Optimal Allocation with Costly Verification[1]

Elchanan Ben-Porath[2] Eddie Dekel[3] Barton L. Lipman[4]

First Draft
August 2012

[1]We thank Ricky Vohra and numerous seminar audiences for helpful comments. We also thank the National Science Foundation, grants SES–0820333 (Dekel) and SES–0851590 (Lipman), and the US–Israel Binational Science Foundation (Ben-Porath and Lipman) for support for this research. Lipman also thanks Microsoft Research New England for their hospitality while this draft was in progress.

[2]Department of Economics and Center for Rationality, Hebrew University. Email: benporat@math.huji.ac.il

[3]Economics Department, Northwestern University, and School of Economics, Tel Aviv University. Email: dekel@northwestern.edu.

[4]Department of Economics, Boston University. Email: blipman@bu.edu.

Abstract

A principal (dean) has an object (job slot) to allocate to one of I agents (departments). Each agent has a strictly positive value for receiving the object. Each agent also has private information which determines the value to the principal of giving the object to him. There are no monetary transfers but the principal can check the value associated with any individual at a cost which may vary across individuals. We characterize the class of optimal Bayesian mechanisms, that is, mechanisms which maximize the expected value to the principal from his assignment of the good minus the costs of checking values. One particularly simple mechanism in this class, which we call the favored–agent mechanism, specifies a threshold value v^* and a favored agent i^*. If all agents other than i^* report values below v^*, then i^* receives the good and no one is checked. Otherwise, whoever reports the highest value is checked with probability 1 and receives the good iff her report is confirmed. We show that *all* optimal mechanisms are essentially randomizations over optimal favored–agent mechanisms.

1 Introduction

Consider the problem of the head of an organization — say, a dean — who has an indivisible resource or good (say, a job slot) that can be allocated to any of several divisions (departments) within the organization (university). Naturally, the dean wishes to allocate this slot to that department which would fill the position in the way which best promotes the interests of the university as a whole. Each department, on the other hand, would like to hire in its own department and puts less, perhaps no, value on hires in other departments. The problem faced by the dean is made more complex by the fact that each department has much more information regarding the availability of promising candidates and the likelihood that these candidates will produce valuable research, teach well, and more generally be of value to the university.

The standard mechanism design approach to this situation would construct a mechanism whereby each department would report its type to the dean. Then the slot would be allocated and various monetary transfers made as a function of these reports. The drawback of this approach is that the monetary transfers between the dean and the departments are assumed to have no efficiency consequences. In reality, the monetary resources each department has is presumably chosen by the dean in order to ensure that the department can achieve certain goals the dean sees as important. To take back such funds as part of an allocation of a job slot would undermine the appropriate allocation of these resources. In other words, such monetary transfers are part of the overall allocation of all resources within the university and hence do have important efficiency consequences. We focus on the admittedly extreme case where no monetary transfers are possible at all.

Of course, without some means to ensure incentive compatibility, the dean cannot extract any information from the departments. In many situations, it is natural to assume that the head of the organization can demand to see documentation which proves that the division or department's claims are correct. Processing such information is costly, though, to the dean and departments and so it is optimal to restrict such information requests to the minimum possible.

Similar problems arise in areas other than organizational economics. For example, governments allocate various goods or subsidies which are intended not for those willing and able to pay the most but for the opposite group. Hence allocation mechanisms based on auctions or similar approaches cannot achieve the government's goal, often leading to the use of mechanisms which rely instead on some form of verification instead.[1]

[1]Banerjee, Hanna, and Mullainathan (2011) give the example of a government that wishes to allocate free hospital beds. Their focus is the possibility that corruption may emerge in such mechanisms where it becomes impossible for the government to entirely exclude willingness to pay from playing a role in the allocation. We do not consider such possibilities here.

As another example, consider the problem of choosing which of a set of job applicants to hire for a job with a predetermined salary. Each applicant wants the job and presents claims about his qualifications for the job. The person in charge of hiring can verify these claims but doing so is costly.

We characterize optimal mechanisms for such settings. We construct an optimal mechanism with a particularly simple structure which we call a *favored–agent mechanism.* There is a threshold value and a favored agent, say i. If each agent other than i reports a value for the good below the threshold, then the good goes to the favored agent and no documentation is required. If some agent other than i reports a value above the threshold, then the agent who reports the highest value is required to document his claims. This agent receives the good iff his claims are verified and the good goes to any other agent otherwise.

In addition, we show that *every* optimal mechanism is essentially a randomization over optimal favored–agent mechanisms. In this sense, we can characterize the full set of optimal mechanisms by focusing entirely on favored–agent mechanisms. By "essentially," we mean that any optimal mechanism has the same reduced form (see Section 2 for definition) as such a randomization up to sets of measure zero. An immediate implication is that if there is a unique optimal favored–agent mechanism, then there is essentially a unique optimal mechanism.

Finally, we give a variety of comparative statics. In particular, we show that an agent is more likely to be the favored agent the higher is the cost of verifying him, the "better" is his distribution of values, and the less risky is his distribution of values. We also show that the mechanism is, in a sense, *almost* a dominant strategy mechanism and consequently is ex post incentive compatible.

Literature review. Townsend (1979) initiated the literature on the principal–agent model with costly state verification. These models differ from what we consider in that they include only one agent and allow monetary transfers. In this sense, one can see our work as extending the costly state verification framework to multiple agents when monetary transfers are not possible. See also Gale and Hellwig (1985), Border and Sobel (1987), and Mookherjee and Png (1989). Our work is also related to Glazer and Rubinstein (2004, 2006), particularly the former which can be interpreted as model of a principal and one agent with limited but costless verification and no monetary transfers. Finally, it is related to the literature on mechanism design and implementation with evidence — see Green and Laffont (1986), Bull and Watson (2007), Deneckere and Severinov (2008), Ben-Porath and Lipman (2011), Kartik and Tercieux (2011), and Sher and Vohra (2011). With the exception of Sher and Vohra, these papers focus more on general issues of mechanism design and implementation in these environments rather than on specific mechanisms and allocation problems. Sher and Vohra do consider a specific allocation question, but it is a bargaining problem between a seller and a buyer, very different from what is considered here.

The remainder of the paper is organized as follows. In the next section, we present the model. Section 3 contains the characterization of the class of optimal compatible mechanisms, showing all optimal mechanisms are essentially randomizations over optimal favored–agent mechanisms. Since these results show that we can restrict attention to favored–agent mechanisms, we turn in Section 4 to characterizing the set of best mechanisms in this class. In Section 5, we give comparative statics and some examples. In Section 6, we sketch the proof of our uniqueness result and discuss several other issues. Section 7 concludes. Proofs not contained in the text are in the Appendix.

2 Model

The set of agents is $\mathcal{I} = \{1, \ldots, I\}$. There is a single indivisible good to allocate among them. The value to the principal of assigning the object to agent i depends on information which is known only to i. Formally, the value to the principal of allocating the good to agent i is t_i where t_i is private information of agent i. The value to the principal of assigning the object to no one is normalized to zero. We assume that the t_i's are independently distributed. The distribution of t_i has a strictly positive density f_i over the interval $T_i \equiv [\underline{t}_i, \bar{t}_i]$ where $0 \leq \underline{t}_i < \bar{t}_i < \infty$. (All results extend to allowing the support to be unbounded above.) We use F_i to denote the corresponding distribution function. Let $T = \prod_i T_i$.

The principal can *check* the type of agent i at a cost $c_i > 0$. We interpret checking as requiring documentation by agent i to demonstrate what his type is. If the principal checks some agent, she learns that agent's type. The cost c_i is interpreted as the direct cost to the principal of reviewing the information provided plus the costs to the principal associated with the resource cost to the agent of providing this documentation. The cost to the agent of providing documentation is zero. To understand this, think of the agent's resources as allocated to activities which are either directly productive for the principal or which provide information for checking claims. The agent is indifferent over how these resources are used since they will all be used regardless. Thus by directing the agent to spend resources on providing information, the principal loses some output the agent would have produced with the resources otherwise while the agent's utility is unaffected.[2] In Section 6, we show one way to generalize our model to allow agents to bear some costs

[2]One reason this assumption is a convenient simplification is that dropping it allows a "back door" for transfers. If agents bear costs of providing documentation, then the principal can use threats to require documentation as a way of "fining" agents and thus to help achieve incentive compatibility. This both complicates the analysis and indirectly introduces a form of the transfers we wish to exclude.

of providing documentation which does not change our results qualitatively.

We assume that every agent strictly prefers receiving the object to not receiving it. Consequently, we can take the payoff to an agent to be the probability he receives the good. The intensity of the agents' preferences plays no role in the analysis, so these intensities may or may not be related to the types.[3] We also assume that each agent's reservation utility is less than or equal to his utility from not receiving the good. Since monetary transfers are not allowed, this is the worst payoff an agent could receive under a mechanism. Consequently, individual rationality constraints do not bind and so are disregarded throughout.

