University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 266

Delegating Performance Evaluation

Igor Letina, Shuo Liu and Nick Netzer

October 2017

Delegating Performance Evaluation

Igor Letina, Shuo Liu and Nick Netzer*

This Version: October 2017
First Version: November 2016

Abstract

We study optimal incentive contracts with multiple agents when performance evaluation is delegated to a reviewer. The reviewer may be biased in favor of the agents, but the degree of the bias is unknown to the principal. We show that a contest, which is a contract in which the principal determines a set of prizes to be allocated to the agents, is optimal. By using a contest, the principal can commit to sustaining incentives despite the reviewer's potential leniency bias. The optimal effort profile can be uniquely implemented by an all-pay auction with a cap, and it can also be implemented by a nested Tullock contest. Our analysis has implications for applications as diverse as the design of worker compensation, the awarding of research grants, and the allocation of foreign aid.

Keywords: performance evaluation, delegation, optimality of contests
JEL: D02, D82, M52

---

*Letina: Department of Economics, University of Bern, Switzerland. Liu and Netzer: Department of Economics, University of Zurich, Switzerland. Email: igor.letina@vwi.unibe.ch, shuo.liu@econ.uzh.ch, and nick.netzer@econ.uzh.ch. We are grateful for useful comments and suggestions to Jean-Michel Benkert, Elchanan Ben-Porath, Lorenzo Casaburi, Yeon-Koo Che, David Dorn, Christian Ewerhart, Sergiu Hart, Andreas Hefti, Johannes Hörner, Navin Kartik, Christian Kellner, Botond Kőszegi, Wladislaw Mill, Harry Pei, Ron Siegel, Florian Scheuer, Armin Schmutzler, Philipp Strack, Dezsö Szalay, and seminar participants at Bar Ilan University, Carnegie Mellon University, Columbia University, Hebrew University of Jerusalem, Toulouse School of Economics, the Universities of Bern, Cologne and Zurich, 20th Colloquium on Personnel Economics, 28th International Conference on Game Theory, SIOE Annual Conference 2017, SIRE Workshop on Recent Developments in Behavioural Economics and Mechanism Design, Swiss Economists Abroad Meeting 2016, Swiss Theory Day 2016, Verein für Socialpolitik Meeting 2016, Verein für Socialpolitik Theoretischer Ausschuss 2017, Zurich Workshop in Economics 2016, and Zurich Workshop on Applied Behavioral Theory 2017. Shuo Liu would like to acknowledge the hospitality of Columbia University, where some of this work was carried out, and financial support by the Swiss National Science Foundation (Doc. Mobility grant P1ZHP1_168260).

1 Introduction

Principals often lack the information or expertise needed to make appropriate decisions. A common response to this problem is to delegate the decision to a better informed party. For example, funding agencies delegate the choice of research projects which will be funded to an expert committee. Within a firm, the CEO usually delegates to a mid-level manager the decision regarding the assignment of bonuses to subordinates. Humanitarian aid is distributed by specialized agencies on behalf of the donor countries.

If the preferences of the principal and the expert who makes the decision are not aligned, then delegation can lead to distorted decisions. The principal can attempt to influence the decision by limiting the set of outcomes from which the expert can select. The existing literature on optimal delegation studies how this *delegation set* should be designed.[1] In these papers, the principal wants to base her decision on some stochastic state of nature, the value of which is known only to the expert. Crucially, this state of nature is assumed to be exogenous. However, in the examples above, the state of nature (the quality of research projects, the performance of employees, the cooperativeness of receiving countries) is determined in part in anticipation of the decision that the expert will make. As a matter of fact, the goal of the principal is exactly to incentivize the agents to exert effort. For example, the goal of the funding agencies is to stimulate creation of high quality research. Similarly, bonuses in firms are instruments that incentivize employees to work hard, and aid is in part allocated to bring about reforms.

In this paper, we study the optimal delegation problem for performance evaluation. A *principal* wishes to incentivize *agents* to exert costly effort. The efforts are not observable to the principal. However, an expert, which we will from now on refer to as the *reviewer*, can costlessly observe the exerted efforts.[2] The principal thus delegates the decision on how to reward the agents to the reviewer, but possibly restricts the set of allowable decisions. The reviewer's preferences may not be perfectly aligned with the principal. While the reviewer takes into account the effect of his actions on the principal's payoff, maybe because he owns shares of the company or he cares intrinsically, he may also care about the agents. For instance, as we will discuss below, there is ample evidence that managers care about the payoffs of their subordinates. Exactly how much the reviewer cares about the agents is the reviewer's private information. Importantly, a reviewer who cares sufficiently much about the agents will be reluctant to punish them even if they do not exert sufficient effort. Anticipating this, the agents will exert less effort. The principal thus has to design the delegation set in a way that restricts the scope of possible leniency of the reviewer.

[1]See, for example, Holmström (1977, 1984), Melumad and Shibano (1991), Alonso and Matouschek (2008), Armstrong and Vickers (2010), Amador and Bagwell (2013), and Frankel (2014).

[2]We do not consider the problem of incentivizing the reviewer to exert costly effort in order to learn the state of nature. This is an interesting but distinct incentive problem which is studied in Aghion and Tirole (1997), Strausz (1997), Szalay (2005), Rahman (2012), and Pei (2015b).

One could also imagine that the principal tries to correct the distortions by paying transfers to the reviewer conditional on the action that he takes. However, contingent transfers are often not observed in reality. Committees deciding which research projects get funded are not paid conditional on how many projects they approve or reject. Mid-level managers do not get paid differently depending on how they allocate bonuses among their subordinates. In fact, paying an expert for performing a particular evaluation is often referred to as a conflict of interest and is explicitly forbidden. The delegation approach, which rules out direct monetary incentives, is therefore particularly plausible for our setting of performance evaluation.[3]

Our first main result is that a contest among the agents is an optimal delegation mechanism. That is, the principal defines a set of prizes and the reviewer only decides how to allocate these prizes to the agents. The reviewer does not have the additional freedom to choose the overall size or the split of the agents' compensation. This strongly limits the degree of leniency he can exercise. In particular, the reviewer is always forced to punish some agents by assigning them a small prize, which is crucial for the preservation of incentives. Without this commitment, the reviewer would be lenient and the agents would shirk. The downside of the contest mechanism is that a small prize has to be assigned (at random) even when all agents provide the sufficient level of effort. When the agents are risk-averse, this will be inefficient.

This result is interesting for several reasons. First, contests are a commonly used and often-studied incentive scheme, and much work has been devoted to their optimal design.[4] There is, however, not much work on the more general question whether and under which conditions contests are actually optimal mechanisms.[5] Exceptions are the seminal paper of Lazear and Rosen (1981), as well as papers which stress that contests can filter out common shocks when agents are risk-averse (Green and Stokey, 1983; Nalebuff and Stiglitz, 1983) or ambiguity-averse (Kellner, 2015). In our model, an important reason for contests to be optimal is that they act as a commitment device. A contest provides two types of commitment. It commits the principal to the announced prizes and thus prevents manipulation of the sum of payments made to the agents. The literature has observed previously that this "commitment to pay" can be beneficial when the agents' efforts are

[3]The absence of monetary transfers in delegation mechanisms mirrors the incentive structure assumed in models of cheap talk (Crawford and Sobel, 1982) or bayesian persuasion (Kamenica and Gentzkow, 2011). Another instrument that the delegation approach rules out in our setting is the possibility of direct communication between the principal and the agents.

[4]For instance, Glazer and Hassin (1988) and Moldovanu and Sela (2001) study the design of prizes for a given contest success function. Jia, Skaperdas, and Vaidya (2013) survey papers that study the design of a contest success function for given prizes.

[5]Prendergast (1999, p. 36) writes: "Rather surprisingly, there is very little work devoted to understanding why this is the case, i.e., why the optimal means of providing incentives within large firms (at least for white-collar workers) seems to be tournaments rather than the other means suggested in the previous sections." We find this still to be the case in the years since Prendergast published his paper.

not verifiable. For instance, Malcomson (1984, 1986) argues that piece-rate contracts are not credible in that case, as the principal would always claim low performance ex post in order to reduce payments, while a contest remains credible.[6] However, this credibility can also be achieved by simply committing to a total sum of payments without setting fixed prizes. In fact, as we will show, such a scheme would outperform a contest when the agents are risk-averse, by removing uncertainty from equilibrium payments.[7] Hence the second type of commitment, the above described "commitment to punish," is crucial in explaining the optimality of contests with fixed prizes.[8] Second, a contest is a remarkably simple mechanism. Even though we allow for arbitrary stochastic delegation mechanisms with possibly sophisticated transfer rules, the optimum can be achieved by a simple mechanism characterized by a prize profile and a suggestion how to distribute the prizes in response to the agents' efforts. The principal does not attempt to screen the reviewer's type, in spite of the fact that the first-best may be achievable if the type were known to the principal. This makes the strategic considerations of the agents simple. In particular, their behavior does not depend on beliefs regarding the reviewer's type. This robustness property is important because principal and agents may well have different beliefs (for instance, in the example with managers allocating bonuses, it seems reasonable to assume that the employees working directly with a manager have more precise information about their manager's type than a CEO does).

Our second result characterizes the prize structure of an optimal contest. Given n agents, an optimal contest will have $n-1$ equal positive prizes and one zero prize. Thus, while the contest acts as a commitment to punish, the punishment is kept at the minimum required to incentivize effort. The delegation set forces the reviewer to allocate a zero prize to only one agent, so that the optimal contest exhibits a "loser-takes-nothing" rather than a "winner-takes-all" structure. In equilibrium, when all agents have provided sufficient effort, the reviewer randomly chooses the agent who receives the zero prize. Thus all agents are facing the same risk. If agents are risk-averse, they respond to this risk by reducing the amount of effort they are willing to exert. A corollary of this result is that the first-best

[6]Similarly, Carmichael (1983) considers a setting where the final output is verifiable but depends on the efforts of both the principal and the agents. With a contract that pays agents based on total output, the principal has an incentive to reduce own effort in order to reduce the payments to agents.

[7]Levin (2002) shows that a constant sum of payments arises in equilibrium of a repeated game, where the principal's wage promises have to be self-enforcing. A contest-type compensation structure arises in an optimal equilibrium only if the agents are risk-neutral.

[8]The problem of committing to punishment is related to Konrad (2001) and Netzer and Scheuer (2010). They study the problem of a planner who would like to implement redistribution after agents have chosen their actions, the anticipation of which may destroy incentives to choose costly but socially desirable actions. In the context of optimal income taxation, Konrad (2001) shows that private information about labor productivity provides a commitment against excessive redistribution. In the context of insurance and labor markets, Netzer and Scheuer (2010) show that adverse selection provides a commitment by generating separating market equilibria. In both cases, agents who choose socially less desirable actions are punished by having to forego information rents.

is implementable if and only if the agents are risk-neutral. We also show that an optimal contest implements an outcome close to the first-best if the agents' risk-aversion is moderate or the number of agents is large.

Our third result shows that a familiar all-pay auction with a cap implements the optimum, and it does so in unique equilibrium.[9] As in a standard all-pay auction with $n-1$ identical prizes, the agent with the lowest effort receives the zero prize, and ties are broken randomly. However, efforts are capped at the desired equilibrium level. This removes the possibility for the agents to exert slightly more than the equilibrium effort in order to guarantee themselves a positive prize with probability one. In addition to the all-pay auction, we show as our fourth result that the optimum can also be achieved with an imperfectly discriminating contest, such as the well-known Tullock contest. The Tullock-type contest success function arises endogenously as part of an optimal contract in our analysis. In summary, the optimum can be achieved with several of the commonly studied formats of contests (see Konrad, 2009). This shows that the essential feature of our main result is the fixed profile of prizes, and not the exact form of the contest success function.

