \\ &> \eta(h_\theta)^\varepsilon \left\{W(\hat{a} \circ \theta) - W(a \circ \theta) + W(\hat{a}' \circ \theta') - W(a' \circ \theta')\right\} \geq 0. \end{aligned} \tag{3}$$

The strict inequality follows from the fact that $h_{\theta'} > h_\theta$ and Equation (2). The weak inequality follows from Equation (1) and the strict supermodularity of W. All that remains is to establish that $\hat{g}$ is incentive-compatible. Since g is incentive compatible, it suffices to prove that $\widehat{G}_i(\theta_i^*) = G_i(\theta_i^*)$, for all i and $\theta_i^* \in \Theta_i$.

For any $i \notin I$, this is immediate since, by construction, $\hat{g}_i(\cdot) = g_i(\cdot)$. For $i \in I$, $\hat{g}_i(\cdot)$ differs from $g_i(\cdot)$ at θ and θ' if and only if a_i' was an incorrect response to θ_i' and a_i was a correct response to θ_i. Suppose θ_i is a high type; the case where θ_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 $a_i = \hat{a}_i' = 1$. So, $\widehat{G}_i(\theta_i) - G_i(\theta_i) = \frac{\eta}{\mu(\theta')}\mu_{-i}(\theta'_{-i}) - \frac{\eta}{\mu(\theta)}\mu_{-i}(\theta_{-i}) = 0$. Hence, $\hat{g}$ is incentive-compatible and we have a contradiction.

The second part of Definition 3 is also satisfied, as proven in Lemma 1. □

The proof of Theorem 2 and Proposition 1 is completed by the following lemma.

Lemma 1. *Every optimal mechanism is within-state assortative, if W is strictly supermodular.*

Proof. Let g^* be an optimal mechanism, and assume by way of contradiction that it is not within-state assortative. Then there exists $\theta \in \Theta$ and $a, a' \in A$ such that $g^*(\theta)[a] > 0, g^*(\theta)[a'] > 0$ and $a \circ \theta \not\geq a' \circ \theta$ and $a \circ \theta \not\leq a' \circ \theta$. Now construct $\hat{a}, \hat{a}' \in A$ as follows

$$\hat{a}_i = \begin{cases} a_i & \text{if } a_i\theta_i \geq a_i'\theta_i \\ a_i' & \text{otherwise} \end{cases} \quad \text{and } \hat{a}_i' = \begin{cases} a_i & \text{if } a_i\theta_i < a_i'\theta_i \\ a_i' & \text{otherwise} \end{cases}.$$

By construction we have $\hat{a} \circ \theta = a \circ \theta \vee a' \circ \theta$ and $\hat{a}' \circ \theta = a \circ \theta \wedge a' \circ \theta$. Let $\varepsilon > 0$ to be smaller than both $g^*(\theta)[a]$ and $g^*(\theta)[a']$. Define

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

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 $G_i(\tilde{\theta}_i) = G_i^*(\tilde{\theta}_i)$. Therefore by Fact 1, g is incentive compatible.

Let $W(g^*)$ and $W(g)$ be the principal's expected payoffs from g^* and g, respectively. Then:

$$W(g^*) - W(g) = \varepsilon \left\{ W(a \circ \theta) + W\left(a' \circ \theta\right) - W(\hat{a} \circ \theta) - W\left(\hat{a}' \circ \theta\right) \right\} < 0.$$

The strict inequality follows from the strict supermodularity of W. This contradicts the optimality of g^*. Note that with weak supermodularity, the above would also be a weak inequality and there would always exist an optimal mechanism which is within-state assortative. □

Note that the statements of Proposition 1 and Lemma 1 assume strict supermodularity of W. This ensures that the *every* optimal mechanism is assortative. If we had weak supermodularity the only change would be that the inequality in Equation 3 would be weak. This means that under weak supermodularity, *there exists* an optimal assortative mechanism. The continuity argument above holds as stated.

3.2. Strategic Favoritism and Meritocracy

One novel feature in this mechanism design problem, relative to the theory of comparative statics and related persuasion problems, is that the marginal probabilities of taking action 1 on each agent are not fixed. They are a choice object which is also determined by the principal's supermodular objective function. As a result, optimal mechanisms can display **strategic favoritism** by taking action 1 more often on a less productive agent. By "more productive" we mean an unambiguously better distribution of type quality, while assuming the agents are symmetric in all other ways. Hence, favoritism involves rewarding an agent who is worse for the principal's own aims.