In its most general form, a mechanism can be quite complex, allowing the principal to decide which agents to check as a function of the outcome of previous checks and cheap talk statements for multiple stages before finally allocating the good or deciding to not allocate it at all. Without loss of generality, we can restrict attention to truth telling equilibria of mechanisms where each agent sends a report of his type to the principal who is committed to (1) a probability distribution over which agents (if any) are checked as a function of the reports and (2) a probability distribution over which agent (if any) receives the good as a function of the reports and the outcome of checking. While this does not follow from the usual Revelation Principle directly (as the usual version does not apply to games with verification), the argument is similar. Fix a dynamic mechanism and any equilibrium. The equilibrium defines a function from type profiles into probability distributions over outcomes. More specifically, an outcome is a sequence of checks and an allocation of the good (perhaps to no one). Replace this mechanism with a direct mechanism where agents report types and the outcome (or distribution over outcomes) given a vector of type reports is what would happen in the equilibrium if this report were true. Clearly, just as in the usual Revelation Principle, truth telling is an equilibrium of this mechanism and this equilibrium yields the same outcome as the original equilibrium in the dynamic mechanism. We can replace any outcome which is a *sequence* of checks with an outcome where exactly these checks are done *simultaneously*. All agents and the principal are indifferent between these two outcomes. Hence the altered form of the mechanism where we change outcomes in this way also has a truth telling equilibrium and yields an outcome which is just as good for the principal as the original equilibrium of the dynamic mechanism.

Given that we focus on truth telling equilibria, all situations in which agent i's report is checked and found to be false are off the equilibrium path. The specification of the mechanism for such a situation cannot affect the incentives of any agent $j \neq i$ since agent j will expect i's report to be truthful. Thus the specification only affects agent

[3] In other words, suppose we let the payoff of i from receiving the good be $\bar{u}_i(t_i)$ and let his utility from not receiving it be $\underline{u}_i(t_i)$ where $\bar{u}_i(t_i) > \underline{u}_i(t_i)$ for all i and all t_i. Then it is simply a renormalization to let $\bar{u}_i(t_i) = 1$ and $\underline{u}_i(t_i) = 0$ for all t_i.

i's incentives to be truthful. Since we want i to have the strongest possible incentives to report truthfully, we may as well assume that if i's report is checked and found to be false, then the good is given to agent i with probability 0. Hence we can further reduce the complexity of a mechanism to specify which agents are checked and which agent receives the good as a function of the reports, where the latter applies only when the checked reports are accurate.

Finally, it is not hard to see that any agent's incentive to reveal his type is unaffected by the possibility of being checked in situations where he does not receive the object regardless of the outcome of the check. That is, if an agent's report is checked even when he would not receive the object if found to have told the truth, his incentives to report honestly are not affected. Since checking is costly for the principal, this means that if the principal checks an agent, then (if he is found to have been honest), he must receive the object with probability 1.

Therefore, we can think of the mechanism as specifying two probabilities for each agent: the probability he is awarded the object without being checked and the probability he is awarded the object conditional on a successful check. Let $p_i(t)$ denote the total probability i is assigned the good and $q_i(t)$ the probability i is assigned the good and checked. So a mechanism is a $2I$ tuple of functions, $(p_i, q_i)_{i\in\mathcal{I}}$ where $p_i : T \to [0,1]$, $q_i : T \to [0,1]$, $\sum_i p_i(t) \leq 1$ for all $t \in T$, and $q_i(t) \leq p_i(t)$ for all $i \in \mathcal{I}$ and all $t \in T$. Henceforth, the word "mechanism" will be used only to denote such a tuple of functions, generally denoted (p,q) for simplicity.

The principal's objective function is

$$\mathrm{E}_t \left[\sum_i (p_i(t)t_i - q_i(t)c_i) \right].$$

The incentive compatibility constraint for i is then

$$\mathrm{E}_{t_{-i}} p_i(t) \geq \mathrm{E}_{t_{-i}} \left[p_i(\hat{t}_i, t_{-i}) - q_i(\hat{t}_i, t_{-i}) \right], \quad \forall \hat{t}_i, t_i \in T_i, \quad \forall i \in \mathcal{I}.$$

Given a mechanism (p,q), let

$$\hat{p}_i(t_i) = \mathrm{E}_{t_{-i}} p_i(t)$$

and

$$\hat{q}_i(t_i) = \mathrm{E}_{t_{-i}} q_i(t).$$

The $2I$ tuple of functions $(\hat{p}, \hat{q})_{i\in\mathcal{I}}$ is the *reduced form* of the mechanism (p,q). We say that (p^1, q^1) and (p^2, q^2) are *equivalent* if $\hat{p}^1 = \hat{p}^2$ and $\hat{q}^1 = \hat{q}^2$ up to sets of measure zero. It is easy to see that we can write the incentive compatibility constraints and the objective function of the principal as a function only of the reduced form of the mechanism. Hence if (p^1, q^1) is an optimal incentive compatible mechanism, (p^2, q^2) must be as well. Therefore, we can only identify the optimal mechanism up to equivalence.

3 The Sufficiency of Favored–Agent Mechanisms

Our main result in this section is that we can restrict attention to a class of mechanisms we call *favored–agent mechanisms*. To be more specific, first we show that there is always a favored–agent mechanism which is an optimal mechanism. Second, we show that every Bayesian optimal mechanism is equivalent to a randomization over favored–agent mechanisms. Hence to compute the set of optimal mechanisms, we can simply optimize over the much simpler class of favored–agent mechanisms. In the next section, we use this result to characterize optimal mechanisms in more detail.

To be more precise, we say that (p, q) is a *favored–agent mechanism* if there exists a *favored agent* $i^* \in \mathcal{I}$ and a *threshold* $v^* \in \mathbf{R}_+$ such that the following holds up to sets of measure zero. First, if $t_i - c_i < v^*$ for all $i \neq i^*$, then $p_{i^*}(t) = 1$ and $q_i(t) = 0$ for all i. That is, if every agent other than the favored agent reports a "value" $t_i - c_i$ below the threshold, then the favored agent receives the object and no agent is checked. Second, if there exists $j \neq i^*$ such that $t_j - c_j > v^*$ and $t_i - c_i > \max_{k \neq i}(t_k - c_k)$, then $p_i(t) = q_i(t) = 1$ and $p_k(t) = q_k(t) = 0$ for all $k \neq i$. That is, if any agent other than the favored agent reports a value above the threshold, then the agent with the highest reported value (regardless of whether he is the favored agent or not) is checked and, if his report is verified, receives the good.

Note that this is a very simple class of mechanisms. Optimizing over this set of mechanisms simply requires us to pick one of the agents to favor and a number for the threshold, as opposed to $2I$ probability distributions.

Theorem 1. *There always exists a Bayesian optimal mechanism which is a favored–agent mechanism.*

A very incomplete intuition for this result is the following. For simplicity, suppose $c_i = c$ for all i and suppose $T_i = [0, 1]$ for all i. Clearly, the principal would ideally give the object to the agent with the highest t_i. Of course, this isn't incentive compatible as each agent would claim to have type 1. By always checking the agent with the highest report, the principal can make this allocation of the good incentive compatible. So this is a feasible mechanism.

Suppose the highest reported type is below c. Obviously, it's better for the principal to not to check in this case since it costs more to check than it could possibly be worth. Thus we can improve on this mechanism by only checking the agent with the highest report when that report is above c, giving the good to no one (and checking no one) when the highest report is below c. It is not hard to see that this mechanism is incentive compatible and, as noted, an improvement over the previous mechanism.

However, we can improve on this mechanism as well. Obviously, the principal could select any agent at random if all the reports are below c and give the good to that agent. Again, this is incentive compatible. Since all the types are positive, this mechanism improves on the previous one.

The principal can do still better by selecting the "best" person to give the good to when all the reports are below c. To think more about this, suppose the principal gives the good to agent 1 if all reports are below c. Continue to assume that if any agent reports a type above c, then the principal checks the highest report and gives the good to this agent if the report is true. This mechanism is clearly incentive compatible. However, the principal can also achieve incentive compatibility and the same allocation of the good while saving on checking costs: he doesn't need to check 1's report when he is the only agent to report a type above c. To see why this cheaper mechanism is also incentive compatible, note that if everyone else's type is below c, 1 gets the good no matter what he says. Hence he only cares what happens if at least one other agent's report is above c. In this case, he will be checked if he has the high report and hence cannot obtain the good by lying. Hence it is optimal for him to tell the truth.

This mechanism is the favored–agent mechanism with 1 as the favored agent and $v^* = 0$. Of course, if the principal chooses the favored agent and the threshold v^* optimally, he must improve on this payoff.

This intuition does not show that some more complex mechanism cannot be superior, so it is far from a proof. Indeed, we prove this result as a corollary to the next theorem, a result whose proof is rather complex.