We then consider extensions where the reviewer only imperfectly observes individual efforts of the agents. We show that, by using stochastic allocation rules, the principal can often still implement the optimal allocation with a contest. Next, we discuss how our analysis can be extended to settings with non-separable preferences and favoritism of the reviewer. We also generalize the model to allow for heterogeneous effort cost functions and show that our main results remain robust. Finally, we consider a model of cheap talk with limited commitment (see Kolotilin, Li, and Li, 2013) where the reviewer does not make the allocation decision but communicates the observed effort levels to the principal. We again show that our results for the delegation model continue to hold in the cheap talk model.

Our contribution is related to three strands of literature: on the optimality of contests, on biased reviewers, and on optimal delegation. A more detailed discussion of this literature is postponed to Section 5. A closely related paper is Frankel (2014). Like in our paper, he considers a multidimensional delegation problem with uncertainty about the expert's preferences. He assumes that the state of the world is exogenous and not affected by the choice of the delegation mechanism. In contrast, our prime concern is how the delegation mechanism affects the state of the world, i.e., how it provides incentives for agents to exert effort. Another difference is that Frankel (2014) derives max-min mechanisms, which are optimal for the worst possible realization of the expert's bias, while we are interested in mechanisms that maximize the principal's expected payoff given her beliefs about the expert's type. Frankel (2014) shows that, when the set of possible preferences of the expert

[9] As we will explain in Section 3, uniqueness only refers to the agents' choice of efforts in the given contest. The reviewer will be indifferent among several actions, and in particular a "babbling" equilibrium exists where his assignment of prizes is unresponsive to the agents' efforts and they exert zero effort.

is rich enough, a ranking mechanism is max-min optimal. When the multidimensional state of the world is an effort profile, like in our setting, this ranking mechanism would correspond to a standard all-pay auction (see Siegel, 2009, for a general treatment of contests with all-pay structure). Such an all-pay auction without a cap is not optimal in our setting with endogenous efforts, but there is a range of different contests that are optimal. Furthermore, since the principal's payoff turns out to be independent of the reviewer's type in these optimal contests, they not only maximize expected payoffs but are also max-min optimal. Another closely related paper is Strausz (1997), who considers the delegation of performance evaluation for a single agent. Monitoring is costly and produces verifiable signals in his setting. Strausz (1997) shows that delegation generates optimal incentives for costly monitoring and the disclosure of signals.

Our model applies to many situations where a principal wants to incentivize agents but cannot directly supervise them. Here we will discuss two possible applications, which are meant to illustrate the range and scale of our model. One application is the design of performance evaluation schemes in firms. The performance evaluation scheme is designed by the CEO, but the CEO does not observe the individual efforts of the employees to which the scheme applies. Hence, the actual performance evaluation is delegated to the employees' supervisor. By virtue of working closely with the employees, the supervisor observes their efforts but also cares about their payoffs. There is ample evidence (both empirical and experimental) that supervisors tend to be too lenient when judging the performance of their subordinates, and that the degree of leniency varies and depends on (among other things) social ties between the supervisor and the team.[10] Our results have direct implications for the controversial debate over the use of the so-called "forced rankings," a review system which was most famously used by General Electric under Jack Welch during their fast growth in the 1980s and 90s.[11] Our contribution to this discussion is (*i*) to show that forced rankings are optimal for motivating effort under the assumptions of our model, (*ii*) to show how optimal forced rankings should be constructed, and (*iii*) to show that some elements of forced rankings which are usually criticized are actually necessary for incentivizing effort. In particular, forced rankings are criticized for forcing managers to assign low rankings even when all workers are performing well: "What happens if you're working with a superstar team? You've just forced a distribution that doesn't

[10]For example, Bol (2011) and Breuer, Nieken, and Sliwka (2013) find evidence of leniency bias which depends on the strength of the employee-manager relationship. Bol (2011) cites studies documenting leniency bias going back to the 1920s, while citations to similar findings in the 1940s can be found in Prendergast (1999). Berger, Harbring, and Sliwka (2013) find experimental evidence of leniency bias, and Bernardin, Cooke, and Villanova (2000) document that the degree of leniency bias in an experiment depends on personality traits of the reviewer. More generally, Cappelen, Hole, Sørensen, and Tungodden (2007) show experimentally that individuals exhibit a variety of different fairness preferences.

[11]For example, see "'Rank and Yank' Retains Vocal Fans" (L. Kwoh, *The Wall Street Journal*, January 31, 2012) and "For Whom the Bell Curve Tolls" (J. McGregor, *The Washington Post*, November 20, 2013).

exist. You create this stupid world where [great] people are punished."[12] Similarly, Brad Smart who worked with Jack Welch on developing GE's forced ranking system criticized GE's decision to assign 10% of the workers a low evaluation: "To force those distributions when the percentages don't meet the reality is nuts."[13] Our results show that, far from being "stupid" or "nuts," not rewarding some workers even when they perform well is necessary, since if the managers were given an option to reward all workers, they may choose it irrespective of actual performance, which would destroy any incentive effect of the evaluation system.

A very different situation for which our model offers insights is foreign aid. Donors have been trying for decades to use foreign aid to incentivize reforms in recipient countries, but there is little empirical evidence that it has been effective (see e.g. Easterly, 2003; Rajan and Subramanian, 2008). In response, funding agencies and governments have tried to improve mechanisms for the allocation of foreign aid in ways that link aid to improvement in governance and other policy reforms. One early approach has been the so-called "conditional aid," where donors promise to withdraw future aid if the agreed policy reforms have not been achieved. However, the donors' threats to withdraw aid were not credible and, unsurprisingly, reforms were usually not carried out. As Easterly (2009) somewhat amusingly points out, the World Bank conditioned aid on the same agricultural policy reform in Kenya five separate times – and the conditions were violated each time. Svensson (2003) proposes a solution to this problem. Instead of allocating the budget for each country separately, similar countries could be pooled together and the total budget for all these countries could be allocated to a single aid officer. This way, if one country does not reform, the aid officer has the option of reallocating the aid from that country to another. Our paper points to a potential problem with this approach and offers a solution. A benevolent aid officer may still be tempted to split the aid more or less equally among the countries, their efforts towards reform notwithstanding. Our paper suggests that holding a contest among recipient countries can overcome this problem. That is, instead of giving the aid officer full discretion over the total budget for multiple countries, the budget could be partitioned into fixed "prizes" that the officer allocates to the countries. Obviously, it may be politically difficult to implement a contest where a country receives zero aid even if it invested effort in reforms. However, some variant of our mechanism, where all countries receive aid but some countries receive "bonus aid" through a contest might be both politically feasible and desirable from the incentive point of view.

The remainder of the paper is organized as follows. Section 2 describes the model. In Section 3 we show that the set of optimal contracts contains a contest, we characterize

[12]Quote of a management adviser in "For Whom the Bell Curve Tolls" (J. McGregor, *The Washington Post*, November 20, 2013).

[13]In "'Rank and Yank' Retains Vocal Fans" (L. Kwoh, *The Wall Street Journal*, January 31, 2012).

all optimal contests, and we deal with different implementations of the optimal outcome. In Section 4 we develop several extensions of the baseline model. Section 5 contains a discussion of the related literature, and Section 6 concludes. All proofs are relegated to Appendices A and B.

2 The Model

2.1 Environment

A principal contracts with a set of agents $I = \{1, ..., n\}$ where $n \geq 2$. Each agent $i \in I$ chooses an effort level $e_i \geq 0$ and obtains a monetary transfer $t_i \geq 0$. The agents have an outside option of zero. The payoff of agent i is given by

$$\pi_i(e_i, t_i) = u(t_i) - c(e_i).$$

The utility function $u : \mathbb{R}_+ \to \mathbb{R}$ is twice differentiable, strictly increasing, weakly concave, and satisfies $u(0) = 0$. The cost function $c : \mathbb{R}_+ \to \mathbb{R}$ is twice differentiable, strictly increasing, strictly convex, and satisfies $c(0) = 0$, $c'(0) = 0$, and $\lim_{e_i \to \infty} c'(e_i) = \infty$. The assumption of additive separability of transfers and efforts is standard in contract theory, mechanism design, and contest theory. We will discuss the robustness of our results with respect to non-separable and also asymmetric preferences in Section 4.

We denote effort profiles by $e = (e_1, ..., e_n) \in E$ and transfer profiles by $t = (t_1, ..., t_n) \in T$. We assume that $E = \mathbb{R}^n_+$ and $T = \{t \in \mathbb{R}^n_+ \mid \sum_{i=1}^n t_i \leq \bar{T}\}$, where $\bar{T} > 0$ can be arbitrarily large. Our results hold no matter whether $\bar{T}$ is eventually binding or not. The payoff of the principal from an allocation (e, t) is

$$\pi_P(e, t) = z(e) - \sum_{i=1}^n t_i,$$

where $z : E \to \mathbb{R}_+$ is interpreted as the production function that converts efforts into output. For clarity of exposition we will focus only on the case where $z(e) = \sum_{i=1}^n e_i$. Our main results continue to hold if we assume more generally that z is symmetric, weakly concave, and strictly increasing in each of its arguments.

Example. We will use a parameterized example to illustrate our results throughout the paper. In this example, each of the n agents has the payoff function

$$\pi_i(e_i, t_i) = t_i^\alpha - \gamma e_i^\beta,$$

where $0 < \alpha \leq 1$ parameterizes risk-aversion, $\beta > 1$ describes the degree of cost convexity, and $\gamma > 0$ determines the relative weight of effort costs. We will always assume that $\bar{T}$ is large enough to be non-binding in the example. The first-best effort level e^{FB} is what the principal would demand from each agent if she could perfectly control effort and would only have to compensate the agent for his cost, thus paying $t^{FB} = u^{-1}(c(e^{FB}))$. In our example, maximization of $e - u^{-1}(c(e))$ yields

$$e^{FB} = \left(\frac{\alpha}{\beta\gamma^{1/\alpha}}\right)^{\frac{\alpha}{\beta-\alpha}} \text{ and } t^{FB} = \left(\frac{\alpha}{\beta\gamma^{1/\beta}}\right)^{\frac{\beta}{\beta-\alpha}}.$$

The principal's first-best profit is $n(e^{FB} - t^{FB})$. □

The effort exerted by the agents is not verifiable to outside parties (e.g. a court) and is not observable to the principal. However, the efforts can be observed by a reviewer. Consequently, the evaluation of the agents' performance and the decision on how to reward the agents are delegated to the reviewer. In line with the literature on delegation, we assume that the principal does not pay the reviewer based on the decision made (but can in principle pay a fixed fee to the reviewer, which we normalize to zero). Specifically, the payoff of the reviewer from an allocation (e, t) is given by

$$\pi_R(e, t, \theta) = \pi_P(e, t) + \theta \sum_{i=1}^{n} \pi_i(e, t),$$

where θ is a parameter that captures how much the reviewer cares about the well-being of the agents, and thus by how much the reviewer's preferences are misaligned with those of the principal. The parameter θ can be thought of as a fundamental preference or as a reduced-form representation of concerns due to other interactions with the agents. We assume that θ is private information of the reviewer, observable neither to the principal nor to the agents. It is drawn according to a commonly known continuous distribution with full support on $\Theta = [\underline{\theta}, \bar{\theta}]$, where $\underline{\theta} < \bar{\theta}$. We describe this distribution by an (absolutely continuous) probability measure τ over Θ. Our results will be independent of the shape as well as the location of this distribution. In particular, τ could be arbitrarily close to a probability measure with atoms, and our results would also continue to hold if the support Θ was a union of countably many disjoint intervals.