Rather than giving the most general treatment, we demonstrate these phenomena in a parametric example with two agents and two possible types: a high type (H) and a low type (L) with $H = 1$ and $L = -1$. We will consider the following principal payoff function

$$W(a_1\theta_1, a_2\theta_2) = \left[(a_1\theta_1 + 2)(a_2\theta_2 + 2)\right]^\kappa.$$

It is straightforward to see that the above is supermodular and that complementarities increase with the parameter $\kappa \geq 1$. The function is also symmetric, in the sense that permuting the arguments of W leaves payoffs unchanged. Fix $\mu_1(H) = 0.5$ and let $\mu_2(H) = \mu_2 > 0.5$ vary, so that agent 2 is unambiguously better for the principal.

Recall that we use notation HH to denote state (H, H) and 01 to indicate action vector $(0, 1)$. For any $\kappa \geq 1$ the solution to the principal's problem has the feature that at states HH and LL the principal chooses action vector 11.[7] This maximizes the principal's payoff in the best possible state which is extremely valuable. However, taking action 11 in state HH, means that action 1 must be taken for both agents with at least probability one-half. For this to be incentive compatible, each agent must receive action 1 with the same probability when they report "low". The principal therefore needs to make some mistakes, which are the least costly in state LL (because her gain from responding correctly to a low-type agent is lower than doing so in any other state).

The next lemma characterizes the two possible action vectors the principal chooses in states HL and LH.

Lemma 2. *There is always an optimal mechanism such that* $g(LH)[01] = (1-\mu_2)/\mu_2$ *and either*

(i) $g(HL)[01] = 1$ *and* $g(LH)[11] = (2\mu_2 - 1)/\mu_2$, *or*

(ii) $g(HL)[01] = (1-\mu_2)/\mu_2$, $g(HL)[00] = (2\mu_2 - 1)/\mu_2$ *and* $g(LH)[10] = (2\mu_2 - 1)/\mu_2$.

Furthermore, mechanism (i) is better than (ii) if $(\mu_2)3^\kappa > (1-\mu_2)(4^\kappa - 2^\kappa)$.

Proof. See Appendix. □

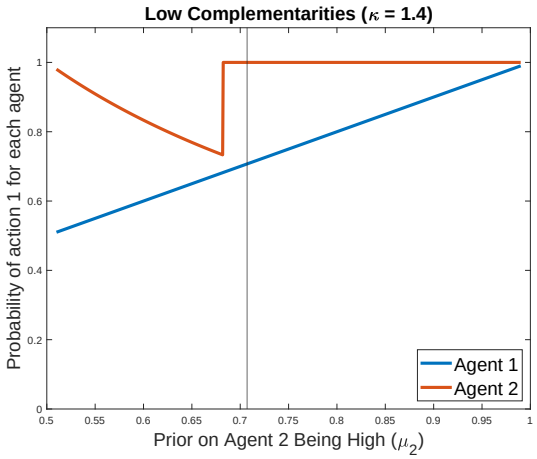
The conditional expected payoffs computed in the proof imply that condition 1 of lemma 2 gives the principal a higher payoff than condition 2 when $(\mu_2)3^\kappa > (1-\mu_2)(4^\kappa - 2^\kappa)$, that is, when κ is low or μ_2 is sufficiently close to 1.

We now turn to how frequently the principal optimally takes action 1 on each agent. A mechanism that satisfies condition 1 of lemma 2, $a_1 = 1$ is chosen with probability μ_2, while $a_2 = 1$ with probability 1. Thus, whenever condition 1 is optimal, the principal takes action 1 on the better agent more frequently; a regime we term **meritocratic**.

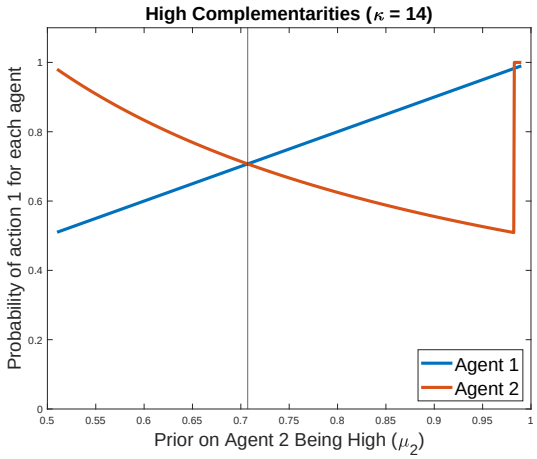
Under a mechanism which satisfies condition 2 however, $a_1 = 1$ is chosen with probability μ_2, but $a_2 = 1$ is chosen with probability $1/(2\mu_2)$. If $\mu_2 < 1/\sqrt{2}$ the principal takes action 1 on the