Let $\mathcal{F}$ denote the set of favored–agent mechanisms and let $\mathcal{F}^*$ denote the set of optimal favored–agent mechanisms. By Theorem 1, if a favored–agent mechanism is better for the principal than every other favored–agent mechanism, then it must be better for the principal than every other incentive compatible mechanism, whether in the favored–agent class or not. Hence every mechanism in $\mathcal{F}^*$ is an optimal mechanism even without the restriction to favored–agent mechanisms.

Given two mechanisms, (p^1, q^1) and (p^2, q^2) and a number $\lambda \in (0, 1)$, we can construct a new mechanism, say (p^λ, q^λ), by $(p^\lambda, q^\lambda) = \lambda(p^1, q^1) + (1 - \lambda)(p^2, q^2)$, where the right–hand side refers to the pointwise convex combination of these functions. The mechanism (p^λ, q^λ) is naturally interpreted as a random choice by the principal between the mechanisms (p^1, q^1) and (p^2, q^2). It is easy to see that if (p^k, q^k) is incentive compatible for $k = 1, 2$, then (p^λ, q^λ) is incentive compatible. Also, the principal's payoff is linear in (p, q), so if both (p^1, q^1) and (p^2, q^2) are optimal for the principal, it must be true that (p^λ, q^λ) is optimal for the principal. It is easy to see that this implies that any mechanism in the convex hull of $\mathcal{F}^*$, denoted $\text{conv}(\mathcal{F}^*)$, is optimal.

Finally, as noted in Section 2, if a mechanism (p, q) is optimal, then any mechanism equivalent to it in the sense of having the same reduced form up to sets of measure zero must also be optimal. Hence Theorem 1 implies that any mechanism equivalent to a mechanism in $\text{conv}(\mathcal{F}^*)$ must be optimal. The following theorem shows the stronger result that this is precisely the set of optimal mechanisms.

Theorem 2. *A mechanism is optimal if and only if it is equivalent to some mechanism in* $\text{conv}(\mathcal{F}^*)$*.*

Section 6 contains a sketch of the proof of this result.

Theorem 2 says that all optimal mechanisms are, essentially, favored–agent mechanisms or randomization over such mechanisms. Hence we can restrict attention to favored–agent mechanisms without loss of generality. This result also implies that if there is a unique optimal favored–agent mechanism, then $\text{conv}(\mathcal{F}^*)$ is a singleton so that there is essentially a unique optimal mechanism.

4 Optimal Favored–Agent Mechanisms

We complete the specification of the optimal mechanism by characterizing the optimal threshold and the optimal favored agent. We show that conditional on the selection of the favored agent, the optimal favored–agent mechanism is unique. After characterizing the optimal threshold given the choice of the favored agent and showing this result, we consider the optimal selection of the favored agent.

For each i, define t_i^* by

$$\text{E}(t_i) = \text{E}(\max\{t_i, t_i^*\}) - c_i. \tag{1}$$

It is easy to show that t_i^* is well–defined. To see this, let

$$\psi_i(t^*) = \text{E}(\max\{t_i, t^*\}) - c_i.$$

Clearly, $\psi_i(\underline{t}_i) = \text{E}(t_i) - c_i < \text{E}(t_i)$. For $t^* \geq \underline{t}_i$, ψ_i is strictly increasing in t^* and goes to infinity as $t^* \to \infty$. Hence there is a unique $t_i^* > \underline{t}_i$.[4]

It will prove useful to give two alternative definitions of t_i^*. Note that we can rearrange the definition above as

$$\int_{\underline{t}_i}^{t_i^*} t_i f_i(t_i)\, dt_i = t_i^* F_i(t_i^*) - c_i$$

[4]Note that if we allowed $c_i = 0$, we would have $t^* = \underline{t}_i$. This fact together with what we show below implies the unsurprising observation that if all the costs are zero, the principal always checks the agent who receives the object and gets the same payoff as under complete information.

or

$$t_i^* = \mathrm{E}[t_i \mid t_i \leq t_i^*] + \frac{c_i}{F_i(t_i^*)}. \tag{2}$$

Finally, note that we could rearrange the next–to–last equation as

$$c_i = t_i^* F_i(t_i^*) - \int_{\underline{t}_i}^{t_i^*} t_i f_i(t_i)\, dt_i = \int_{\underline{t}_i}^{t_i^*} F_i(\tau)\, d\tau.$$

So a final definition of t_i^* is

$$\int_{\underline{t}_i}^{t_i^*} F_i(\tau)\, d\tau = c_i. \tag{3}$$

Given any i, let $\mathcal{F}_i$ denote the set of favored agent mechanisms with i as the favored agent.

Theorem 3. *The unique best mechanism in $\mathcal{F}_i$ is obtained by setting the threshold v^* equal to $t_i^* - c_i$.*

Proof. For notational convenience, let the favored agent i equal 1. Contrast the principal's payoff to thresholds $t_1^* - c_1$ and $\hat{v}^* > t_1^* - c_1$. Let t_{-1} denote the profile of types of $j \neq 1$ and let $x = \max_{j \neq 1}(t_j - c_j)$ — that is, the highest value of (and hence reported by) one of the other agents. Then the principal's payoff as a function of the threshold and x is given by

	$x < t_1^* - c_1 < \hat{v}^*$	$t_1^* - c_1 < x < \hat{v}^*$	$t_1^* - c_1 < \hat{v}^* < x$
$t_1^* - c_1$	$\mathrm{E}(t_1)$	$\mathrm{E}\max\{t_1 - c_1, x\}$	$\mathrm{E}\max\{t_1 - c_1, x\}$
$\hat{v}^*$	$\mathrm{E}(t_1)$	$\mathrm{E}(t_1)$	$\mathrm{E}\max\{t_1 - c_1, x\}$

To see this, note that if $x < t_1^* - c_1 < \hat{v}^*$, then the principal gives the object to agent 1 without a check using either threshold. If $t_1^* - c_1 < \hat{v}^* < x$, then the principal give the object to either 1 or the highest of the other agents with a check and so receives a payoff of either $t_1 - c_1$ or x, whichever is larger. Finally, if $t_1^* - c_1 < x < \hat{v}^*$, then with threshold $t_1^* - c_1$, the principal's payoff is the larger of $t_1 - c_1$ and x, while with threshold $\hat{v}^*$, she gives the object to agent 1 without a check and has payoff $\mathrm{E}(t_1)$.

Recall that $t_1^* > \underline{t}_1$. Hence $t_1 < t_1^*$ with strictly positive probability. Therefore, for $x > t_1^* - c_1$, we have

$$\mathrm{E}\max\{t_1 - c_1, x\} > \mathrm{E}\max\{t_1 - c_1, t_1^* - c_1\}.$$

But the right–hand side is $\mathrm{E}\max\{t_1, t_1^*\} - c_1$ which equals $\mathrm{E}(t_1)$ by our first definition of t_i^*. Hence given that 1 is the favored agent, the threshold $t_1^* - c_1$ weakly dominates than any larger threshold. A similar argument shows that the threshold $t_1^* - c_1$ weakly dominates any smaller threshold, establishing that it is optimal.

To see that the optimal mechanism in this class is unique, note that the comparison of threshold $t_1^* - c_1$ to a larger threshold v^* is strict unless the middle column of the table above has zero probability. That is, the only situation in which the principal is indifferent between the threshold $t_1^* - c_1$ and the larger threshold v^* is when the allocation of the good and checking decisions are the same with probability 1 given either threshold. That is, indifference occurs only when changes in the threshold do not change (p, q). Hence there is a unique best mechanism in $\mathcal{F}_i$. ▮

Given that the best mechanism in each $\mathcal{F}_i$ is unique, it remains only to characterize the optimal choice or choices of i.

Theorem 4. *The optimal choice of the favored agent is any i with $t_i^* - c_i = \max_j (t_j^* - c_j)$.*

Proof. For notational convenience, number the agents so that 1 is any i with $t_i^* - c_i = \max_j t_j^* - c_j$ and let 2 denote any other agent so $t_1^* - c_1 \geq t_2^* - c_2$. First, we show that the principal must weakly prefer having 1 as the favored agent at a threshold of $t_2^* - c_2$ to having 2 as the favored agent at this threshold. If $t_1^* - c_1 = t_2^* - c_2$, this argument implies that the principal is indifferent between having 1 and 2 as the favored agents, so we then turn to the case where $t_1^* - c_1 > t_2^* - c_2$ and show that it must always be the case that the principal strictly prefers having 1 as the favored agent at threshold $t_1^* - c_1$ to favoring 2 with threshold $t_2^* - c_2$, establishing the claim.

So first let us show that it is weakly better to favor 1 at threshold $t_2^* - c_2$ than to favor 2 at the same threshold. First, note that if any agent other than 1 or 2 reports a value above $t_2^* - c_2$, the designation of the favored agent is irrelevant since the good will be assigned to the agent with the highest reported value and this report will be checked. Hence we may as well condition on the event that all agents other than 1 and 2 report values below $t_2^* - c_2$. If this event has zero probability, we are done, so we may as well assume this probability is strictly positive. Similarly, if both agents 1 and 2 report values above $t_2^* - c_2$, the object will go to whichever reports a higher value and the report will be checked, so again the designation of the favored agent is irrelevant. Hence we can focus on situations where at most one of these two agents reports a value above $t_2^* - c_2$ and, again, we may as well assume the probability of this event is strictly positive.