2.2 Implementation with Credible Contracts

The timing is as follows. First, the principal delegates the evaluation and remuneration of the agents to the reviewer by designing a set D of possible actions. An action is a probability measure $\mu \in \Delta T$ on the set of transfer profiles, describing the potentially stochastic payments made to the agents. Given the delegation set, the agents choose their efforts simultaneously. The reviewer then observes the efforts (and his type) and chooses an action from D to reward or punish the agents.[14]

Since e and θ are observable only to the reviewer, he is always free to choose any action that he prefers. We model this by defining a *contract* $\Phi = (\mu^{e,\theta})_{(e,\theta)\in E\times\Theta}$ as a collection of probability measures $\mu^{e,\theta} \in \Delta T$, one for each $(e,\theta) \in E \times \Theta$. The interpretation is that the principal suggests that a reviewer of type θ should reward an effort profile e by transfers according to $\mu^{e,\theta}$.[15] The following incentive constraint makes sure that the reviewer indeed has an incentive to follow this suggestion:

$$\Pi_R(e, \mu^{e,\theta}, \theta) \geq \Pi_R(e, \mu^{e',\theta'}, \theta) \quad \forall (e,\theta), (e',\theta') \in E \times \Theta, \tag{IC-R}$$

where

$$\Pi_R(e, \mu^{e',\theta'}, \theta) = \mathbb{E}_{\mu^{e',\theta'}}\left[\pi_P(e,t) + \theta \sum_{i=1}^{n} \pi_i(e_i, t_i)\right].$$

We say that a contract Φ is *credible* if it satisfies (IC-R). Given a credible contract, the delegation set is implicitly given by $D = \{\mu \in \Delta T \mid \exists (e,\theta) \ s.t. \ \mu = \mu^{e,\theta}\}$.

Denote by $\sigma_i \in \Delta\mathbb{R}_+$ agent i's mixed strategy for his effort provision. We also write $e_i \in \Delta\mathbb{R}_+$ for Dirac measures that represent pure strategies. Strategy profiles are given by $\sigma = (\sigma_1, ..., \sigma_n) \in (\Delta\mathbb{R}_+)^n$. We also use σ to denote the induced product measure in ΔE. We say a contract Φ implements a strategy profile σ if it is credible and satisfies

$$\Pi_i((\sigma_i, \sigma_{-i}), \Phi) \geq \Pi_i((\sigma_i', \sigma_{-i}), \Phi) \quad \forall \sigma_i' \in \Delta\mathbb{R}_+, \forall i \in I, \tag{IC-A}$$

where

$$\Pi_i(\sigma, \Phi) = \mathbb{E}_\sigma\left[\mathbb{E}_\tau\left[\mathbb{E}_{\mu^{e,\theta}}\left[u(t_i)\right]\right]\right] - \mathbb{E}_{\sigma_i}\left[c(e_i)\right].$$

Since a deviation to an effort of zero always guarantees each agent a payoff of at least zero,

[14] A different timing would be that the reviewer observes his type already before he interacts with the agents and observes their efforts. In that case, the principal could offer a menu of delegation sets from which the reviewer selects one before observing the agents' efforts. Our results are robust to this different timing provided that the reviewer's selection of a delegation set is not observable to the agents. If it was observable, additional signalling issues would arise, because the agents may coordinate on different effort profiles contingent on the reviewer's observable action. We refrain from studying these issues here, but we remark that, even in that case, the additional gain (if any) from a more complicated mechanism is limited if the agents' risk-aversion is moderate or the number of agents is large (see Section 4).

[15] This formulation does not preclude the possibility that a reviewer randomizes over actions, because the randomization over probability measures can instead be written as a compound measure that is chosen with probability one. We impose the following regularity condition on contracts: for each measurable set $A \subseteq T$, $\mu^{e,\theta}(A)$ is a measurable function of (e,θ). This ensures that expected payoffs are well-defined in contracts.

the agents' participation constraints can henceforth be ignored.[16]

The principal maximizes her expected payoff by choosing a contract Φ to implement some strategy profile σ. Formally, the principal's problem is given by

$$\max_{(\sigma,\Phi)} \Pi_P(\sigma,\Phi) \quad s.t. \quad \text{(IC-R)}, \text{(IC-A)}, \tag{P}$$

where

$$\Pi_P(\sigma,\Phi) = \mathbb{E}_\sigma\left[\sum_{i=1}^n e_i\right] - \mathbb{E}_\sigma\left[\mathbb{E}_\tau\left[\mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^n t_i\right]\right]\right].$$

A contract Φ^* is *optimal* if there exists σ^* such that (σ^*,Φ^*) solves (P).

Finally, we introduce a specific class of contracts that will be referred to as *contests*. A contest is described by a collection of prizes and a contest success function. Put differently, a contest commits to a profile of prizes $y = (y_1, ..., y_n) \in T$, some of which could be zero, and specifies how these prizes are allocated to the n agents as a function of their efforts. More formally, let $P(y)$ denote the set of permutations of y.[17] Then a contest C_y with prize profile y is a contract that satisfies (i) $\mu^{e,\theta}(P(y)) = 1$ for all $(e,\theta) \in E \times \Theta$, and (ii) $\mu^{e,\theta} = \mu^{e,\theta'}$ for all $\theta, \theta' \in \Theta$ and $e \in E$. This formulation includes all common contests, such as all-pay auctions and Tullock contests.

Note that every contest is credible. This is because, once the agents' efforts are sunk, any permutation of the prizes generates the same payoff for the principal and the same sum of utilities for the agents. Formally, in any contest C_y we obtain that

$$\Pi_R(e,\mu^{e',\theta'},\theta) = \sum_{i=1}^n e_i - \sum_{i=1}^n y_i + \theta\left(\sum_{i=1}^n u(y_i) - \sum_{i=1}^n c(e_i)\right)$$

is independent of (e',θ'). However, the set of credible contracts is substantially larger than the set of contests. For instance, it is possible to select from a much larger set of transfer profiles, not just permutations of a given prize profile, and still keep both the expected sum of transfers and the expected sum of the agents' utilities constant (see Section 4 for an example). Furthermore, credible contracts do not have to be independent of θ but can screen different reviewer types.

[16] Formally, constraints (IC-R) and (IC-A) characterize Perfect Bayesian equilibria of the following game. Given a delegation set D, the agents first simultaneously choose their efforts e. Nature then determines the reviewer's type θ. The reviewer finally observes e and θ and chooses from D. Constraint (IC-R) prescribes sequential rationality for the reviewer's (singleton) information sets, while (IC-A) prescribes sequential rationality (with weakly consistent beliefs) for the information sets in which the agents choose.

[17] Profile t is a permutation of y if there exists a bijective mapping $s : I \to I$ such that $t_i = y_{s(i)}$ $\forall i \in I$.

3 Optimal Contracts

3.1 The Optimality of Contests

To illustrate the key incentive problem in our model, suppose first that the preference parameter θ was known to the principal and the agents. The following example shows that, in this case, there may exist a credible contract which is not a contest but which implements the first-best effort levels and extracts the entire surplus.

Example. Consider our previous example for the special case of $n = 2$. Suppose the reviewer's type was common knowledge. First assume $\theta = 0$, so that there is also no misalignment of preferences between the principal and the reviewer. Consider a contract Φ^{FB} where, if both agents exert e^{FB}, each of them is paid t^{FB}. If one agent deviates, that agent is paid 0 while the non-deviating agent is paid $2t^{FB}$. In case both agents deviate, they are again each paid t^{FB}. It is easy to verify that this contract is credible, because the sum of transfers is constant across (t^{FB}, t^{FB}), $(2t^{FB}, 0)$ and $(0, 2t^{FB})$, which makes the reviewer indifferent between these transfer profiles. It is also easy to verify that this contract implements (e^{FB}, e^{FB}), because both agents receive a payoff of zero in equilibrium and a payoff of at most zero after any unilateral deviation. Thus the first-best is achievable. Observe that Φ^{FB} is not a contest, because the three transfer profiles are not all permutations of each other. We will show in Section 3.2 that the first-best is not achievable by a contest if the agents are risk-averse. Hence, Φ^{FB} performs strictly better than any contest in this example. This shows that non-verifiability of effort alone does not make contests optimal.[18]

Now assume that $\theta > 0$ and adjust the contract Φ^{FB} as follows. The payment $2t^{FB}$ to a non-deviating agent is replaced by some t^{nd}, while everything else is kept unchanged. If t^{nd} is chosen such that the reviewer is indifferent between the transfer profiles (t^{FB}, t^{FB}), $(t^{nd}, 0)$, and $(0, t^{nd})$, credibility is restored and the first-best can be implemented. For instance, with $\alpha = 1/2$, $\beta = 2$, and $\gamma = 1$ we have $e^{FB} \approx 0.63$ and $t^{FB} \approx 0.16$. For a reviewer of known type $\theta = 3$ we would then obtain $t^{nd} \approx 1.15$.[19] This shows that the misalignment of preferences per se does not make contests optimal either. $\square$

The contracts described in the example no longer work if θ is the reviewer's private information. Just consider the optimal contract for type $\theta = 0$. Any reviewer with type

[18] The argument is related to MacLeod (2003), who considers an environment with a principal, a single agent, and non-verifiable performance signals. He shows that, if the principal can commit to an unbalanced budget, she can credibly punish a shirking agent. In our example, the transfer to the non-deviating agent resolves the budget surplus problem.

[19] The indifference condition is $-2t^{FB} + \theta 2u(t^{FB}) = -t^{nd} + \theta u(t^{nd})$. Given our parameters, it has a second solution $t^{nd} \approx 3.72$, which would work as well. Note that the indifference condition is not guaranteed to have a solution for all parameter values, so our simple construction of a first-best contract does not work for all values of $\theta > 0$.

$\theta' > 0$ will strictly prefer to allocate (t^{FB}, t^{FB}), no matter what efforts the agents have exerted. This illustrates the leniency bias and the need for commitment.

Our first main result shows that, despite the fact that the set of possible contracts is very large, optimal contracts with uncertainty about θ take a very simple form.

Theorem 1 *The set of optimal contracts contains a contest.*

We will establish Theorem 1 by proving a series of six lemmas. Since we have shown that the principal may be able to implement the first-best if she knew the reviewer's private type θ, it would seem reasonable to expect that the principal could benefit from screening these types. In fact, it is not difficult to construct contracts that screen the reviewer's type, by varying the sum of transfers and the sum of utilities given to the agents. However, Lemmas 1 - 3 below demonstrate that it is not possible for the principal to benefit from screening. Lemmas 4 - 5 then show that the principal cannot benefit from implementing mixed or asymmetric effort profiles. Finally, Lemma 6 shows that using contests is without loss of generality.

To begin with, we fix an arbitrary contract $\Phi = (\mu^{e,\theta})_{(e,\theta)\in E\times\Theta}$ and denote

$$S_t(e,\theta) = \mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^{n} t_i\right],\ S_u(e,\theta) = \mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^{n} u(t_i)\right]$$

and

$$S(e,\theta) = -S_t(e,\theta) + \theta S_u(e,\theta).$$

We can then rewrite the credibility constraint (IC-R) as

$$-S_t(e,\theta) + \theta S_u(e,\theta) \geq -S_t(e',\theta') + \theta S_u(e',\theta') \ \ \forall(e,\theta),(e',\theta') \in E\times\Theta.$$

Our first lemma provides a characterization of this multidimensional constraint.

Lemma 1 *A contract Φ is credible if and only if the conditions (i) - (iii) hold:*

(i) $\forall\theta\in\Theta$, $S(e,\theta) = S(e',\theta)$ $\forall e,e'\in E$.