[7]This is not true when μ_2 is sufficiently close to 0 and κ close to 1 (in this case always choosing action 0 for agent 2 is best). We focus on the case where $\mu_2 > 1/2$ as this provides greatest insight, while minimizing the number of cases to consider.

better agent (agent 2) more often and we again have meritocracy. However when $\mu_2 > 1/\sqrt{2}$ then the worse agent, agent 1, gets action 1 more often; a regime we call strategic favoritism.



Panel A: Lower Complementarities



Panel B: Higher Complementarities

The above figure plots the probability of taking action 1 on each agent as a function of μ_2 for different choices of κ. When κ is low, as illustrated in Panel A, the principal will prefer a mechanism which satisfies condition 1 of lemma 2 (where agent 2 is chosen with probability 1) before μ_2 reaches the cutoff value $1/\sqrt{2}$ and thus only meritocracy is possible. When κ is high, complementarities are larger, the principal prefers a mechanism which satisfies condition 2 of lemma 2 for a larger set of μ_2 and we have a large region of strategic favoritism, as Panel B shows. The magnitude of favoritism is the difference in the probability of taking action 1 on the worse agent as opposed to the better agent, and can be quite large. For example, when $\kappa = 14$ and $\mu_2 = 0.98$, agent 1, the worse agent, who is the high type with probability 0.5, gets $a_1 = 1$ with 98% of the time, while agent 2 gets $a_2 = 1$ with probability 51%. The magnitude of favoritism tends to 1/2 as $\mu_2 \to 1$, and for any $\mu_2 < 1$ if κ is large enough this can indeed be optimal.

To gain an intuition into strategic favoritism, let's first think about why the principal faces a trade-off between these two extreme points in the first place. Since she is taking action 11 in state HH, this is very likely to be realized for agent 1 if he reports "high", as agent 2 is most probably a high type. This doesn't mean that agent 1 will be get action 1 more often, since, if agent 2 is very likely to be good, the principal could take action 1 on him with probability 1. That is exactly what condition 1 of lemma 2 implies.

To see exactly how condition 2 of lemma 2 can be better for the principal, consider the differences in actions taken. In state HL, rather than the worst possible payoff being realized with probability 1 (the principal incorrectly responds to both agents), the principal now takes action 00 with positive probability and correctly responds to agent 2 when he is the low type, while incorrectly responding to agent 1. In state LH, a mechanism satisfying condition 1 correctly responds to agent 2 with some probability when he is the high type, while incorrectly responding to agent 1. This effectively presents a choice to the principal between a lottery which correctly responds to agent 2 when is the low type, as opposed to one that correctly responds when he is the high type. Holding μ_2 fixed, so that the agents' incentives are not affected, as κ increases condition 2 of lemma 2 becomes increasingly attractive as correctly responding when agent 2 is the low type carries a 4^κ term, instead of a 3^κ term.

4. Robustness

We consider two robustness exercises. In section 4.1, we relax the assumption that the agents draw their types independently and show that assortative mechanisms, which satisfy all incentive constraints strictly, are almost optimal. In section 4.2, we show how these mechanisms can be made robust to coalitions of players agreeing on a reporting strategy.

4.1. Robustness of Assortative Mechanisms

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 an assortative 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 an assortative mechanism.

Our baseline model with independent priors has an unusual feature: everywhere binding incentive constraints. This is because agent preferences are type-independent. Relaxing the independence of priors assumption leads to different types of the same agent having different incentives, due to differing beliefs about the type distributions, and hence reports, of 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.

The second result in this section shows that the optimal mechanism with independent types is strictly incentive-compatible and almost optimal on an open set of non-independent priors.

In this section, we let the state space $\Theta = \{L, H\}^n$, so that each agent has two types. The general space of prior beliefs is $\Upsilon = \Delta\Theta$. Let $\Upsilon_I \subset \Upsilon$ be the space of independent priors. We write $\mu(\theta_{-i})$ for the overall probability of θ_{-i} and $\mu(\theta_{-i}|t)$ for the conditional probability of θ_{-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 μ on each individual type of each agent.

Definition 4. *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 Υ_p be the space of all priors with positive dependence.

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).[8] Therefore, positive dependence requires that adding correlation makes higher θ_{-i} more likely when agent i is a high type and less likely when agent i is a low type.