If both agents 1 and 2 report values below $t_2^* - c_2$, then no one is checked under either mechanism. In this case, the good goes to the agent who is favored under the mechanism. So suppose 1's reported value is above $t_2^* - c_2$ and 2's is below. If 1 is the favored agent, he gets the good without being checked, while he receives the good with a check if 2 were favored. The case where 2's reported value is above $t_2^* - c_2$ and 1's is below is symmetric. For brevity, let $\hat{t}_1 = t_2^* - c_2 + c_1$. Note that 1's report is below the threshold iff $t_1 - c_1 < t_2^* - c_2$ or, equivalently, $t_1 < \hat{t}_1$. Given the reasoning above, we see that under threshold $t_2^* - c_2$, it is weakly better to have 1 as the favored agent if

$$
\begin{aligned}
F_1(\hat{t}_1)F_2(t_2^*)\mathrm{E}[t_1 \mid t_1 \le \hat{t}_1] + [1 - F_1(\hat{t}_1)]F_2(t_2^*)\mathrm{E}[t_1 \mid t_1 > \hat{t}_1] \\
+ F_1(\hat{t}_1)[1 - F_2(t_2^*)]\left\{\mathrm{E}[t_2 \mid t_2 > t_2^*] - c_2\right\} \\
\ge F_1(\hat{t}_1)F_2(t_2^*)\mathrm{E}[t_2 \mid t_2 \le t_2^*] + [1 - F_1(\hat{t}_1)]F_2(t_2^*)\left\{\mathrm{E}[t_1 \mid t_1 > \hat{t}_1] - c_1\right\} \\
+ F_1(\hat{t}_1)[1 - F_2(t_2^*)]\mathrm{E}[t_2 \mid t_2 > t_2^*].
\end{aligned}
$$

If $F_1(\hat{t}_1) = 0$, then this equation reduces to

$$F_2(t_2^*)\mathrm{E}[t_1 \mid t_1 > \hat{t}_1] \ge F_2(t_2^*)\left\{\mathrm{E}[t_1 \mid t_1 > \hat{t}_1] - c_1\right\},$$

which must hold. If $F_1(\hat{t}_1) > 0$, then we can rewrite the equation as

$$\mathrm{E}[t_1 \mid t_1 \le \hat{t}_1] + \frac{c_1}{F_1(\hat{t}_1)} - c_1 \ge \mathrm{E}[t_2 \mid t_2 \le t_2^*] + \frac{c_2}{F_2(t_2^*)} - c_2.$$

From equation (2), the right–hand side equation (4) is $t_2^* - c_2$. Hence we need to show

$$\mathrm{E}[t_1 \mid t_1 \le \hat{t}_1] + \frac{c_1}{F_1(\hat{t}_1)} - c_1 \ge t_2^* - c_2. \tag{4}$$

Recall that $t_2^* - c_2 \le t_1^* - c_1$ or, equivalently, $\hat{t}_1 \le t_1^*$. Hence from equation (1), we have

$$\mathrm{E}(t_1) \ge \mathrm{E}[\max\{t_1, \hat{t}_1\}] - c_1.$$

A similar rearrangement to our derivation of equation (2) yields

$$\mathrm{E}[t_1 \mid t_1 \le \hat{t}_1^*] + \frac{c_1}{F_1(\hat{t}_1^*)} \ge \hat{t}_1.$$

Hence

$$\mathrm{E}[t_1 \mid t_1 \le \hat{t}_1^*] + \frac{c_1}{F_1(\hat{t}_1^*)} - c_1 \ge \hat{t}_1 - c_1 = t_2^* - c_2 + c_1 - c_1 = t_2^* - c_2,$$

implying equation (4). Hence as asserted, it is weakly better to have 1 as the favored agent with threshold $t_2^* - c_2$ than to have 2 as the favored agent with this threshold.

Suppose that $t_1^* - c_1 = t_2^* - c_2$. In this case, an argument symmetric to the one above shows that the principal weakly prefers favoring 2 at threshold $t_1^* - c_1$ to favoring 1 at the same threshold. Hence the principal must be indifferent between favoring 1 or 2 at threshold $t_1^* - c_1 = t_2^* - c_2$.

We now turn to the case where $t_1^* - c_1 > t_2^* - c_2$. The argument above is easily adapted to show that favoring 1 at threshold $t_2^* - c_2$ is strictly better than favoring 2 at this threshold if the event that $t_j - c_j < t_2^* - c_2$ for every $j \ne 1, 2$ has strictly positive probability. To see this, note that if this event has strictly positive probability, then the claim follows iff

$$\begin{aligned} &F_1(\hat{t}_1)F_2(t_2^*)\mathrm{E}[t_1 \mid t_1 \le \hat{t}_1] + [1 - F_1(\hat{t}_1)]F_2(t_2^*)\mathrm{E}[t_1 \mid t_1 > \hat{t}_1] \\ &\qquad + F_1(\hat{t}_1)[1 - F_2(t_2^*)] \{\mathrm{E}[t_2 \mid t_2 > t_2^*] - c_2\} \\ &\qquad > F_1(\hat{t}_1)F_2(t_2^*)\mathrm{E}[t_2 \mid t_2 \le t_2^*] + [1 - F_1(\hat{t}_1)]F_2(t_2^*) \left\{\mathrm{E}[t_1 \mid t_1 > \hat{t}_1] - c_1\right\} \\ &\qquad + F_1(\hat{t}_1)[1 - F_2(t_2^*)]\mathrm{E}[t_2 \mid t_2 > t_2^*]. \end{aligned}$$

If $F_1(\hat{t}_1) = 0$, this holds iff $F_2(t_2^*)c_1 > 0$. By assumption, $c_i > 0$ for all i. Also, $\underline{t}_2 < t_2^*$, so $F_2(t_2^*) > 0$. Hence this must hold if $F_1(\hat{t}_1) = 0$. If $F_1(\hat{t}_1) > 0$, then this holds if equation (4) holds strictly. It is easy to use the argument above and $t_1^* - c_1 > t_2^* - c_2$ to show that this holds.

So if the event that $t_j - c_j < t_2^* - c_2$ for every $j \neq 1, 2$ has strictly positive probability, the principal strictly prefers having 1 as the favored agent to having 2. Suppose, then, that this event has zero probability. That is, there is some $j \neq 1, 2$ such that $t_j - c_j \geq t_2^* - c_2$ with probability 1. In this case, the principal is indifferent between having 1 as the favored agent at threshold $t_2^* - c_2$ versus favoring 2 at this threshold. However, we now show that the principal must strictly prefer favoring 1 with threshold $t_1^* - c_1$ to either option and thus strictly prefers having 1 as the favored agent.

To see this, recall from the proof of Theorem 3 that the principal strictly prefers favoring 1 at threshold $t_1^* - c_1$ to favoring him at a lower threshold v^* if there is a positive probability that $v^* < t_j - c_j < t_1^* - c_1$ for some $j \neq 1$. Thus, in particular, the principal strictly prefers favoring 1 at threshold $t_1^* - c_1$ to favoring him at $t_2^* - c_2$ if there is a $j \neq 1, 2$ such that the event $t_2^* - c_2 < t_j - c_j < t_1^* - c_1$ has strictly positive probability. By hypothesis, there is a $j \neq 1, 2$ such that $t_2^* - c_2 < t_j - c_j$ with probability 1, so we only have to establish that for this j, we have a positive probability of $t_j - c_j < t_1^* - c_1$. Recall that $\underline{t}_j - c_j < t_j^* - c_j$ by definition of t_j^*. By hypothesis, $t_j^* - c_j < t_1^* - c_1$. Hence we have $t_j - c_j < t_1^* - c_1$ with strictly positive probability, completing the proof. ▌

Summarizing, we see that the set of optimal favored–agent mechanisms is easily characterized. A favored–agent mechanism is optimal if and only if the favored agent i satisfies $t_i^* - c_i = \max_j t_j^* - c_j$ and the threshold v^* satisfies $v^* = \max_j t_j^* - c_j$. Thus the set of optimal mechanisms is equivalent to picking a favored–agent mechanism with threshold $v^* = \max_j t_j^* - c_j$ and randomizing over which of the agents i with $t_i^* - c_i$ equal to this threshold to favor. Loosely speaking, for "generic" distributions and checking costs, there will be a unique i with $t_i^* - c_i = \max_j t_j^* - c_j$ and hence a unique optimal mechanism.