(ii) $\forall e\in E$, $S_u(e,\theta)$ *is non-decreasing in* θ.

(iii) $\forall e\in E$, $S(e,\theta) = S(e,\underline{\theta}) + \int_{\underline{\theta}}^{\theta} S_u(e,s)ds$ $\forall\theta\in\Theta$.

Conditions (ii) and (iii) are familiar from the mechanism design literature. They have to hold separately for each fixed effort profile e. Condition (i) concerns the effort dimension and shows that the payoff of any reviewer has to be constant for any reported e.

Given the characterization provided by Lemma 1, the next lemma states an important implication of the credibility constraint. Not only does $S(e, \theta)$ have to be constant across different profiles e, also its constituent parts $S_t(e, \theta)$ and $S_u(e, \theta)$ cannot vary with e.

Lemma 2 *A contract Φ is credible only if there exists a pair of functions $x : \Theta \to \mathbb{R}_+$ and $\hat{x} : \Theta \to \mathbb{R}_+$ such that, $\forall e \in E$,*

$$S_t(e, \theta) = x(\theta), \; S_u(e, \theta) = \hat{x}(\theta)$$

for almost all $\theta \in \Theta$.

We now show that there is no gain for the principal from screening the reviewer's private type by using a complex contract where $\mu^{e,\theta}$ varies with θ. Put differently, the principal can without loss of generality design the delegation set in a way such that all reviewers select the same actions.

Lemma 3 *For every contract Φ that implements a strategy profile σ, there exists a contract $\hat{\Phi}$ that also implements σ, yields the same expected payoff to the principal, and, $\forall \theta, \theta' \in \Theta$, satisfies $\hat{\mu}^{e,\theta} = \hat{\mu}^{e,\theta'}$ $\forall e \in E$.*

The proof of the lemma is constructive and shows how the contract $\hat{\Phi}$ without screening can be obtained from an arbitrary contract Φ. Given this result, we from now on focus without loss of generality on contracts where the agents' transfers depend on their efforts only, which we write as $\Phi = (\mu^e)_{e \in E}$. The next lemma shows that the principal does not benefit from implementing mixed strategies.

Lemma 4 *For every contract Φ that implements a strategy profile σ, there exists a contract $\hat{\Phi}$ that implements the pure-strategy profile $\bar{e} = (\bar{e}_1, ..., \bar{e}_n)$, where $\bar{e}_i = \mathbb{E}_{\sigma_i}[e_i]$ $\forall i \in I$, and yields the same expected payoff to the principal.*

The intuition behind this result is simple: any randomness in transfers that is achieved by mixed strategies can equivalently be generated by the contract. On the other hand, since c is convex, the agents benefit from exerting the average effort $\bar{e}_i$ instead of σ_i, while the principal is indifferent as to whether she obtains the efforts in expectation or deterministically.

The next lemma states that, in the current symmetric setting, it is without loss to restrict attention to the implementation of symmetric effort profiles.

Lemma 5 *For every contract Φ that implements a pure-strategy profile $\bar{e} = (\bar{e}_1, ..., \bar{e}_n)$, there exists a contract $\hat{\Phi}$ that implements the symmetric pure-strategy profile $\hat{e} = (\hat{e}_1, ..., \hat{e}_n)$, where $\hat{e}_1 = \ldots = \hat{e}_n = \frac{1}{n} \sum_{i=1}^{n} \bar{e}_i$, and yields the same expected payoff to the principal.*

The next lemma completes the proof of Theorem 1 by demonstrating that the principal can achieve the same payoff with a contest as with any non-screening contract that implements a symmetric pure-strategy effort profile. Thus, the principal can obtain her maximal payoff with a contest.[20]

Lemma 6 *For every contract Φ that implements a symmetric pure-strategy profile $\hat{e}$, there exists a contest C_y that also implements $\hat{e}$ and yields the same expected payoff to the principal.*

To prove this lemma, we construct a contest which implements the effort profile $\hat{e}$. This contest features $n-1$ identical prizes and one prize that is smaller. The small prize is used to punish agents who deviate in either direction from $\hat{e}$. In equilibrium, when the effort profile $\hat{e}$ is realized, the small prize is allocated randomly among the agents. As we will show in the next section, this particular prize structure is a general feature of optimal contests, while the specific (non-monotonic) allocation rule is not required to achieve the optimum.

3.2 Optimal Contests

From the previous section, we know that the principal can restrict attention to contests when designing an optimal contract. In this section, we characterize general features of all optimal contests. When describing a contest C_y, in the following we always assume w.l.o.g. that the prize profile y is ordered such that $y_1 \geq y_2 \geq \ldots \geq y_n$.

Theorem 2 *A contest is optimal if and only if the conditions (i) and (ii) hold:*

(i) The prizes satisfy $y_n = 0$ and $\sum_{k=1}^{n} y_k = x^*$, with $x^* = \min\{\bar{x}, \bar{T}\}$ and $\bar{x}$ given by

$$u'\left(\frac{\bar{x}}{n-1}\right) = c'\left(c^{-1}\left(\frac{n-1}{n}u\left(\frac{\bar{x}}{n-1}\right)\right)\right).$$

If the agents are risk-averse, then the prize profile is unique and given by

$$y = (x^*/(n-1), \ldots, x^*/(n-1), 0).$$

(ii) The contest implements $(e^*, \ldots, e^*)$, where e^* is given by

$$e^* = c^{-1}\left(\frac{n-1}{n}u\left(\frac{x^*}{n-1}\right)\right).$$

[20] To be exact, Theorem 1 follows only after it has been shown that problem (P) has a solution, so that an optimal contract exists. This will be shown in the proof of Theorem 2 in the next section.

Condition (i) in the theorem shows that the lowest prize will be zero in any optimal contest. This is not obvious, since the agents can be risk-averse and in equilibrium all agents face the risk of receiving the zero prize. The intuition is that in equilibrium an agent receives the zero prize with probability $1/n$, while a shirking agent would receive the zero prize with strictly larger probability (possibly one). Thus, any increase in y_n decreases the difference between the equilibrium and the deviation payoffs, and therefore decreases the amount of effort that can be demanded in equilibrium. When the agents are risk-averse, the optimal prize profile will feature $n-1$ identical positive prizes in addition to the zero prize, which keeps the risk imposed on the agents in equilibrium at a minimum. This prize structure is not novel but has been found to be optimal in different settings (e.g. Glazer and Hassin, 1988; Nalebuff and Stiglitz, 1983). We also characterize the optimal total prize sum x^*, which is given by the point $\bar{x}$ where marginal cost and benefit of inducing effort are equalized, or by the exogenous budget $\bar{T}$ whenever it is sufficiently tight.

Condition (ii) in the theorem shows that every optimal contest extracts the entire surplus from the agents, because it implements a symmetric pure-strategy effort profile such that each agent's equilibrium payoff is zero. This condition puts limits on the set of possible contest success functions which can be used to achieve the optimum. We will explore these limits in the following subsections.

Having characterized the optimal contests, we turn to the question of efficiency loss.

Corollary 1 *If the agents are risk-neutral, the principal can achieve the first-best. If the agents are risk-averse, the principal cannot achieve the first-best.*

The efficiency loss is driven entirely by risk-aversion of the agents. The loss is a direct consequence of the necessity to assign a prize of zero. Since the principal has to commit to the low prize, it must be delivered even in equilibrium. Risk-averse agents have to be compensated for this, which increases the cost of inducing effort. Hence, the commitment problem prevents the principal from achieving the first-best. However, the loss will be small if the agents are only mildly risk-averse, as the following example illustrates.

Example. Consider again our example for $n = 2$. Applying the results from Theorem 2, it can be shown that $e^* = 2^{\frac{\alpha-1}{\beta-\alpha}} e^{FB}$ and $x^* = 2^{\frac{\beta-1}{\beta-\alpha}} t^{FB}$ holds in any optimal contest. The ratio of second-best to first-best profits of the principal is therefore

$$R = \frac{2e^* - x^*}{2e^{FB} - 2t^{FB}} = \frac{2^{\frac{\beta-1}{\beta-\alpha}}(e^{FB} - t^{FB})}{2(e^{FB} - t^{FB})} = 2^{\frac{\alpha-1}{\beta-\alpha}}.$$

This ratio is increasing in α, with $R \to 1$ in the limit as $\alpha \to 1$, so second-best profits approach first-best profits when the agents' risk-aversion vanishes. □

The loss will also be small for any given risk-aversion if there are many agents, because the probability of not receiving any of the $n-1$ identical prizes is $1/n$ in equilibrium and vanishes as the number of agents grows. This has been observed before by Glazer and Hassin (1988), and similar arguments can be found in Green and Stokey (1983) and Nalebuff and Stiglitz (1983). Here, these arguments imply that delegation comes with little loss compared to an unconstrained mechanism design approach with monetary transfers if risk-aversion is small or if the number of agents is large.

3.3 Unique Implementation

We say a contract Φ *uniquely implements* some pure-strategy effort profile e if it (i) implements e and (ii) does not implement any other (possibly mixed) strategy profile $\sigma \neq e$. The next theorem states that the second-best effort profile from Theorem 2 can be uniquely implemented by an all-pay auction with a cap.

An all-pay auction is one of the canonical contest types (see Konrad, 2009, Ch. 2.1). It is perfectly discriminating, in the sense that the agent with the highest effort wins the highest prize with probability one, the agent with the second highest effort wins the second prize, and so on. Ties are broken randomly. A cap is a maximum admissible effort level (see Che and Gale, 1998). In our setting, a cap could be implemented in two different ways. If efforts are not physically capped, the reviewer can still be instructed to not differentiate effort levels at or above the cap, i.e., an agent who exerts effort exactly at the cap and an agent who exerts effort above the cap are treated the same and have the same chance of winning each of the prizes. A simpler implementation is to put an actual upper bound on the effort that each agent can provide, e.g. by enforcing maximal work hours, limiting the accumulation of overtime, or imposing page limits and deadlines on grant proposals. See Gavious, Moldovanu, and Sela (2002) for many other examples of actual caps that are imposed in different contests.

The following result shows that a cap at the optimal effort level e^* generates a unique equilibrium in the all-pay auction, which is in pure strategies.[21]

Theorem 3 *The effort profile $(e^*, ..., e^*)$ is uniquely implemented by an all-pay auction with prize profile $y = (x^*/(n-1), ..., x^*/(n-1), 0)$ and a cap at e^*.*

To see why all agents exerting e^* is an equilibrium, observe that upward deviations are impossible (or ineffective) while downward deviations guarantee the zero prize. The intuition for the result that no other pure-strategy equilibria exist is similar to that for all-pay auctions without caps. For every effort profile $e \neq (e^*, ..., e^*)$, either an upward deviation discretely increases the probability of winning, or a downward deviation decreases

[21] Note that the all-pay auction without a cap does not have an equilibrium in pure strategies. Therefore, the max-min optimal ranking mechanism in Frankel (2014) would not be optimal in our setting with endogenous efforts, because it is unable to implement the optimal effort profile.

costs without changing the probability of winning. The last step in the proof is to show that the cap destroys any potential mixed-strategy equilibrium.

An existing literature has investigated the effect of caps on equilibria and revenues in all-pay auctions (see Che and Gale, 1998, for an early contribution and Olszewski and Siegel, 2017, for a very general recent treatment). In particular, it is known that caps can generate pure-strategy equilibria. In our setting, e^* is the largest possible cap for which a pure-strategy equilibrium still exists. This is intuitive. The principal wants to make the contest as competitive as possible without destroying the pure-strategy equilibrium.