Theorem 3. *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 an assortative mechanism g and a sequence of mechanisms $\{g^m\}_{m=1}^\infty$, such that g is optimal at μ, g^m is optimal at μ^m and $g^m \to g$.*

Proof. See Appendix. □

A similar argument would prove that optimal mechanisms move upper hemicontinuously at any independent prior in the space Υ_I. However, upper hemicontinuity by itself does not guarantee that optimal mechanisms with small amounts of positive dependence are close to an assortative mechanism. This gap occurs because Theorem 2 only established an optimal mechanism is assortative and not that every optimal mechanism is assortative. Theorem 3 explicitly addresses this gap by showing that, with small a amount of positive dependence, there exists an optimal mechanism that is close to an assortative mechanism. Even though they are close to an assortative mechanism, some, but by no means all, incentive constraints bind.

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

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. Let $P_i^g = E_\theta [G_i(\theta)]$ be the overall probability that g sets $a_i = 1$. We say that a mechanism g is trivial if there exists an i such that either $P_i^g = 1$ or $P_i^g = 0$, that is a mechanism is trivial if there exists an agent who either always or never receives action 1.

Theorem 4. *Take any $\mu \in \Upsilon_I$ and $\varepsilon > 0$; let g be any optimal assortative mechanism at μ. If g is nontrivial, then there exists an open set of priors Υ_O, with μ on the boundary of Υ_O, such that for any $\nu \in \Upsilon_O$, g is strictly incentive-compatible and furthermore*

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

Proof. See Appendix. □

Theorem 4 shows that an assortative mechanism optimal at independent prior μ becomes strictly incentive-compatible and almost optimal for some open set of nearby priors Υ_O. Strict incentives makes the mechanism robustly incentive-compatible even if we allow agents to make small mistakes in their reports. The fact that μ is not in the set Υ_O follows from the fact that incentive-compatible mechanisms can never provide strict incentives at an independent prior. The best result we could hope for is that μ is on the boundary of Υ_O, which is what Theorem 4 proves. Observe that if g was a trivial mechanism, the theorem above would still hold if we remove any agents who either always or never receive action 1 and consider the mechanism induced on the remaining agents.

4.2. Coalition-Proofness

A potential concern with the optimal direct mechanism is that agents may be able to collude.[9] For example, in the assortative mechanism discussed in Section 2.1, the agents could always get action 1 if they both agree to report H. This section shows how we can amend the optimal direct mechanism to make it robust to collusion with virtually the same payoff to the principal.

The coalition-proof mechanism we propose is inspired by techniques from the virtual implementation literature (e.g., Abreu and Matsushima (1992)), where with arbitrarily high probability, the optimal direct mechanism is implemented. Our assumption that the agents' preferences are type-independent, means that our setting fails to satisfy Abreu-Matsushima measurability,[10] and those

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

[10]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.

results cannot be directly applied. We add ideas from the classic implementation literature, notably integer games (Maskin (1999)), to get around this. Our results also differ from these literatures since we focus on immunity from coalitional 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 is not 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 θ_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 5. *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 σ_i *such that the following three conditions hold:*

(1) *For every* $i \notin I$, σ_i *is a truthful reporting strategy.*
(2) *For all* $i \in I$, σ_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.*

A mechanism is **coalition-proof** if there are no any valid coalitions.

Theorem 5. *For any* $\varepsilon \in (0, 1)$, *there exists a mechanism* g^ε *which is coalition proof. Furthermore, as* $\varepsilon \to 0$, g^ε *converges to the optimal assortative mechanism in both payoffs and outcomes.*

Proof. See Appendix. □

The idea behind coalition proof mechanisms is that with a large probability they implements the optimal direct mechanism, which only relies on each agent's report about his own type. However, with the remaining probability these mechanisms ask agents to play a betting game that rewards agents for correctly guessing other agents' reports.[11] However, only the bet of the agent who reports the highest integer 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 are calibrated so

[11]Because we have no transfers, agents are rewarded by the principal taking action 1 on them with some probability. Higher rewards correspond to higher probability.

that agents break even on every possible bet when everyone is reporting truthfully. When collusion occurs, agents in the deviating coalition can make a strict gain from betting, and hence all agents will desire to report $z_i > 0$, which destroys possible collusion.