5 Comparative Statics and Examples

Our characterization of the optimal favored agent and threshold makes it easy to give comparative statics. Recall our third expression for t_i^* which is

$$\int_{\underline{t}_i}^{t_i^*} F_i(\tau)\, d\tau = c_i. \tag{5}$$

Hence an increase in c_i increases t_i^*. Also, from our first definition of t_i^*, note that $t_i^* - c_i$ is that value of v_i^* solving $\mathrm{E}(t_i) = \mathrm{E}\max\{t_i - c_i, v_i^*\}$. Obviously for fixed v_i^*, the right–hand side is decreasing in c_i, so $t_i^* - c_i$ must be increasing in c_i. Hence, all else equal, the higher is c_i, the more likely i is to be selected as the favored agent. To see the intuition, note that if c_i is larger, then the principal is less willing to check agent i's report. Since the favored agent is the one the principal checks least often, this makes it more desirable to make i the favored agent.

It is also easy to see that a first–order or second–order stochastic dominance shift upward in F_i reduces the left–hand side of equation (5) for fixed t_i^*, so to maintain the equality, t_i^* must increase. Therefore, such a shift makes it more likely than i is the favored agent and increases the threshold in this case. Hence both "better" (FOSD) and "less risky" (SOSD) agents are more likely to be favored.

The intuition for the effect of a first–order stochastic dominance increase in t_i is clear. If agent i is more likely to have high value, he is a better choice to be the favored agent.

The intuition for why less risky agents are favored is not as immediate. One way to see the idea is to suppose that there is one agent whose type is completely riskless — i.e., is known to the principal. Obviously, there is no reason for the principal to check this agent since his type is known. Thus setting him as the favored agent — the least likely agent to be checked — seems natural.

We illustrate with two examples. First, suppose we have two agents where $t_1 \sim U[0,1]$, $t_2 \sim U[0,2]$ and $c_1 = c_2 = c$. It is easy to calculate t_i^*. From equation (1), we have

$$\mathrm{E}(t_i) = \mathrm{E}\max\{t_i, t_i^*\} - c.$$

For $i = 1$, if $t_1^* < 1$, it must solve

$$\frac{1}{2} = \int_0^{t_1^*} t_1^*\, ds + \int_{t_1^*}^1 s\, ds - c$$

or

$$\frac{1}{2} = (t_1^*)^2 + \frac{1}{2} - \frac{(t_1^*)^2}{2} - c$$

so

$$t_1^* = \sqrt{2c}.$$

This holds only if $c \leq 1/2$ so that $t_1^* \leq 1$. Otherwise, $\mathrm{E}\max\{t_1, t_1^*\} = t_1^*$, so $t_1^* = (1/2)+c$. Hence

$$t_1^* = \begin{cases} \sqrt{2c}, & \text{if } c \leq 1/2 \\ (1/2) + c, & \text{otherwise.} \end{cases}$$

A similar calculation shows that

$$t_2^* = \begin{cases} 2\sqrt{c}, & \text{if } c \leq 1 \\ 1 + c, & \text{otherwise.} \end{cases}$$

It is easy to see that $t_2^* > t_1^*$ for all $c > 0$, so 2 is the favored agent. The optimal threshold value is

$$t_2^* - c = \begin{cases} 2\sqrt{c} - c, & \text{if } c \leq 1 \\ 1, & \text{otherwise.} \end{cases}$$

Note that if $2\sqrt{c} \geq 1$ (or $c \geq 1/4$), then the threshold value $2\sqrt{c} - c \geq 1 - c$. In this case, the fact that t_1 is always less than 1 implies that $t_1 - c_1 \leq v^*$ with probability 1. In this case, the favored agent mechanism corresponds to simply giving the good to agent 2 independently of the reports. If $c \in (0, 1/4)$, then there are type profiles for which agent 1 receives the good, specifically those with $t_1 > \max\{2\sqrt{c}, t_2\}$.

For a second example, suppose again we have two agents, but now $t_i \sim U[0,1]$ for $i = 1, 2$. Assume $c_2 > c_1 > 0$. In this case, calculations similar to those above show that

$$t_i^* = \begin{cases} \sqrt{2c_i}, & \text{if } c_i \leq 1/2 \\ (1/2) + c_i, & \text{otherwise} \end{cases}$$

so

$$t_i^* - c_i = \begin{cases} \sqrt{2c_i} - c_i, & \text{if } c_i \leq 1/2 \\ (1/2), & \text{otherwise.} \end{cases}$$

It is easy to see that $\sqrt{2c_i} - c_i$ is an increasing function for $c_i \in (0, 1/2)$. Thus if $c_1 < 1/2$, we must have $t_2^* - c_2 > t_1^* - c_1$, so that 2 is the favored agent. If $c_1 \geq 1/2$, then $t_1^* - c_1 = t_2^* - c_2 = 1/2$, so the principal is indifferent over which agent should be favored. Note that in this case, the cost of checking is so high that the principal never checks, so that the favored agent simply receives the good independent of the reports. Since the distributions of t_1 and t_2 are the same, it is not surprising that the principal is indifferent over who should be favored in this case. It is not hard to show that when $c_1 < 1/2$ so that 2 is the favored agent, 2's payoff is higher than 1's. That is, it is advantageous to be favored. Note that this implies that agents may have incentives to increase the cost of being checked in order to become favored, an incentive which is costly for the principal.

6 Discussion

6.1 Proof Sketch

In this section, we sketch the proof of Theorem 2. It is easy to see that Theorem 1 is a corollary.

First, it is useful to rewrite the optimization problem as follows. Recall that $\hat{p}_i(t_i) = \mathrm{E}_{t_{-i}} p_i(t_i, t_{-i})$ and $\hat{q}_i(t_i) = \mathrm{E}_{t_{-i}} q_i(t_i, t_{-i})$. We can write the incentive compatibility constraint as

$$\hat{p}_i(t_i') \geq \hat{p}_i(t_i) - \hat{q}_i(t_i), \quad \forall t_i, t_i' \in T_i.$$

Clearly, this holds if and only if

$$\inf_{t_i' \in T_i} \hat{p}_i(t_i') \geq \hat{p}_i(t_i) - \hat{q}_i(t_i), \quad \forall t_i \in T_i.$$

Letting $\varphi_i = \inf_{t_i' \in T_i} \hat{p}_i(t_i')$, we can rewrite the incentive compatibility constraint as

$$\hat{q}_i(t_i) \geq \hat{p}_i(t_i) - \varphi_i, \quad \forall t_i \in T_i.$$

Because the objective function is strictly decreasing in $\hat{q}_i(t_i)$, this constraint must bind, so $\hat{q}_i(t_i) = \hat{p}_i(t_i) - \varphi_i$. Hence we can rewrite the objective function as

$$\begin{aligned}
\mathrm{E}_t \left[\sum_i p_i(t) t_i - c_i \sum_i q_i(t) \right] &= \sum_i \mathrm{E}_{t_i} \left[\hat{p}_i(t_i) t_i - c_i \hat{q}_i(t_i) \right] \\
&= \sum_i \mathrm{E}_{t_i} \left[\hat{p}_i(t_i)(t_i - c_i) + \varphi_i c_i \right] \\
&= \mathrm{E}_t \left[\sum_i [p_i(t)(t_i - c_i) + \varphi_i c_i] \right].
\end{aligned}$$

Some of the arguments below will use the reduced form probabilities and hence rely on the first expression for the payoff function, while others focus on the "nonreduced" mechanism and so rely on the second expression.

Summarizing, we can replace the choice of p_i and q_i functions for each i with the choice of a number $\varphi_i \in [0, 1]$ for each i and a function $p_i : T \to [0, 1]$ satisfying $\sum_i p_i(t) \leq 1$ and $\mathrm{E}_{t_{-i}} p_i(t) \geq \varphi_i \geq 0$. Note that this last constraint implies $\mathrm{E}_t p_i(t) \geq \varphi_i$, so

$$\sum_i \varphi_i \leq \sum_i \mathrm{E}_t p_i(t) = \mathrm{E}_t \sum_i p_i(t) \leq 1.$$

Hence the constraint that $\varphi_i \leq 1$ cannot bind and so can be ignored.

The remainder of the proof sketch is more complex and so we introduce several simplifications for simplicity. First, the proof sketch assumes $I = 2$ and $c_i = c$ for all i. The equal costs assumption implies that the threshold value v^* can be thought of as defining a threshold type t^* to which we compare the t_i reports.

Second, we will consider the case of finite type spaces and disregard certain boundary issues. To explain, the statement of Theorem 2 is made cleaner by our use of a continuum of types. Without this, we would have "boundary" types where there is some arbitrariness to the optimal mechanism, making a statement of uniqueness more complex. On the other hand, the continuum of types greatly complicates the proof of our characterization of optimal mechanisms. In this proof sketch, we explain how the proof would work if we focused on finite type spaces, ignoring what happens at boundaries. The proof in the appendix can be seen as a generalization of these ideas to continuous type spaces.