3.4 Implementation in Tullock Contests

The rent-seeking literature commonly studies contests that are imperfectly discriminating, which means that higher effort translates smoothly into a higher winning probability (see again Konrad, 2009). For the symmetric case with n agents but only one prize, such contests are often characterized by a contest success function

$$p_i(e) = \frac{f(e_i)}{\sum_{j\in I} f(e_j)}, \tag{1}$$

which determines the probability that agent i wins the prize as a function of the effort profile, where f is continuous, strictly increasing and satisfies $f(0) = 0$ (Skaperdas, 1996). If all agents exert zero effort, each of them wins with equal probability. With more than one prize, as in our optimal contests, the contest success function can be applied in a nested fashion (see Clark and Riis, 1996): the first prize is allocated according to (1) among all n agents, the second prize is allocated according to (1) restricted to those $n-1$ agents who have not received the first prize, and so on.

Tullock contests are a special case for $f(e_i) = e_i^r$, where $r \geq 0$ is a parameter measuring the randomness of the allocation rule. In particular, if $r = 0$ the winners are determined randomly irrespective of the exerted efforts. On the other hand, as $r \to \infty$ the contest approaches the perfectly discriminating all-pay auction in which an agent exerting more effort wins a higher prize for sure.

The contest theory literature has developed foundations for various functional forms of the contest success function. For a comprehensive survey see Jia et al. (2013). Our next result contributes to this literature by showing that a specific contest success function arises as part of an optimally designed delegation mechanism.

Theorem 4 *The effort profile $(e^*, ..., e^*)$ is implemented by a nested contest with prize profile $y = (x^*/(n-1), ..., x^*/(n-1), 0)$ and the contest success function (1) for*

$$f(e_i) = c(e_i)^{r^*(n)} \text{ with } r^*(n) = \frac{n-1}{H_n - 1},$$

where $H_n = \sum_{k=1}^{n} 1/k$ is the n-th harmonic number.

The optimal contest success function incorporates the agents' cost function and thus depends on the effort technology. It can be thought of as a generalization of the Tullock contest success function to settings with non-linear cost. Furthermore, it always reduces to the traditional Tullock shape $f(e_i) = e_i^r$ when the cost function is $c(e_i) = \gamma e_i^\beta$, as in our running example.

The optimal randomness parameter $r^*(n)$ exhibits some noteworthy features. If there are two agents and hence one prize, we obtain $r^*(2) = 2$. It is known that this is the largest value of the parameter for which the contest still has a pure-strategy equilibrium (see e.g. Ewerhart, 2017). Picking up our earlier discussion, the contest is therefore again as competitive as possible without destroying the pure-strategy equilibrium. We conjecture that a similar reasoning explains the value of $r^*(n)$ for $n > 2$, but little is known about the equilibrium structure of general nested Tullock contests. The parameter $r^*(n)$ is strictly increasing in the number of agents, which means that the optimal contest becomes more discriminating as n grows. It holds that $r^*(n) \to \infty$ in the limit as $n \to \infty$, so that the contest approaches the all-pay auction when the number of agents becomes large.

Theorem 4 shows that the optimum can be achieved by using an appropriately designed imperfectly discriminating contest. This contest is strictly monotonic, in the sense that higher effort always translates into strictly higher expected monetary payments.

4 Extensions

4.1 Imperfect Effort Observation

The assumption that the reviewer perfectly observes the individual effort of each agent can be seen as a strong one. In this subsection, we investigate two different observational constraints that may appear realistic. We will study a setting where only effort differences between the agents but not the levels are observable, and one where only noisy signals of the efforts are available. We restrict attention to the case of two agents throughout this subsection.

We first assume that the reviewer does not observe the effort profile (e_1, e_2) but only the difference $\Delta e = e_1 - e_2$. We then define a *contest with additive noise* as a contest in which the optimal prize x^* is given to agent 1 if and only if $\Delta e + \tilde{\epsilon} \geq 0$, where $\tilde{\epsilon}$ is a random variable (see, for instance, Lazear and Rosen, 1981). Intuitively, agent 1 receives the prize whenever the effort difference Δe is larger than a randomly determined number. Such a contest can be conducted if only Δe is observable, but of course also if the entire profile e can be observed. Our next result shows that a contest with additive noise implements the optimum for an appropriate choice of the distribution of $\tilde{\epsilon}$.

Proposition 1 *The effort profile (e^*, e^*) is implemented by a contest with additive noise for $\tilde{\epsilon} \sim \mathcal{U}[-c(e^*)/c'(e^*), c(e^*)/c'(e^*)]$.*

The randomness in the allocation rule ensures that unilateral deviations from (e^*, e^*) are not profitable, because the winning probability adapts appropriately. The distribution can neither be too noisy, which would create incentives to deviate to smaller effort levels, nor to concentrated, which would create incentives to deviate to larger effort levels. Che and Gale (2000) provide a general treatment of contests with additive uniform noise. They show that these contests will often not have a symmetric pure-strategy equilibrium. The uniform distribution described in our Proposition 1 is chosen precisely to avoid this problem, i.e., to guarantee that the optimal pure-strategy effort profile is an equilibrium.

We next assume that the reviewer observes only noisy signals of the individual agents' efforts. The usual interpretation is that agent i exerts effort e_i but the final observable output is a random variable $\tilde{e}_i$ that depends on e_i. In fact, the principal may care about output rather than effort, but since her payoffs are unaffected by noise with zero mean, here we focus on randomness in the reviewer's observation of efforts. Since effort is non-negative, we assume that noise is multiplicative, such that

$$\tilde{e}_i = e_i \tilde{\eta}_i$$

for a non-negative random variable $\tilde{\eta}_i$. To ensure closed-form tractability, we assume that the effort cost function is given by $c(e_i) = \gamma e_i^\beta$ for some $\beta > 1$ and $\gamma > 0$. We furthermore assume that the pair $(\tilde{\eta}_1, \tilde{\eta}_2)$ follows a bivariate log-normal distribution,

$$(\tilde{\eta}_1, \tilde{\eta}_2) \sim \ln \mathcal{N} \left[\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right].$$

It is possible – but not necessary for our analysis – to impose parameter constraints that guarantee symmetry and/or that the expected value of $\tilde{e}_1 + \tilde{e}_2$ equals $e_1 + e_2$.

We now define a *contest with multiplicative noise* as a contest in which the optimal prize x^* is given to agent 1 if and only if $\tilde{\eta}\tilde{e}_1/\tilde{e}_2 \geq 1$, where $\tilde{\eta}$ is a non-negative random variable (see, for instance, Jia et al., 2013). Intuitively, agent 1 receives the prize whenever the observed effort ratio $\tilde{e}_1/\tilde{e}_2$ is larger than a randomly determined number. As before, such a contest could also be conducted if there is actually no noise in the observation. Our next result shows that a contest with multiplicative noise can indeed implement the optimum when observation is not too noisy.

Proposition 2 *If* $\sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \leq 2/(\pi\beta^2)$*, then the effort profile* (e^*, e^*) *is implemented by a contest with multiplicative noise for* $\tilde{\eta} \sim \ln \mathcal{N}[\nu_\eta, \sigma_\eta^2]$ *with*

$$\nu_\eta = \nu_2 - \nu_1 \quad \text{and} \quad \sigma_\eta^2 = \frac{2}{\pi\beta^2} - (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}).$$

As argued before, an appropriate level of randomness in the allocation guarantees implementation of the optimum. Noisy observation of efforts already generates some baseline randomness and can thus help to achieve the principal's goal. The condition $\sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \leq 2/(\pi\beta^2)$ just requires that this noise is not too strong. For instance, if the random variables $\tilde{\eta}_i$ are i.i.d., the condition simplifies to $\sigma_i^2 \leq 1/(\pi\beta^2)$. Positive correlation effectively reduces the observational noise and slackens it further. If it is satisfied, then the randomness due to noisy effort observation can be raised to the appropriate level by an additional random component in the contract. With $\tilde{\eta}$ as specified in the proposition, the compound random variable $(\tilde{\eta}_2/\tilde{\eta}_1)/\tilde{\eta}$ follows a log-normal distribution with location parameter $\nu = 0$ and scale parameter $\sigma^2 = 2/(\pi\beta^2)$. As we show in the proof, (e^*, e^*) is an equilibrium for the resulting stochastic allocation process.[22]

4.2 Non-Separability and Favoritism

The assumption that the agents have additively separable utility functions serves as a natural starting point. It was used at several points in the analysis, but it matters mostly for our characterization of the credibility constraint (IC-R). Here, separability implies that the agents' sunk efforts do not influence the reviewer's preferences over the distribution of the prizes. Without separability, different prize allocations may lead to different sums of the agents' utilities, which implies that the reviewer may no longer be indifferent between all possible permutations.

In order to keep comparisons simple, we introduce non-separability directly into the reviewer's utility function, leaving the agents' utility functions unchanged. For instance, non-separability can be captured by modifying the reviewer's payoff function to

$$\pi_R(e, t, \theta) = \pi_P(e, t) + \theta \sum_{i=1}^{n} h(\pi_i(e, t)),$$

[22]The formulation of the proposition allows for $\sigma_\eta^2 = 0$, by which we mean that $\tilde{\eta}$ is degenerate and takes the value e^{ν_η} with probability one. We are not aware of an explicit treatment of the multiplicative log-normal noise model in the literature, but of course it can be transformed into a specific probit model with additive normal noise (Dixit, 1987).

for a strictly concave function $h : \mathbb{R} \to \mathbb{R}$. Whenever $\theta > 0$, which we will assume for the following discussion, this transformation implies a concern for equality of the entire utilities of the agents ("wide bracketing"), rather than just their utilities from monetary transfers ("narrow bracketing"). In particular, the reviewer will prefer giving larger prizes to agents who have exerted higher efforts.[23]

For a contest to remain credible, it now has to be perfectly discriminating. This rules out contest success functions such as the Tullock form investigated in Section 3.4. However, the all-pay auction discussed in Section 3.3 will continue to work.[24]

Favoritism of the reviewer is another potential problem. It can be captured by agent-specific functions h_i in the above expression. The reviewer will then prefer giving larger prizes to agents he likes better, putting incentives at risk. To address this problem, a blind reviewing process could make sure that the reviewer observes the chosen efforts but not the identity of the agent who chose each effort. Blind reviews are indeed used for performance evaluation (for example, see Goldin and Rouse, 2000). Garbling the reviewer's information in such a way comes at no loss with an optimal contest, but it helps to sustain credibility despite possible asymmetries in how the reviewer wants to treat the agents.

These arguments show that contests are robust or can be made robust with respect to non-separable preferences or favoritism. One may also ask if the principal can exploit non-separable preferences or favoritism of the reviewer by writing a contract which is not a contest. The answer to this question will depend on the exact nature of the principal's knowledge. With precise knowledge of preferences, it may indeed be the case that a standard contest is no longer optimal. We leave this extension to future research. However, it is worth pointing out again that a robust contest will often achieve an outcome very close to the first-best, leaving little additional gain from designing a more complex mechanism. This is the case if the number of agents is sufficiently large (Glazer and Hassin, 1988; Green and Stokey, 1983; Nalebuff and Stiglitz, 1983) or if the agents' risk-aversion is moderate. We illustrate this in the following example.

Example. Consider our running example and fix $\beta = 2$ and $\gamma = 1$. Figure 1 depicts the percentage of first-best payoffs that the principal can achieve with an optimal contest, as a function of the risk-aversion parameter α and for several values of n. Two observations

[23] See Corchon and Dahm (2011) for a characterization of contest success functions which arise if a contest organizer ex post allocates prizes to maximize some non-separable, non-expected utility function. Similarly, Corchon and Dahm (2010) investigate contest success functions which arise when a contest organizer of unknown type freely allocates prizes ex post.