5. Conclusion

We study a screening problem with type-independent preferences. While natural in a number of economic environments, they pose a challenge to standard screening techniques as the principal's marginal action on each agent cannot depend on that agent's reported type. However, a principal with a supermodular objective function cares about the joint distribution of actions and types, not just the marginal, and can benefit from an assortative mechanism. Such mechanisms correctly respond to agents at states with many high types and incorrectly respond at states with many low types. Assortative mechanisms allow the principal to accept the failure necessary for incentive compatibility when the payoff consequences are 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 demonstrate that optimal mechanisms can display strategic favoritism by rewarding an inferior agent with a better marginal action. Such behavior is sometimes attributed to biases, but we show it can arise from purely strategic considerations and is optimal for an unbiased principal.

The basic ideas apply beyond the supermodularity assumption: whenever the principal's utility function is not additively separable, the principal faces a non-trivial mechanism design problem. The principal is able to exploit the fact that agents have thick indifference curves relative to the principal. We focused on supermodularity because of potential applications and to relate to the comparative statics literature. Generalizing our results to a other non-separable environments 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 assortative mechanisms would still operate even if the preferences of agents did depend on their type. Determining how this 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.

Arieli, I. and Y. Babichenko (2019). Private bayesian persuasion. *Journal of Economic Theory 182*, 185–217.

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.

Levy, G., I. Moreno de Barreda, and R. Razin (2021). Persuasion with correlation neglect: A full manipulation result. *Working Paper*.

Levy, G. and R. Razin (2015). Correlation neglect, voting behavior, and information aggregation. *American Economic Review 105*(4), 1634–45.

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.

Meyer, M. and B. Strulovici (2012). Increasing interdependence of multivariate distributions. *Journal of Economic Theory 147*(4), 1460–1489.

Meyer, M. and B. Strulovici (2017). Beyond correlation: Measuring interdependence through complementarities. *Working Paper*.

Müller, A. and M. Scarsini (2000). Some remarks on the supermodular order. *Journal of Multivariate Analysis 73*(1), 107–119.

Ortoleva, P. and E. Snowberg (2015). Overconfidence in political behavior. *American Economic Review 105*(2), 504–35.

Radner, R. (1981). Monitoring cooperative agreements in a repeated principal-agent relationship. *Econometrica: Journal of the Econometric Society*, 1127–1148.

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.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *The Annals of Probability*, 814–827.

A. Proofs From Section 3

A.1. Proof of Theorem 1

We start by formally defining an anonymous mechanism. For any bijection $\sigma : \{1, ..., n\} \to \{1, ..., n\}$ we let $\sigma(\theta) = \left(\theta_{\sigma(1)}, \theta_{\sigma(2)}, ..., \theta_{\sigma(n)}\right) \in \Theta$ and $\sigma(a) = \left(a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)}\right) \in A$.

Definition 6. *We say a mechanism g is anonymous if for any bijection $\sigma : \{1, ..., n\} \to \{1, ..., n\}$, and $g(\theta)[a] = g(\sigma(\theta))[\sigma(a)]$, for all $\theta \in \Theta$, $a \in A$.*

We claim there always exists an anonymous assortative mechanism that is optimal. To prove this take an ε-perturbation of the principal's payoff function as in the proof of Theorem 2. This perturbation maintains the symmetry of the payoff function. Let g be an optimal mechanism in this perturbed setting. Let Σ be the set of all bijections from $\{1, ..., n\} \to \{1, ..., n\}$. For each $\sigma \in \Sigma$, define mechanism g^σ by $g^\sigma(\theta)[a] = g(\sigma(\theta))[\sigma(a)]$. By the symmetry of the setting, g^σ is optimal for every $\sigma \in \Sigma$ because it has the same incentive compatibility properties and principal payoff as g. Now define g^*_ε as

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

The mechanism g^*_ε is a mixture of optimal mechanisms and is therefore optimal itself. By construction g^*_ε is anonymous. Using the same proof as in Theorem 2, we can show every optimal mechanisms in the ε-perturbed setting with strict supermodularity is an assortative mechanism, including g^*_ε. Take ε to zero while constructing a sequence of optimal, 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 assortative.

Let g be an optimal, anonymous assortative mechanism. The remainder of this proof involves showing that g has all the properties stated in Theorem 1.

Part 2 of Definition 3 (within-state assortative) 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 θ, 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 ($\overrightarrow{\mathbf{0}}$).