The proof sketch has five steps. First, we observe that every optimal mechanism is monotonic in the sense that higher types are more likely to receive the object. That is, for all i, $t_i > t'_i$ implies $\hat{p}_i(t_i) \geq \hat{p}_i(t'_i)$. To see the intuition, suppose we have an optimal mechanism which violates this monotonicity property so that we have types t_i and t'_i such that $\hat{p}_i(t_i) < \hat{p}_i(t'_i)$ even though $t_i > t'_i$. To simplify further, suppose that these two types have the same probability. Then consider the mechanism p^* which is the same as this one *except* we flip the roles of t_i and t'_i. That is, for any type profile $\hat{t}$ where $\hat{t}_i \notin \{t_i, t'_i\}$, we let $p^*_i(\hat{t}) = p_i(\hat{t})$. For any type profile of the form (t_i, t_{-i}) we assign the p's the original mechanism assigned to (t'_i, t_{-i}) and conversely. Since the probabilities of these types are the same, our independence assumption implies that for every $j \neq i$, agent j is unaffected by the change in the sense that $\hat{p}^*_j = \hat{p}_j$. Obviously, $\hat{p}^*_i(t_i) \geq \hat{p}^*_i(t'_i) = \hat{p}_i(t_i)$. Since the original mechanism was feasible, we must have $\hat{p}_i(t_i) \geq \varphi_i$, so this mechanism must be feasible. It is easy to see that this change improves the objective function, so the original mechanism could not have been optimal.

This monotonicity property implies that any optimal mechanism has the property that there is a cutoff type, say $\hat{t}_i \in [\underline{t}_i, \bar{t}_i]$, such that $\hat{p}_i(t_i) = \varphi_i$ for $t_i < \hat{t}_i$ and $\hat{p}_i(t_i) > \varphi_i$ for $t_i > \hat{t}_i$.

The second step shows that if we have a type profile $t = (t_1, t_2)$ such that $t_2 > t_1 > \hat{t}_1$, then the optimal mechanism has $p_2(t) = 1$. To see this, suppose to the contrary that $p_2(t) < 1$. Then we can change the mechanism by increasing this probability slightly and lowering the probability of giving the good to 1 (or, if the probability of giving it to 1 was 0, lowering the probability that the good is not given to either agent). Since $t_1 > \hat{t}_1$, we have $\hat{p}_1(t_1) > \varphi_1$ before the change, so if the change is small enough, we still satisfy this constraint. Since $t_2 > t_1$, the value of the objective function increases, so the original mechanism could not have been optimal.

The third step is to show that for a type profile $t = (t_1, t_2)$ such that $t_1 > \hat{t}_1$ and

$t_2 < \hat{t}_2$, we must have $p_1(t) = 1$. To see this, consider the point labeled $\alpha = (\tilde{t}_1, \tilde{t}_2)$ in Figure 1 below where $\tilde{t}_1 > \hat{t}_1$ while $\tilde{t}_2 < \hat{t}_2$. Suppose that at α, player 1 receives the good with probability strictly less than 1. It is not hard to see that at any point directly below α but above $\hat{t}_1$, such as the one labeled $\beta = (t_1', \tilde{t}_2)$, player 1 must receive the good with probability zero. To see this, note that if 1 did receive the good with strictly positive probability here, we could change the mechanism by lowering this probability slightly, giving the good to 2 at β with higher probability, and increasing the probability with which 1 receives the good at α. By choosing these probabilities appropriately, we do not affect $\hat{p}_2(\tilde{t}_2)$ so this remains at φ_2. Also, by making the reduction in p_1 small enough, $\hat{p}_1(t_1')$ will remain above φ_1. Hence this new mechanism would be feasible. Since it would switch probability from one type of player 1 to a higher type, the new mechanism would be better than the old one, implying the original one was not optimal.[5]

Similar reasoning implies that for every $t_1 \neq t_1^*$, we must have $\sum_i p_i(t_1, \tilde{t}_2) = 1$. Otherwise, the principal would be strictly better off increasing $p_2(t_1, \tilde{t}_2)$, decreasing $p_2(\tilde{t}_1, \tilde{t}_2)$, and increasing $p_1(\tilde{t}_1, \tilde{t}_2)$. Again, if we choose the sizes of these changes appropriately, $\hat{p}_2(\tilde{t}_2)$ is unchanged but $\hat{p}_1(\tilde{t}_1)$ is increased, an improvement.

[5]Since $\hat{p}_2(\tilde{t}_2)$ is unchanged, the ex ante probability of type $\tilde{t}_1$ getting the good goes up by the same amount that the ex ante probability of the lower type t_1' getting it goes down.

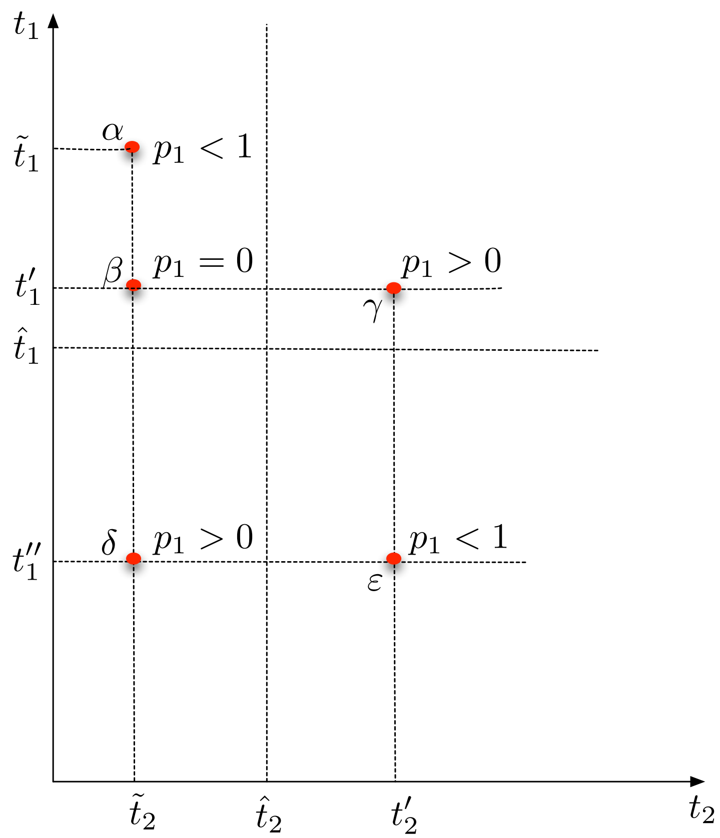

Figure 1

Since player 1 receives the good with zero probability at β but type t'_1 does have a positive probability overall of receiving the good, there must be some point like the one labeled $\gamma = (t'_1, t'_2)$ where 1 receives the good with strictly positive probability. We do not know whether t'_2 is above or below $\hat{t}_2$ — the position of γ relative to this cutoff plays no role in the argument to follow.

Finally, there must be a t''_1 (not necessary below $\hat{t}_1$) corresponding to points δ and ε where p_1 is strictly positive at δ and strictly less than 1 at ε. To see that such a t''_1 must exist, suppose not. Then for all $t''_1 \neq \tilde{t}_1$, either $p_1(t_1, \tilde{t}_2) = 0$ or $p_1(t_1, t'_2) = 1$. Since $\sum_i p_i(t_1, \tilde{t}_2) = 1$ for all $t_1 \neq \tilde{t}_1$, this implies that for all $t_1 \neq \tilde{t}_1$, either $p_2(t_1, \tilde{t}_2) = 1$

or $p_2(t_1, t_2') = 0$. Either way, $p_2(t_1, \tilde{t}_2) \geq p_2(t_1, t_2')$ for all $t_1 \neq t_1'$. But we also have $p_2(t_1', \tilde{t}_2) = 1 > 1 - p_1(t_1', t_2') \geq p_2(t_1', t_2')$. So $\hat{p}_2(\tilde{t}_2) > \hat{p}_2(t_2')$. But $\hat{p}_2(\tilde{t}_2) = \varphi_2$, so this implies $\hat{p}_2(t_2') < \varphi_2$, which violates the constraints on our optimization problem.

From this, we can derive a contradiction to the optimality of the mechanism. At γ, lower p_1 and increase p_2 by the same amount. At ε, raise p_1 and lower p_2 in such a way that $\hat{p}_2(t_2')$ is unchanged. In doing so, keep the reduction of p_1 at γ small enough that $\hat{p}_2(t_1')$ remains above φ_1. This is clearly feasible. Now that we have increased p_1 at ε, we can lower it at δ while increasing p_2 in such a way that $\hat{p}_1(t_1'')$ remains unchanged. Finally, since we have lowered p_1 at δ, we can increase it at α, now lowering p_2, in such a way that $\hat{p}_2(\tilde{t}_2)$ is unchanged.

Note the overall effect: $\hat{p}_1$ is unaffected at t_1'' and lowered in a way which retains feasibility at t_1'. $\hat{p}_2$ is unchanged at $\tilde{t}_2$ and at t_2'. Hence the resulting p is feasible. But we have shifted some of the probability that 1 gets the object from γ to α. Since 1's type is higher at α, this is an improvement, implying that the original mechanism was not optimal.