[24] This argument rests on the interpretation of the cap as an actual bound on efforts. If efforts above e^* are possible, a reviewer with non-separable utility may not want to follow the instruction to not differentiate between effort levels at and above the cap. Even in that case, credibility of the contest can be restored using tools from *information design*. The principal could structure the observation process such that deviations above e^* become unobservable to the reviewer (e.g. by forbidding the manager to call the workers on the weekend) and conduct an otherwise standard all-pay auction. She could also try to add the right amount of noise to make a perfectly discriminating contest optimal (see Proposition 2).

are immediate. First, as $\alpha \to 1$ the share of the first-best payoffs that the principal can capture converges to one. Second, for any given level of risk-aversion, the principal obtains a larger share with a larger number of agents. The example also shows that, even for a modest number of agents, the principal obtains a substantial share of the first-best payoffs by running an optimal contest. For instance, already for $n = 6$ the principal captures more than 90% of the first-best payoffs for *any* $\alpha \in (0, 1)$. $\square$

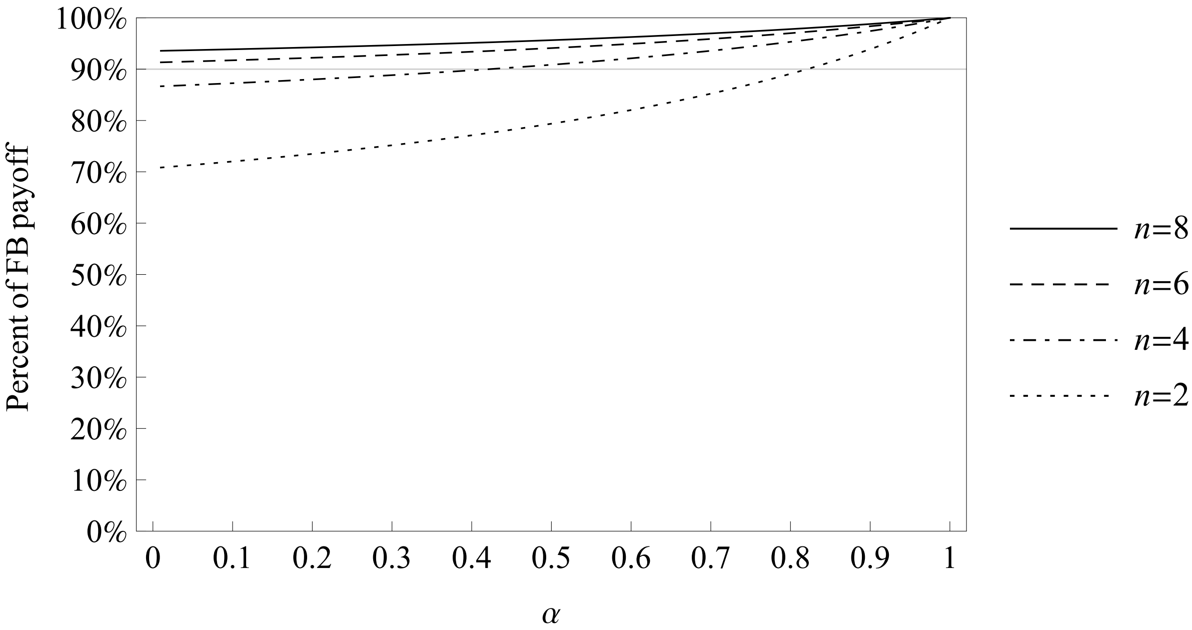

Figure 1: Share of first-best payoffs with an optimal contest.

4.3 Heterogeneous Abilities

We now consider a variation of the basic model in which the payoff of agent i is given by

$$\pi_i(e_i, t_i) = u(t_i) - c_i(e_i),$$

where the cost functions c_i satisfy our previous assumptions but can be different across agents. We first show that our main result is fully robust for the case of two agents.

Proposition 3 *Suppose $n = 2$. For any profile of effort cost functions (c_1, c_2), the set of optimal contracts contains a contest.*

To understand the logic of the result, first note that our earlier Lemmas 1 to 4 continue to hold under cost heterogeneity for any n. Hence it is still without loss of generality to consider contracts that do not screen the reviewer's type, implement a possibly asymmetric pure-strategy effort profile, and distribute a fixed expected sum of transfers x. For any such contract, we then show that a contest with prize vector $y = (x, 0)$ can implement the same effort profile when there are two agents. In contrast to our previous results for the symmetric case, the agents' winning probabilities will typically not be identical in equilibrium, because the agents may have to be compensated for different effort costs. The following example illustrates this point more explicitly.

Example. Consider a variant of our running example with $n = 2$, $\alpha = 1$, and heterogeneous effort costs $c_i(e_i) = \gamma_i e_i^\beta$ for $\gamma_1 < \gamma_2$. It can be shown that it is optimal for the principal to implement the (first-best) effort levels

$$e_1^* = \left(\frac{1}{\beta\gamma_1}\right)^{\frac{1}{\beta-1}} \quad \text{and} \quad e_2^* = \left(\frac{1}{\beta\gamma_2}\right)^{\frac{1}{\beta-1}}.$$

This can be achieved by a contest with prize vector $y = (x^*, 0)$ for $x^* = c_1(e_1^*) + c_2(e_2^*)$. In equilibrium, agent i receives the prize x^* with probability $p_i^* = c_i(e_i^*)/(c_1(e_1^*) + c_2(e_2^*))$. Hence the more efficient agent 1 is asked for higher equilibrium effort ($e_1^* > e_2^*$). In this example, the resulting difference in equilibrium efforts is so large that agent 1 bears a higher total cost ($c_1(e_1^*) > c_2(e_2^*)$). To compensate this cost difference, the prize x^* must be allocated to agent 1 with a probability larger than 1/2. $\square$

Generalizing the above argument to the case of $n > 2$ agents and a contest with $n - 1$ equal prizes faces the difficulty that some agents may have substantially higher effort costs in equilibrium than others and cannot be compensated even if they win one of the identical prizes for sure. Our next result rests on the insight that effort profiles for which the agents' costs are so strongly heterogeneous cannot be optimal if the agents' cost functions are not strongly heterogeneous. In that case, a contest with $n - 1$ equal prizes and one prize of zero is still optimal. To formalize this idea, we fix any sequence of cost function profiles $(c_1^m, \ldots, c_n^m)_{m\in\mathbb{N}}$ such that, for each $i \in I$, the sequence $(c_i^m)_{m\in\mathbb{N}}$ converges uniformly to some common cost function c as $m \to \infty$.

Proposition 4 *Suppose $(c_1^m, \ldots, c_n^m) \to (c, \ldots, c)$ uniformly. Then there exists $\underline{m} \in \mathbb{N}$ such that for all $m \geq \underline{m}$ the set of optimal contracts contains a contest.*

A "loser-gets-nothing" contest thus remains optimal whenever the differences in abilities are not too large. Again, the optimal contest will typically ask for different effort levels from different agents, and it allocates the zero prize with non-identical probabilities across the agents in equilibrium. For example, suppose that the cost heterogeneity is small enough for Proposition 4 to apply. Denote by $(e_1^*, ..., e_n^*)$ the optimal effort profile and let $\underline{e}^* = \min_{i\in I} e_i^*$. This optimum can be implemented by a modified all-pay auction with a cap. For any real effort profile $e = (e_1, \ldots, e_n)$, we first compute a "virtual effort profile" $\hat{e} = (\hat{e}_1, \ldots, \hat{e}_n)$ by $\hat{e}_i = (\underline{e}^*/e_i^*)e_i$ for every agent. Then the prizes are allocated as in an all-pay auction with cap $\underline{e}^*$ in which $\hat{e} = (\hat{e}_1, \ldots, \hat{e}_n)$ is taken as the actual efforts exerted by the agents. In equilibrium, when $e = (e_1^*, \ldots, e_n^*)$ and thus $\hat{e} = (\underline{e}^*, \ldots, \underline{e}^*)$, ties are broken such that agent i receives one of the positive prizes with probability $p_i^* = c_i(e_i^*)/u(x^*/(n-1))$, which is exactly enough to compensate for his effort cost. Thus the auction (*i*) handicaps agents who have to provide higher efforts by reducing their performance to lower virtual efforts, and (*ii*) compensates those agents who experience higher equilibrium effort costs by breaking ties with higher probability in their favor.

One may conjecture that a contract with more than two prize levels becomes optimal when the agents are very different from each other. This turns out to be true. Our next result shows that, for arbitrary heterogeneity in the agents' effort costs, a *generalized contest* is always optimal. Such a contest is described by two prize profiles y and y^d, where the prizes y are allocated in equilibrium and the prizes y^d are allocated after deviations. In essence, the reviewer gets to choose between two different contests and finds it optimal to select one of them in equilibrium, while the other one remains an unused (but credible) off-equilibrium threat. The special case of a standard contest arises when $y = y^d$.

Proposition 5 *For any profile of effort cost functions* $(c_1, \ldots, c_n)$*, the set of optimal contracts contains a generalized contest.*

As the proof of Proposition 5 reveals, the optimal equilibrium prize profile y no longer features $n-1$ equal prizes when costs are sufficiently heterogeneous. Instead, there are multiple different prize levels which reflect the heterogeneity in equilibrium effort costs.

While real-world contests often feature only two prize levels (for instance, the size of research grants is often fixed, students sometimes receive only pass-fail grades, tenure is either granted or declined, and workers are promoted or not), there are also contests with multiple prize levels. GE under Jack Welch separated their employees into three performance levels,[25] and grades are often given on a scale from A to D. According to our theory, multiple prize levels are offered as a response to heterogeneity in the agents' abilities. The prediction of our model is that multiple prize levels will be more likely in contests which are held among a population of agents where heterogeneity is high (for example, students in a public school), and that two prize levels will be more likely if the population of agents is more homogeneous (for example, junior analysts in a consulting firm).[26]

---

[25] See "'Rank and Yank' Retains Vocal Fans" (L. Kwoh, *The Wall Street Journal*, January 31, 2012).

[26] There are other reasons for using multiple prize levels. Moldovanu and Sela (2001) find, in a model with incomplete information and risk-neutral agents, that multiple prizes can be optimal for effort maximization when cost functions are convex. Olszewski and Siegel (2017b) develop a novel approach to contest design, which can be used for very general classes of large contests with many agents. They characterize the distribution of prizes which maximizes the aggregate effort exerted by the agents. Among other results, they find that multiple prizes of different levels are optimal when agents have convex costs or are risk-averse.

4.4 Cheap Talk

In the main model, we assumed that the principal delegates to the reviewer the decision on how to reward the agents. In this section, we consider a cheap talk model where the reviewer only reports the observed efforts to the principal, who then decides how to reward the agents. Similar to Kolotilin et al. (2013), we allow the principal to ex ante limit the set from which she can take her action ex post.

The timing is as follows. First, the principal commits to a set $D \subseteq \Delta T$ of possible actions. Next, the agents choose their efforts simultaneously. The reviewer then observes the efforts (and his type) and reports back to the principal. After receiving the report, the principal chooses an action from D to reward or punish the agents.

Our previous analysis can be modified to capture this cheap talk setting. Given a credible contract as defined before, we reinterpret $\mu^{e,\theta}$ as the report of a type-θ reviewer who has observed e, rather than as the action that the reviewer can take himself. This report can be interchangeably interpreted as a direct communication of (e, θ) or as a recommendation to pay the agents according to $\mu^{e,\theta}$. The following additional constraint then ensures that the principal always has an incentive to follow the reviewer's recommendation:

$$\pi_P(e, \mu^{e,\theta}) \geq \pi_P(e, \mu^{e',\theta'}) \ \ \forall (e, \theta), (e', \theta') \in E \times \Theta, \tag{IC-P}$$

where $\pi_P(e, \mu^{e',\theta'}) = \mathbb{E}_{\mu^{e',\theta'}} [\pi_P(e, t)]$. We are interested in problem (P) with the additional cheap talk constraint (IC-P). The solution to this problem describes the optimum that the principal can achieve in the cheap talk setting.[27]

Given this formulation of the problem, it is obvious that the cheap talk setting is weakly less permissive than the delegation setting. In the presence of a commitment problem, keeping the authority to make decisions may harm the principal. The following example illustrates possible consequences of the additional constraint (IC-P).