Recall that h_θ denotes the number of high types at θ. Using what we established in the previous paragraph along with the definition of an anonymous assortative 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 $h_\theta - 1 = m^L$ 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 the definition of m^L and m^H, 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 θ 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. □

We further 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 $\overrightarrow{\mathbf{1}}_{\theta'_i=H} = \overrightarrow{\mathbf{1}}$ 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 not both be used with positive probability, since the mechanism wouldn't be within-state assortative. 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.2. Completely Assortative Mechanisms

As mentioned in the text of Section 3, we provide a characterization of completely assortative mechanisms in the two-type symmetric environment. In these mechanisms the principal either correctly responds to every agent or incorrectly responds to every agent. Any completely assortative mechanisms can be fully described by two values, as given in the following definition.

Definition 7. *For $m \in \{0, 1, ..., n\}$ and $q \in [0, 1)$, the **completely assortative 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 probability $1-q$ and q respectively, at states with exactly m high types.*

From Fact 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

$$\sum_{k=m}^{n-1} \mathcal{H}(k) + (1-q)\,\mathcal{H}(m-1) = \sum_{k=0}^{m-1} \mathcal{H}(k) + q\mathcal{H}(m), \tag{4}$$

where $\mathcal{H}(k) := \binom{n-1}{k} p^k (1-p)^{n-1-k}$, for $k \in \{0, ..., n\}$.[12] Fixing any agent, $\mathcal{H}(k)$ is the probability that exactly k of the $n-1$ other agents are type H.

The left-hand side and right-hand side of Equation (4) 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 (4). In other words, for a fixed two-type-symmetric environment, there is exactly one completely assortative mechanism that is incentive-compatible. The following proposition characterizes when that mechanism is optimal. Recall that we let the indicator functions $\overrightarrow{\mathbf{1}}_{\theta_i=H}$ and $\overrightarrow{\mathbf{1}}_{\theta_i=L}$, respectively, denote correctly responding to every agent and incorrectly responding to every agent, respectively.

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

Proposition 2. *Assume a two-type symmetric environment, and let (m^*, q^*) be the unique solution to Equation (4). 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 and is proved as a corollary to Theorem 1.

A.3. Proof of Lemma 2

In states HL and LH the principal's choices are constrained by incentive compatibility. The incentive compatibility constraint for agent 1, $G_1(H) = G_1(L)$, implies $\mu_2 g_1(HH)+(1-\mu_2)\,g_1(HL) = \mu_2 g_1(LH)+(1-\mu_2)\,g_1(LL)$. For agent 2, incentive compatibility implies $g_2(HL) = g_2(LH)$. Using the fact that $g_1(HH) = g_1(LL) = 1$, we can write these constraints in terms of $g(\theta)[a]$, the probability of playing action a in state θ, as follows

$$\begin{aligned} g(HL)[11] + g(HL)[10] + \frac{2\mu_2-1}{1-\mu_2} &= \frac{\mu_2}{1-\mu_2}\left\{g(LH)[11] + g(LH)[10]\right\}, \text{ and} \\ g(HL)[11] + g(HL)[01] &= g(LH)[11] + g(LH)[01]. \end{aligned}$$

Furthermore, by Theorem 2 we know that the optimal mechanism must be within-state assortative. This implies that $g_1(HL)[11] = 0$ or $g_1(HL)[00] = 0$ (or both), and similarly for state LH, which further restricts the number of cases.

We can begin to characterize the extreme points of this constraint set, using all of the above. While listing the extreme points, we also compute the expected payoff for the principal, conditional on states HL or LH being realized. This will allow us to focus on the extreme points which maximize the principal's payoffs.

(1) $g(HL)[11] = 1$ and $g(LH)[11] = 1$. Conditional expected payoff: $(1-\mu_2)\,3^\kappa + (\mu_2)\,3^\kappa = 3^\kappa$

(2) $g(HL)[10] = 1$ and $g(LH)[10] = 1$. Conditional expected payoff: $(1-\mu_2)\,6^\kappa + (\mu_2)\,2^\kappa$

(3) $g(HL)[01] = 1$ and $g(LH)[11] = (2\mu_2-1)/\mu_2$, $g(LH)[01] = (1-\mu_2)/\mu_2$. Conditional expected payoff: $(1-\mu_2)\,2^\kappa + (2\mu_2-1)\,3^\kappa + (1-\mu_2)6^\kappa$

When $\mu_2 = 1$, the principal is indifferent between extreme points 1 and 3, but otherwise strictly prefers extreme point 3 to the other two, for all $\kappa \geq 1$.