The fourth step is to show that $\hat{t}_1 = \hat{t}_2$. To see this, suppose to the contrary that $\hat{t}_2 > \hat{t}_1$. Then consider a type profile $t = (t_1, t_2)$ such that $\hat{t}_2 > t_2 > t_1 > \hat{t}_1$. From our second step, the fact that $t_2 > t_1 > \hat{t}_1$ implies $p_2(t) = 1$. However, from our third step, $t_1 > \hat{t}_1$ and $t_2 < \hat{t}_2$ implies $p_1(t) = 1$, a contradiction. Hence there cannot be any such profile of types, implying $\hat{t}_2 \leq \hat{t}_1$. Reversing the roles of the players then implies $\hat{t}_1 = \hat{t}_2$.

Let $t^* = \hat{t}_1 = \hat{t}_2$. This common value of these individual "thresholds" will yield the threshold of our favored–agent mechanism as we will see shortly.

To sum up the first four steps, we can characterize any optimal mechanism by specifying t^*, φ_1, and φ_2. From our second step, if we have $t_2 > t_1 > t^*$, then $p_2(t) = 1$. That is, if both agents are above the threshold, the higher type agent receives the object. From our third step, if $t_1 > t^* > t_2$, then $p_1(t) = 1$. That is, if only one agent is above the threshold, this agent receives the object. Either way, then, if there is at least one agent whose type is above the threshold, the agent with the highest type receives the object. Also, by definition, if $t_i < t^*$, then $\hat{p}_i(t_i) = \varphi_i = \inf_{t_i'} \hat{p}_i(t_i')$. Recall that we showed $\hat{q}_i(t_i) = \hat{p}_i(t_i) - \varphi_i$, so $\hat{q}_i(t_i) = 0$ whenever $t_i < t^*$. That is, if an agent is below the threshold, he receives the good with the lowest possible probability and is not checked.

This implies that $\hat{p}$ is completely pinned down as a function of t^*, φ_1. and φ_2. If $t_i > t^*$, then $\hat{p}_i(t_i)$ must be the probability $t_i > t_j$. If $t_i < t^*$, then $\hat{p}_i(t_i) = \varphi_i$. We already saw that $\hat{q}$ is pinned down by $\hat{p}$ and the φ_i's, so the reduced form is a function only of t^* and the φ_i's. Since we can write the principal's payoff can be written as a function only of the reduced form, this implies that the principal's payoff is completely pinned down once we specify t^* and the φ_i's. It is not hard to see that the principal's payoff is linear in the φ_i's. Because of this and the fact that the set of feasible φ vectors is convex, we see that given v^*, there must be a solution to the principal's problem at an extreme point of the set of feasible (φ_1, φ_2). Furthermore, every optimal choice of the φ's is a randomization over optimal extreme points.

The last step is to show that such extreme points correspond to favored–agent mechanisms. It is not hard to see that at an extreme point, one of the φ_i's is set to zero and the other is "as large as possible," where we clarify the meaning of this phrase below.[6] For notational convenience, consider the extreme point where $\varphi_2 = 0$ and φ_1 is set as high as possible. We now show that this corresponds to the favored–agent mechanism where 1 is the favored agent and the threshold v^* is $t^* - c$.

To see this, first note that the reduced form for an arbitrary t^*, φ_1, and φ_2 is

$$\hat{p}_i(t_i) = \begin{cases} \Pr[t_j < t_i], & \text{if } t_i > t^*, \\ \varphi_i, & \text{otherwise} \end{cases}$$

and $\hat{q}_i(t_i) = \hat{p}_i(t_i) - \varphi_i$. The reduced form of the favored–agent mechanism where 1 is favored, say (p^*, q^*), is given by

$$\hat{p}_1^*(t_1) = \begin{cases} \Pr[t_2 < t_1], & \text{if } t_1 > t^*, \\ \Pr[t_2 < t^*], & \text{otherwise,} \end{cases}$$

$$\hat{q}_1^*(t_1) = \begin{cases} \Pr[t_2 < t_1] - \Pr[t_2 < t^*], & \text{if } t_1 > t^*, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\hat{q}_2^*(t_2) = \hat{p}_2^*(t_2) = \begin{cases} \Pr[t_1 < t_2], & \text{if } t_2 > t^*, \\ 0, & \text{otherwise,} \end{cases}$$

It is easy to see that these are equal if we set $\varphi_1 = \Pr[t_2 < t^*]$ and $\varphi_2 = 0$. It turns out that $\Pr[t_2 < t^*]$ is what "as large as possible" means in this example.

6.2 Almost Dominance and Ex Post Incentive Compatibility

One appealing property of the favored–agent mechanism is that it is *almost* a dominant strategy mechanism. That is, for every agent, truth telling is a best response to *any* strategies by the opponents. It is not always a dominant strategy, however, as the agent may be completely indifferent between truth telling and lies.

To see this, consider any agent i who is not favored and a type t_i such that $t_i - c_i > v^*$. If t_i reports his type truthfully, then i receives the object with strictly positive probability

[6] Under some conditions, $\varphi_i = 0$ for all i is also an extreme point. See the appendix for details.

under a wide range of strategy profiles for the opponents. Specifically, any strategy profile for the opponents with the property that $t_i - c_i$ is the highest report for some type profiles has this property. On the other hand, if t_i lies, then i receives the object with zero probability given any strategy profile for the opponents. This follows because i is not favored and so cannot receive the object without being checked. Hence for such a type, truth telling weakly dominates any lie for t_i.

Continuing to assume i is not favored, consider any t_i such that $t_i - c_i < v^*$. For *any* profile of strategies by the opponents, t_i's probability of receiving the object is zero regardless of his report. To see this, simply note that if i reports truthfully, he cannot receive the good (since it will either go to another nonfavored agent if one has the highest $t_j - c_j$ and reports honestly or to the favored agent). Similarly, if i lies, he cannot receive the object since he will be caught lying when checked. Hence truth telling is *an* optimal strategy for t_i, though it is not weakly dominant since the agent is indifferent over all reports given any strategies by the other agents.

A similar argument applies to the favored agent. Again, if his type satisfies $t_i - c_i > v^*$, truth telling is dominant, while if $t_i - c_i < v^*$, he is completely indifferent over all strategies. Either way, truth telling is an optimal strategy regardless of the strategies of the opponents.

Because of this property, the favored–agent mechanism is ex post incentive compatible. Formally, (p, q) is ex post incentive compatible if

$$p_i(t) \geq p_i(\hat{t}_i, t_{-i}) - q_i(\hat{t}_i, t_{-i}), \quad \forall \hat{t}_i, t_i \in T_i, \quad \forall t_{-i} \in T_{-i}, \quad \forall i \in \mathcal{I}.$$

That is, t_i prefers reporting honestly to lying even conditional on knowing the types of the other agents. It is easy to see that the favored–agent mechanism's almost–dominance property implies this. Of course, the ex post incentive constraints are stricter than the Bayesian incentive constraints, so this implies that that favored–agent mechanism is ex post optimal.

While the almost–dominance property implies a certain robustness of the mechanism, the complete indifference for types below the threshold is troubling. There are simple modifications of the mechanism which do not change its equilibrium properties but make truth telling weakly dominant rather than just almost dominant. For example, suppose there are at least three agents and that every agent i satisfies $\bar{t}_i - c_i > v^*$.[7] Suppose we modify the favored agent mechanism as follows. If an agent is checked and found to have lied, then one of the other agents is chosen at random and his report is checked. If it is truthful, he receives the object. Otherwise, no agent receives it. It is easy to see

[7] Note that if $\bar{t}_i - c_i < v^*$, then the favored agent mechanism *never* gives the object to i, so i's report is entirely irrelevant to the mechanism. Thus we cannot make truth telling dominant for such an agent, but the report of such an agent is irrelevant anyway.

that truth telling is still an optimal strategy and that the outcome is unchanged if all agents report honestly. It is also still weakly dominant for an agent to report the truth if $t_i - c_i > v^*$. Now it is also weakly dominant for an agent to report the truth even if $t_i - c_i < v^*$. To see this, consider such a type and assume i is not favored. Then if t_i lies, it is impossible for him to receive the good regardless of the strategies of the other agents. However, if he reports truthfully, there is a profile of strategies for the opponents where he has a strictly positive probability of receiving the good — namely, where one of the nonfavored agents lies and has the highest report. Hence truth telling weakly dominates any lie. A similar argument applies to the favored agent.

6.3 Extension: When Verification is Costly for Agent

A natural extension to consider is when the process of verifying an agent's claim is also costly for that agent. In our example where the principal is a dean and the agents are departments, it seems natural to say that departments bear a cost associated with providing documentation to the dean.

The main complication associated with this extension is that the agents may now trade off the value of obtaining the object with the costs of verification. An agent who values the object more highly would, of course, be willing to incur a higher expected verification cost to increase his probability of receiving it. Thus the simplification we obtain where we can treat the agent's payoff as simply equal to the probability he receives the object no longer holds.