Example. Reconsider the first-best contract Φ^{FB} for $n = 2$ and a reviewer of known type $\theta = 3$ derived for our parametric example in Section 3. This contract rewards the first-best efforts by the transfer profile (t^{FB}, t^{FB}) and punishes unilateral deviations by one of the profiles $(t^{nd}, 0)$ or $(0, t^{nd})$. Since $2t^{FB} \approx 0.32 < t^{nd} \approx 1.15$, the sum of transfers is not constant across these profiles. Hence the contract violates (IC-P). With this contract, the principal would exhibit a leniency bias. To induce the altruistic reviewer to report

[27] The constraints (IC-R), (IC-A) and (IC-P) again characterize the Perfect Bayesian equilibria of an extensive form game. Consider the game described in footnote 16 and reinterpret the reviewer's choice from D as a recommendation to the principal. Then add a stage where the principal, after observing the recommendation but not the true state (e, θ), makes the choice from D. Despite the complexity of the principal's information sets, many of which are off the equilibrium path, constraint (IC-P) prescribes sequential rationality (given any weakly consistent beliefs) for all these information sets. The reason is that the principal's best responses in these information sets are independent of her beliefs about (e, θ).

a deviator truthfully, the punishment has to be combined with a very large payment t^{nd} to the non-deviating agent. Ex post, the principal would not be willing to carry out this costly punishment. $\square$

The next result shows that, with uncertainty about θ, the principal does not lose by keeping the authority to allocate rewards to the agents.

Proposition 6 *Any credible contract which does not screen reviewer types satisfies (IC-P).*

The result follows from the observation that, by Lemma 2, for each credible contract that does not screen the reviewer's type, there exists a value x such that $\mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^n t_i\right] = x$ for all $(e, \theta) \in E \times \Theta$. But then

$$\pi_P(e, \mu^{e',\theta'}) = \sum_{i=1}^n e_i - \mathbb{E}_{\mu^{e',\theta'}}\left[\sum_{i=1}^n t_i\right] = \sum_{i=1}^n e_i - x$$

is independent of (e', θ'), so the principal always has an incentive to follow the reviewer's recommendation. An immediate corollary of this result is that any contest, and moreover any generalized contest, satisfies the additional constraint (IC-P). Thus, all optimal contests derived for the delegation setting remain optimal in the cheap talk setting.

5 Related Literature

Our contribution is related to three distinct groups of papers. First, the optimal delegation mechanism in our paper is a contest, so we contribute to the literature examining conditions under which contests are optimal incentive mechanisms. Second, we work with a three-tiered hierarchical structure with a lenient reviewer. There are several papers featuring a similar structure. Third, the principal in our model delegates the decision on how to reward the agents to the reviewer. Our paper is therefore related to the literature on optimal delegation. We will discuss the connections and differences of our paper to each of these strands of literature in turn.

Optimality of contests. In their seminal paper, Lazear and Rosen (1981) show that contests can implement the socially optimal effort levels when agents are risk-neutral. They assume perfectly competitive labor markets in which the agents obtain all the surplus. Contests are then among the optimal mechanisms because they can induce first-best effort. At the same time, the set of optimal mechanisms also contains a piece-rate contract, among others. For the case of risk-averse agents, Lazear and Rosen (1981) compare the two specific mechanisms of piece-rate contracts and contests. They show that either of them sometimes dominates the other, but they do not establish results on global optimality.

A defining feature of contests is that the payoff of the agents depends on how well they perform relative to each other. This feature can make contests optimal in the presence of common uncertainty (see Holmstrom, 1982, and Mookherjee, 1984, for a general sufficient statistics approach to relative performance pay). Both Green and Stokey (1983) and Nalebuff and Stiglitz (1983) show that contests can do better than individual contracts when agents are risk-averse and there is a random common shock to their outputs. If the relationship between effort and output is ambiguous and the agents are ambiguity-averse, Kellner (2015) shows that contests can be optimal because they filter out the common ambiguity.

In our paper, contests provide a commitment for lenient reviewers to punish shirking agents. Contests are optimal because they incentivize the agents at least as efficiently as any other contract that can provide such a commitment.

Lenient reviewer. Several papers have looked at a three-tiered hierarchy with a lenient reviewer. Prendergast and Topel (1996) and Giebe and Gürtler (2012) consider a situation where the reviewer is facing a single agent and the principal can write contracts where the reviewer's pay is contingent on his behavior. In Prendergast and Topel (1996), both the reviewer and the principal receive a signal about the worker's effort. Their main result is that leniency need not be costly for the firm, because it can charge the reviewer for exercising leniency. In Giebe and Gürtler (2012), the principal offers a menu of contracts to screen lenient and non-lenient reviewers. They show that if the non-lenient type is common enough, the optimal solution can be to pay a flat wage to the reviewer, and rely on the non-lenient type for punishment of agents who shirk. The main difference between these papers and ours is that we consider multiple agents and do not allow contracts which condition payment to the reviewer on the reported evaluation.

Svensson (2003) applies a model with a lenient reviewer to the design of allocation mechanisms for foreign aid. The principal wants to use aid to incentivize countries to implement reforms. In his model, the principal determines the allocation mechanism, but the aid is then distributed by a country manager whose utility takes into account the well-being of the target countries. Svensson (2003) proposes a mechanism where each country manager is given a budget for several similar countries but has discretion in how to allocate the aid across countries. He shows that under certain conditions this mechanism can incentivize countries to reform. Like in our paper, Svensson (2003) considers multiple agents and does not allow for conditional monetary payments to the reviewer. The main difference is that we solve for the optimal delegation mechanism while Svensson (2003) compares two specific mechanisms.

Optimal delegation. In the usual delegation problem, the agent (the reviewer in our setting) is better informed about some exogenous state of the world. The principal delegates a unidimensional decision to the agent, but restricts the set of actions that the agent can choose. The question is how this set should be designed if the preferences of the principal and the agent are misaligned. The first to formulate the problem and show the existence of a solution was Holmström (1977, 1984), who focussed on interval delegation sets. Melumad and Shibano (1991) show that the optimal delegation set does not necessarily take the form of an interval. Alonso and Matouschek (2008) and Amador and Bagwell (2013) characterize the optimal delegation sets in progressively more general environments and find conditions under which the optimal delegation set is indeed an interval.

The canonical delegation model has been applied and extended in a number of ways.[28] In the multidimensional delegation model by Frankel (2014), which we have already discussed in the Introduction, optimal mechanisms exhibit what he calls the "aligned delegation" property, which means that all agents behave in the same way as the principal would behave. Our optimal mechanisms also satisfy the aligned delegation property, i.e., the equilibrium behavior of reviewers is independent of their type. Krähmer and Kováč (2016) assume that the agent has a privately known type which encodes his ability to interpret the private information he receives later on. They also find that screening is not beneficial in a large range of cases. Tanner (2014) obtains a no-screening result in a standard delegation model with uncertain bias of the agent.

Most papers cited above focus on deterministic delegation mechanisms. Kováč and Mylovanov (2009) and Goltsman, Hörner, Pavlov, and Squintani (2009) allow for stochastic delegation mechanisms and derive conditions under which the optimal mechanism is deterministic. We allow for stochastic mechanisms, and our optimal mechanism is indeed non-deterministic, but in a special sense – the contest prizes are allocated according to a probabilistic contest success function of the agents' efforts.

Finally, instead of delegating the decision to the agent, the principal could ask the agent to report the state of the world and take the action herself. This is the question addressed in the cheap talk literature in the tradition of Crawford and Sobel (1982). Several papers ask if the principal is better off delegating the decision or just asking for advice. Bester and Krähmer (2008) find that, if the agent needs to exert effort after selecting a project, delegation of the project selection is less likely to be optimal. Kolotilin et al. (2013) consider a model of cheap talk where the principal can ex ante commit not to take a certain action ex post. Again, they show that cheap talk with commitment can outperform delegation. On the other hand, Dessein (2002), Krishna and Morgan (2008), and Ivanov (2010) find that, in general, delegation is better than cheap talk. However, Fehr, Herz, and Wilkening (2013) and Bartling, Fehr, and Herz (2014) provide experimental evidence showing that

[28] See Armstrong and Vickers (2010) for an application to merger policy, Pei (2015a) for a model where delegation is used to conceal the principal's private type, Guo (2016) for a model of delegation of experimentation, and Frankel (2017) for a model of delegated hiring decisions.

individuals value decision rights intrinsically, which implies that delegation may not take place even when it is beneficial. These issues do not arise in our model. We can implement in a cheap talk setting exactly the same outcome as with the optimal delegation mechanism, provided the principal can commit to constraining her own actions in the same way as she can constrain the actions of the reviewer.

The main difference between our paper and the above discussed delegation literature is that the state of the world, on which the expert has private information, is exogenous in the delegation literature, while it is endogenous in our paper. In our model, the reviewer observes the efforts exerted by the agents. Since the incentives of the agents depend on the behavior of the reviewer, which in turn depends on the given delegation set, the state of the world is affected by the principal's choice of the delegation set.

6 Conclusion

In this paper we have analyzed a three-tiered structure consisting of a principal, a reviewer, and n agents. The principal designs a reward scheme in order to induce the agents to exert effort. However, the principal does not observe the efforts, so she delegates the allocation of rewards to the reviewer. The reviewer has private information about the utility weights he puts on the payoffs of the principal and of the agents.

Our main result is that a very simple mechanism, a contest, is optimal. By providing a commitment for the reviewer to punish shirking agents, a contest efficiently incentivizes the agents to exert costly effort. We also characterize the set of all optimal contests and show that they have a flat reward structure with $n-1$ equal positive prizes and one zero prize. Finally, we show that the optimum can be achieved with several common contest success functions, including all-pay auctions with caps, nested Tullock contests, and contests with additive or multiplicative noise. Our results are robust in a range of extensions, including imperfect observation of efforts and asymmetric abilities of the agents.

Other interesting questions could be asked in the framework of delegated performance evaluation. Here we will mention three immediate ones. First, the framework developed here can be used to study other forms of reviewer bias. A particularly important bias that could be examined is the gender bias (Goldin and Rouse, 2000). Moreover, it would be possible to examine how different biases like leniency and gender bias interact in shaping an optimal contract. Second, in addition to incentivizing agents, principals will often be interested in screening the abilities of heterogeneous agents in order to be able to assign more responsibilities to more capable agents. The purpose of a tenure or promotion contest is obviously not only to induce hard work, but also to select the right agents for a more advanced position. A question that could be asked in this framework is how screening and provision of incentives interact when both are delegated to potentially biased reviewers.

Finally, the principal also hires the reviewer. At first blush, our results might seem to suggest that the principal would be better off by trying to recruit a selfish reviewer who will not take the well-being of the agents into account. However, that would be the case only if the principal was able to determine the reviewer's type with absolute certainty. If there is a small remaining uncertainty, our results hold and assert that the principal's maximal payoff in an optimal contest does not depend on the reviewer's type. This implies that the principal is free to select the reviewer based on other criteria. For instance, an altruistic mid-level manager may outperform a selfish one in uniting his team to face a common challenge.