(4) $g(HL)[00] = 1$, and $g(LH)[10] = (2\mu_2 - 1)/\mu_2$, $g(LH)[00] = (1-\mu_2)/\mu_2$. Conditional expected payoff: $2(1-\mu_2)4^\kappa + (2\mu_2 - 1)2^\kappa$

(5) $g(HL)[00] = (2\mu_2 - 1)/\mu_2$, $g(HL)[01] = (1-\mu_2)/\mu_2$ and $g(LH)[10] = (2\mu_2 - 1)/\mu_2$, $g(LH)[01] = (1-\mu_2)/\mu_2$. Conditional expected payoff: $(2\mu_2 - 1)(1-\mu_2)/\mu_2 4^\kappa + (1-\mu_2)^2/\mu_2 2^\kappa + (2\mu_2 - 1)2^\kappa + (1-\mu_2)6^\kappa$.

Comparing the two extreme points above, we see that 5 is better for all $\kappa \geq 1$ and $\mu_2 > 1/2$ (as the difference between expected payoffs is $(1-\mu_2)/\mu_2 (2^\kappa - 4^\kappa - 2^\kappa\mu_2 + 6^\kappa\mu_2)$, which is minimized at $\mu_2 = 1/2$ and $\kappa = 1$).[13]

Extreme points 3 and 5 are the two possible solutions for the principal's problem. Comparing the conditional expected payoffs leads to the condition that extreme point 3 is better than 5 if $(\mu_2)3^\kappa > (1-\mu_2)(4^\kappa - 2^\kappa)$.

B. Proofs from Section 4

B.1. Proof of Theorem 3

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

Lemma 3. *Let $\mu \in \Upsilon_I$ and $\mu' \in \Upsilon_p$ such that $\mu = I^{\mu'}$. Then any any assortative mechanism that is incentive compatible at μ is incentive compatible at μ'.*

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, W^ε, 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 μ^m under payoff function W^ε. 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 be upper hemicontinuous with respect to the prior. Hence, g^ε is incentive compatible at μ. Now let $U^{m,\varepsilon}$ and U^ε be the optimal principal payoff at μ^m and μ respectively under W^ε. The payoff from g^ε equals $\lim_{m\to\infty} U^{m,\varepsilon}$. We will show g^ε is optimal at μ under W^ε, i.e.,

$$\lim_{m\to\infty} U^{m,\varepsilon} = U^\varepsilon. \tag{5}$$

[13]Note that when $\mu_2 = 1/2$, the incentive constraint of agent 1 loses the constant term and now $g(H,L)[0,0] = 1$, $g(L,H)[0,0] = 1$ are incentive compatible. This was optimal in the example presented in Section 2.1. The parameter choices in this example were made to minimize the number of cases to consider.

Since g^ε is incentive compatible at μ, it is immediate that $\lim_{m\to\infty} U^{m,\varepsilon} \leq U^\varepsilon$. For each m and ε, let $\widetilde{g}^{m,\varepsilon}$ be an optimal mechanism at prior I^{μ^m} under W^ε with associated payoff $U^{m,\varepsilon,I}$. We can take $\widetilde{g}^{m,\varepsilon}$ to be an assortative mechanism by Theorem 2. Therefore by Lemma 3, $\widetilde{g}^{m,\varepsilon}$ is incentive compatible at μ^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 $\widetilde{g}^{m,\varepsilon}$ converges since $\mathcal{G}$ is compact, and let $\widetilde{g}^\varepsilon = \lim_{m\to\infty} \widetilde{g}^{m,\varepsilon}$. And the fact that $\mu^m \to \mu$ and $\mu \in \Upsilon_I$ implies $I^{\mu^m} \to \mu$. Since the optimal mechanism moves upper hemicontinuously in the domain Υ_I,[14] which implies $\widetilde{g}^\varepsilon$ is optimal at μ under W^ε, and therefore $\lim_{m\to\infty} U^{m,\varepsilon} \geq \lim_{m\to\infty} U^{m,\varepsilon,I} = U^\varepsilon$.

Now that we've shown Equation (5), which implies that g^ε is optimal at μ under W^ε. Note that g^ε is an assortative mechanism by the proof of Theorem 2. 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 ε, and, hence, g is an optimal assortative mechanism under W at μ and g^m is an optimal mechanism under W at μ^m for each m. Moreover, by exchanging the order of the limits we get $\lim_{m\to\infty} g^m = g$, as desired.