On the other hand, we can retain this simplification at the cost of a stronger assumption. To be specific, we can simply assume that the value to the agent of receiving the object is 1 and the value of not receiving it is 0, regardless of his type. For example, if the principal is a dean and the agents are academic departments, this assumption holds if each department's payoff to getting the job slot independent of the value they would produce for the dean.[8] If we make this assumption, the extension to verification costs for the agents is straightforward. We can also allow the cost to the agent of being verified to differ depending on whether the agent lied or not. To see this, let $\hat{c}_i^T$ be the cost incurred by agent i from being verified by the principal if he reported his type truthfully and let $\hat{c}_i^F$ be his cost if he lied. We assume $1 + \hat{c}_i^F > \hat{c}_i^T \geq 0$. (Otherwise, verification costs hurt honest types more than dishonest ones.) Then the incentive compatibility condition

[8]More formally, think of an agent's type as a pair (s_i, t_i) where s_i is the value of the good to the agent and t_i is the value to the principal of giving the good to the agent. Normalize the value to the agent of not receiving the good to be zero. Then if s_i and t_i are independently distributed and $s_i > 0$ with probability 1, we can renormalize the agent's payoff to getting the good to be 1 for every s_i, independent of t_i.

becomes

$$\hat{p}_i(t_i') - \hat{c}_i^T \hat{q}_i(t_i') \geq \hat{p}_i(t_i) - \hat{c}_i^F \hat{q}_i(t_i) - \hat{q}_i(t_i), \ \forall t_i, t_i', \ \forall i.$$

Let

$$\varphi_i = \inf_{t_i'} [\hat{p}_i(t_i') - \hat{c}_i^T \hat{q}_i(t_i')],$$

so that incentive compatibility holds iff

$$\varphi_i \geq \hat{p}_i(t_i) - \hat{c}_i^F \hat{q}_i(t_i) - \hat{q}_i(t_i), \ \forall t_i, \ \forall i.$$

Analogously to the way we characterized the optimal mechanism in Section 6.1, we can treat φ_i as a separate choice variable for the principal where we add the constraint that $\hat{p}_i(t_i') - \hat{c}_i^T \hat{q}_i(t_i') \geq \varphi_i$ for all t_i'.

Given this, $\hat{q}_i(t_i)$ must be chosen so that the incentive constraint holds with equality for all t_i. To see this, suppose to the contrary that we have an optimal mechanism where the constraint holds with strict inequality for some t_i (more precisely, some positive measure set of t_i's). If we lower $\hat{q}_i(t_i)$ by ε, the incentive constraint will still hold. Since this increases $\hat{p}_i(t_i') - \hat{c}_i^T \hat{q}_i(t_i')$, the constraint that this quantity is greater than φ_i will still hold. Since auditing is costly for the principal, his payoff will increase, implying the original mechanism could not have been optimal, a contradiction.

Since the incentive constraint holds with equality for all t_i, we have

$$\hat{q}_i(t_i) = \frac{\hat{p}_i(t_i) - \varphi_i}{1 + \hat{c}_i^F}.$$

Substituting, this implies that

$$\varphi_i = \inf_{t_i'} \left[\hat{p}_i(t_i') - \frac{\hat{c}_i^T}{1 + \hat{c}_i^F} [\hat{p}_i(t_i) - \varphi_i] \right]$$

or

$$\varphi_i = \inf_{t_i'} \left[\left\{ 1 - \frac{\hat{c}_i^T}{1 + \hat{c}_i^F} \right\} \hat{p}_i(t_i) + \frac{\hat{c}_i^T}{1 + \hat{c}_i^F} \varphi_i \right].$$

By assumption, the coefficient multiplying $\hat{p}_i(t_i')$ is strictly positive, so this is equivalent to

$$\left\{ 1 - \frac{\hat{c}_i^T}{1 + \hat{c}_i^F} \right\} \varphi_i = \left\{ 1 - \frac{\hat{c}_i^T}{1 + \hat{c}_i^F} \right\} \inf_{t_i'} \hat{p}_i(t_i'),$$

so $\varphi_i = \inf_{t_i'} \hat{p}_i(t_i')$, exactly as in our original formulation.

The principal's objective function is

$$\begin{aligned}
\mathrm{E}_t \sum_i [p_i(t)t_i - c_i q_i(t)] &= \sum_i \mathrm{E}_{t_i}[\hat{p}_i(t_i)t_i - c_i \hat{q}_i(t_i)] \\
&= \sum_i \mathrm{E}_{t_i}\left[\hat{p}_i(t_i)t_i - \frac{c_i}{1+\hat{c}_i^F}[\hat{p}_i(t_i) - \varphi_i]\right] \\
&= \sum_i \mathrm{E}_{t_i}\left[\hat{p}_i(t_i)(t_i - \tilde{c}_i) + \varphi_i \tilde{c}_i\right]
\end{aligned}$$

where $\tilde{c}_i = c_i/(1 + \hat{c}_i^F)$. This is the same as the principal's objective function in our original formulation but with $\tilde{c}_i$ replacing c_i.

In short, the solution changes as follows. The allocation probabilities p_i are exactly the same as what we characterized but with $\tilde{c}_i$ replacing c_i. The checking probabilities, however, are the earlier ones divided by $1 + \hat{c}_i^F$. Intuitively, since verification imposes costs on the agent in this model, the threat of verification is more severe than in the previous model, so the principal doesn't need to check as often. In short, the new optimal mechanism is still a favored agent mechanism but where the checking which had probability 1 before now has probability $1/(1 + \hat{c}_i^F)$. The optimal choice of the favored agent and the optimal threshold is exactly as before with $\tilde{c}_i$ replacing c_i. Note that agents with low values of $\hat{c}_i^F$ have higher values of $\tilde{c}_i$ and hence are more likely to be favored. That is, agents who find it easy to undergo an audit after lying are more likely to be favored. Note also that $\hat{c}_i^T$ has no effect on optimal mechanism.

7 Conclusion

There are many natural extensions to consider. For example, in the previous subsection, we discussed the extension in which the agents bear some costs associated with verification, but under the restriction that the value to the agent of receiving the object is independent of his type. A natural extension of interest would be to drop this restriction.[9]

A second natural extension would be to allow costly monetary transfers. We argued in the introduction that within organizations, monetary transfers are costly to use and hence have excluded them from the model. It would be natural to model these costs explicitly and determine to what extent the principal allows inefficient use of some resources to obtain a better allocation of other resources.

Another direction to consider is to generalize the nature of the principal's allocation problem. For example, what is the optimal mechanism if the principal has to allocate

[9]See Ambrus and Egorov [2012] for an analysis of how imposing costs on agents can be useful for the principal in a setting without verification.

some *tasks*, as well as some resources? In this case, the agents may prefer to *not* receive certain "goods." Alternatively, there may be some common value elements to the allocation in addition to the private values aspects considered here.

Another natural direction to consider is alternative specifications of the information structure and verification technology. Here each agent knows exactly what value he can create for the principal with the object. Alternatively, the principal may have private information which determines how he interprets an agent's information. Also, it is natural to consider the possibility that the principal partially verifies an agent's report, choosing how much detail to go into. For example, an employer dealing with a job applicant can decide how much checking to do of the applicant's resume and references.

References

[1] Ambrus, A., and G. Egorov, "Delegation and Nonmonetary Incentives," Duke University working paper, July 2012.

[2] Banerjee, A., R. Hanna, and S. Mullainathan, "Corruption," working paper, January 2011.

[3] Border, K., and J. Sobel, "Samurai Accountant: A Theory of Auditing and Plunder," *Review of Economic Studies*, **54**, October 1987, 525–540.

[4] Bull, J., and J. Watson, "Hard Evidence and Mechanism Design," *Games and Economic Behavior*, **58**, January 2007, 75–93.

[5] Deneckere, R. and S. Severinov, "Mechanism Design with Partial State Verifiability," *Games and Economic Behavior*, **64**, November 2008, 487–513.

[6] Gale, D., and M. Hellwig, "Incentive–Compatible Debt Contracts: The One–Period Problem," *Review of Economic Studies*, **52**, October 1985, 647–663.

[7] Glazer, J., and A. Rubinstein, "On Optimal Rules of Persuasion," *Econometrica*, **72**, November 2004, 1715–1736.

[8] Glazer, J., and A. Rubinstein, "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach," *Theoretical Economics*, **1**, December 2006, 395–410.

[9] Green, J., and J.-J. Laffont, "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, **53**, July 1986, 447–456.

[10] Mookherjee, D., and I. Png, "Optimal Auditing, Insurance, and Redistribution," *Quarterly Journal of Economics*, **104**, May 1989, 399–415.

[11] Sher, I., and R. Vohra, "Price Discrimination through Communication," working paper, June 2011.

[12] Townsend, R., "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, **21**, 1979, 265–293.