B.1.1. Proof of Lemma 3

Let $\mu \in \Upsilon_I$ and $\mu' \in \Upsilon_p$ such that $\mu = I^{\mu'}$. Let g be an assortative mechanism at μ. We show g is incentive compatible at μ'. Mechanism g is incentive compatible for agent i at μ' 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 μ which implies

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

Hence it suffices 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 an

[14] A proof of this straightforward fact appeared in an earlier working paper version.

assortative mechanism, $g_i(L, \theta_{-i})$ is decreasing in θ_{-i} and $g_i(H, \theta_{-i})$ is increasing in θ_{-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).

B.2. Proof of Theorem 4

Fix any $\mu \in \Upsilon_I$ and let g to be an optimal assortative mechanism at μ. Let θ^H be the state where every agent is type H, and let θ^L be the state where every agent is type L. Let Θ^{n-1} be the set of n states in which exactly $n-1$ agents are type H. Let Θ^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 μ 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 ν^ε for all $\varepsilon \in (0, \bar{\varepsilon})$. Clearly μ is on the boundary of Υ_O and $I^v = \mu$ for any $v \in \Upsilon$.

Since g is incentive compatible at μ, 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 $\theta_i' \neq \theta_i$

$$\begin{aligned} \sum_{\theta_{-i}} g_i(\theta_i, \theta_{-i}) \nu(\theta_i, \theta_{-i}|\theta_i) &\geq \sum_{\theta_{-i}} g_i(\theta_i, \theta_{-i}) \mu(\theta_i, \theta_{-i}|\theta_i), \text{ and} \\ \sum_{\theta_{-i}} g_i(\theta_i', \theta_{-i}) \mu(\theta_i, \theta_{-i}|\theta_i) &\geq \sum_{\theta_{-i}} g_i(\theta_i', \theta_{-i}) \nu(\theta_i, \theta_{-i}|\theta_i), \end{aligned}$$

with at least one of the above being strict.

We only consider $\theta_i = H$; the case $\theta_i = L$ is similar. The first inequality holds strictly if

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

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

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

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

This inequality holds due to the fact that g is a non-trivial assortative mechanism. In particular, g is an assortative 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 an assortative mechanism. By definition, $\Upsilon_O \subseteq \Upsilon_p$, which implies, using an argument similar to the proof of Theorem 3, that there exists $\Upsilon'_O \subset \Upsilon_O$ such that for any $v \in \Upsilon'_O$, $\max_{g' \in \mathcal{G}} V\left(g' | \nu\right) < V\left(g | \nu\right) + \varepsilon$.

B.3. Proof of Theorem 5

We first formally define our coalition-proof mechanism. For $\varepsilon > 0$, 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. Let $T_i \in \Theta$ and $T_{ij} \in \Theta_j$ denote agent i's report of the type profile and agent j's type, respectively. Let z_i denote the integer reported by agent i and $diag\left(T\right) = \left(T_{11}, T_{22}, ..., T_{nn}\right)$ be the diagonal vector of each agent's report about his own type.

With probability $1 - \varepsilon$, g^ε plays the optimal direct mechanism using reports $diag\left(T\right)$, and with probability ε, g^ε 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 (T, z). 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^ε is coalition-proof for all $\varepsilon > 0$. Fix $\varepsilon \in (0,1)$, and suppose for contradiction that $(\{\sigma_i\}_{i=1}^n, I)$ is a valid coalition deviation for g^ε. For any $t_i \in \Theta_i$, let $\rho_i(t_i)$ denote the probability that agent i reports his own type to be t under strategy σ_i. For any $t_{-i} \in \Theta_{-i}$, we will define $\rho_{-i}(t_{-i}) := \prod_{j \neq i} \rho_j(t_j)$. For each agent i, let $D_i := \max_{\theta_i \in \Theta_i} (\rho_i(\theta_i) - \mu_i(\theta_i))$. 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 take $i \in S$. Otherwise let $i \in I$. In either case, for any $j \neq i$ we have $D_j = 0$ which implies $\rho_j(\theta_j) = \mu_j(\theta_j)$ for all $\theta_j \in \Theta_j$. Since types are independent, for any possible strategy i could employ he receives the same payoff as when all other agents report their types truthfully. Because the optimal direct mechanism is incentive compatible, truthfully reporting his own type is a best response for agent i. Moreover, if agent i reports $z_i > 0$, 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 $(\{\sigma_i\}_{i=1}^n, I)$ 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^ε is coalition proof for all $\varepsilon \in (0, 1)$.

Finally, observe that as ε approaches zero, mechanism g^ε implements the optimal direct mechanism with probability 1, and, because the principal's payoff is bounded, the payoff of g^ε converges to the optimal payoff.