Of course, using a contest can also have drawbacks in some circumstances, for instance when the agents can easily sabotage each other's efforts. However, contests are widely used in reality, and our results offer a novel explanation for their optimality. We also point to new applications where contest-like mechanisms could be profitably implemented. Settings where an intermediary allocates monetary rewards are widespread in the economy, and, as the example of foreign aid illustrates, they can be found in unexpected places.

References

Aghion, P. and J. Tirole (1997): "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105, 1–29.

Alonso, R. and N. Matouschek (2008): "Optimal Delegation," *Review of Economic Studies*, 75, 259–293.

Amador, M. and K. Bagwell (2013): "The Theory of Optimal Delegation with an Application to Tariff Caps," *Econometrica*, 81, 1541–1599.

Armstrong, M. and J. Vickers (2010): "A Model of Delegated Project Choice," *Econometrica*, 78, 213–244.

Bartling, B., E. Fehr, and H. Herz (2014): "The Intrinsic Value of Decision Rights," *Econometrica*, 82, 2005–2039.

Barvinek, E., I. Daler, and J. Francu (1991): "Convergence of Sequences of Inverse Functions," *Archivum Mathematicum*, 27, 201–204.

Baye, M. R., D. Kovenock, and C. G. De Vries (1996): "The All-Pay Auction with Complete Information," *Economic Theory*, 8, 291–305.

Berger, J., C. Harbring, and D. Sliwka (2013): "Performance Appraisals and the Impact of Forced Distribution - An Experimental Investigation," *Management Science*, 59, 54–68.

Bernardin, H. J., D. K. Cooke, and P. Villanova (2000): "Conscientiousness and Agreeableness as Predictors of Rating Leniency," *Journal of Applied Psychology*, 85, 232–236.

Bester, H. and D. Krähmer (2008): "Delegation and Incentives," *RAND Journal of Economics*, 39, 664–682.

Bol, J. C. (2011): "The Determinants and Performance Effects of Managers' Performance Evaluation Biases," *Accounting Review*, 86, 1549–1575.

Breuer, K., P. Nieken, and D. Sliwka (2013): "Social Ties and Subjective Performance Evaluations: An Empirical Investigation," *Review of Managerial Science*, 7, 141–157.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2007): "The Pluralism of Fairness Ideals: An Experimental Approach," *American Economic Review*, 97, 818–827.

Carmichael, H. L. (1983): "The Agent-Agents Problem: Payment by Relative Output," *Journal of Labor Economics*, 1, 50–65.

Chang, S.-H., P. C. Cosman, and L. B. Milstein (2011): "Chernoff-Type Bounds for the Gaussian Error Function," *IEEE Transactions on Communications*, 59, 2939–2944.

Che, Y.-K. and I. L. Gale (1998): "Caps on Political Lobbying," *American Economic Review*, 88, 643–651.

——— (2000): "Difference-Form Contests and the Robustness of All-Pay Auctions," *Games and Economic Behavior*, 30, 22–43.

Clark, D. J. and C. Riis (1996): "A Multi-Winner Nested Rent-Seeking Contest," *Public Choice*, 87, 177–184.

Corchon, L. and M. Dahm (2010): "Foundations for Contest Success Functions," *Economic Theory*, 43, 81–98.

——— (2011): "Welfare Maximizing Contest Success Functions when the Planner Cannot Commit," *Journal of Mathematical Economics*, 47, 309–317.

Crawford, V. P. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

Dessein, W. (2002): "Authority and Communication in Organizations," *Review of Economic Studies*, 69, 811–838.

Dixit, A. (1987): "Strategic Behavior in Contests," *American Economic Review*, 77, 891–898.

Easterly, W. (2003): "Can Foreign Aid Buy Growth?" *Journal of Economic Perspectives*, 17, 23–48.

——— (2009): "Can the West Save Africa?" *Journal of Economic Literature*, 47, 373–447.

Ewerhart, C. (2017): "Revenue Ranking of Optimally Biased Contests: The Case of Two Players," *Economics Letters*, 157, 167–170.

Fehr, E., H. Herz, and T. Wilkening (2013): "The Lure of Authority: Motivation and Incentive Effects of Power," *American Economic Review*, 103, 1325–1359.

Frankel, A. (2014): "Aligned Delegation," *American Economic Review*, 104, 66–83.

——— (2017): "Selecting Applicants," *Mimeo.*

Gavious, A., B. Moldovanu, and A. Sela (2002): "Bid Costs and Endogenous Bid Caps," *RAND Journal of Economics*, 33, 709–722.

Giebe, T. and O. Gürtler (2012): "Optimal Contracts for Lenient Supervisors," *Journal of Economic Behavior & Organization*, 81, 403–420.

Glazer, A. and R. Hassin (1988): "Optimal Contests," *Economic Inquiry*, 26, 133–143.

Goldin, C. and C. Rouse (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 90, 715–741.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009): "Mediation, Arbitration and Negotiation," *Journal of Economic Theory*, 144, 1397–1420.

Green, J. R. and N. L. Stokey (1983): "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91, 349–364.

Guo, Y. (2016): "Dynamic Delegation of Experimentation," *American Economic Review*, 106, 1969–2008.

Holmström, B. (1977): "On Incentives and Control in Organizations," Ph.D. thesis, Stanford University.

——— (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340.

——— (1984): "On the Theory of Delegation," in *Bayesian Models in Economic Theory*, ed. by M. Boyer and R. Kihlstrom, Amsterdam: Elsevier, 115–141.

Ivanov, M. (2010): “Informational Control and Organizational Design,” *Journal of Economic Theory*, 145, 721–751.

Jia, H., S. Skaperdas, and S. Vaidya (2013): “Contest Functions: Theoretical Foundations and Issues in Estimation,” *International Journal of Industrial Organization*, 31, 211–222.

Kamenica, E. and M. Gentzkow (2011): “Bayesian Persuasion,” *American Economic Review*, 101, 2590–2615.

Kellner, C. (2015): “Tournaments as a Response to Ambiguity Aversion in Incentive Contracts,” *Journal of Economic Theory*, 159, 627–655.

Kolotilin, A., H. Li, and W. Li (2013): “Optimal Limited Authority for Principal,” *Journal of Economic Theory*, 148, 2344–2382.

Konrad, K. A. (2001): “Privacy and Time-Consistant Optimal Labor Income Taxation,” *Journal of Public Economics*, 79, 503–519.

——— (2009): *Strategy and Dynamics in Contests*, Oxford University Press.

Kováč, E. and T. Mylovanov (2009): “Stochastic Mechanisms in Settings without Monetary Transfers: The Regular Case,” *Journal of Economic Theory*, 144, 1373–1395.

Krähmer, D. and E. Kováč (2016): “Optimal Sequential Delegation,” *Journal of Economic Theory*, 163, 849–888.

Krishna, V. and J. Morgan (2008): “Contracting for Information under Imperfect Commitment,” *RAND Journal of Economics*, 39, 905–925.

Kwoh, L. (2012, January 31): “’Rank and Yank’ Retains Vocal Fans,” *The Wall Street Journal*, http://goo.gl/2UErjD.

Lazear, E. P. and S. Rosen (1981): “Rank-Order Tournaments as Optimum Labor Contracts,” *Journal of Political Economy*, 89, 841–864.

Levin, J. (2002): “Multilateral Contracting and the Employment Relationship,” *Quarterly Journal of Economics*, 117, 1075–1103.

MacLeod, B. W. (2003): “Optimal Contracting with Subjective Evaluation,” *American Economic Review*, 93, 216–240.

Malcomson, J. M. (1984): “Work Incentives, Hierarchy, and Internal Labor Markets,” *Journal of Political Economy*, 92, 486–507.

——— (1986): "Rank-Order Contracts for A Principal with Many Agents," *Review of Economic Studies*, 53, 807–817.

McGregor, J. (2013, November 20): "For Whom the Bell Curve Tolls," *The Washinton Post*, https://goo.gl/CwdpTO.

Melumad, N. D. and T. Shibano (1991): "Communication in Settings with No Transfers," *RAND Journal of Economics*, 22, 173–198.

Milgrom, P. and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

Moldovanu, B. and A. Sela (2001): "The Optimal Allocation of Prizes in Contests," *American Economic Review*, 91, 542–558.

Mookherjee, D. (1984): "Optimal Incentive Schemes with Many Agents," *Review of Economic Studies*, 51, 433–446.

Nalebuff, B. J. and J. E. Stiglitz (1983): "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *Bell Journal of Economics*, 14, 21–43.

Netzer, N. and F. Scheuer (2010): "Competitive Markets without Commitment," *Journal of Political Economy*, 118, 1079–1109.

Olszewski, W. and R. Siegel (2017a): "Bid Caps in Large Contests," *mimeo.*

——— (2017b): "Performance-Maximizing Contests with Many Contestants," *mimeo.*

Pei, H. D. (2015a): "A Blame Shifting Theory of Delegation," *mimeo.*

——— (2015b): "Communication with Endogenous Information Acquisition," *Journal of Economic Theory*, 160, 132–149.

Pinelis, I. (2002): "L'Hospital Type Rules for Monotonicity, with Applications," *Journal of Inequalities in Pure and Applied Mathematics*, 3.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37, 7–63.

Prendergast, C. and R. H. Topel (1996): "Favoritism in Organizations," *Journal of Political Economy*, 104, 958–978.

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

Rajan, R. G. and A. Subramanian (2008): "Aid and Growth: What Does the Cross-Country Evidence Really Show?" *Review of Economics and Statistics*, 90, 643–665.

Siegel, R. (2009): "All-Pay Contests," *Econometrica*, 77, 71–92.

Skaperdas, S. (1996): "Contest Success Functions," *Economic Theory*, 7, 283–290.

Strausz, R. (1997): "Delegation of Monitoring in a Principal-Agent Relationship," *Review of Economic Studies*, 64, 337–357.

Svensson, J. (2003): "Why Conditional Aid Does Not Work and What Can Be Done About It?" *Journal of Development Economics*, 70, 381–402.

Szalay, D. (2005): "The Economics of Clear Advice and Extreme Options," *Review of Economic Studies*, 72, 1173–1198.

Tanner, N. (2014): "Optimal Delegation Under Uncertain Bias," *mimeo.*

A Appendix

A.1 Proof of Theorem 1

A.1.1 Proof of Lemma 1

If-statement. We first show that (IC-R) is implied by (i) - (iii). Note that (IC-R) can be rewritten as

$$S(e,\theta) - S(e',\theta') \geq (\theta - \theta')S_u(e',\theta') \quad \forall (e,\theta),(e',\theta') \in E \times \Theta.$$

Using (i) and (iii), this is equivalent to the requirement that, $\forall e' \in E$ and $\forall \theta, \theta' \in \Theta$,

$$\int_{\theta'}^{\theta} \left(S_u(e',s) - S_u(e',\theta')\right) ds \geq 0,$$

and this inequality indeed holds since $S_u(e',\theta)$ is non-decreasing in θ by (ii).

Only-if-statement. We now proceed to prove that (IC-R) implies (i) - (iii). Note that for the special case $\theta' = \theta$, (IC-R) is reduced to the requirement that $S(e,\theta) \geq S(e',\theta)$ $\forall e, e' \in E$. Interchanging e and e', we immediately obtain (i). Next, consider the special case where $e' = e$. For this case, (IC-R) requires that, $\forall \theta, \theta' \in \Theta$,

$$S(e,\theta) \geq -S_t(e,\theta') + \theta S_u(e,\theta') \qquad (IC_{\theta,\theta'})$$

and

$$S(e,\theta') \geq -S_t(e,\theta) + \theta' S_u(e,\theta). \qquad (IC_{\theta',\theta})$$

Summing up $(IC_{\theta,\theta'})$ and $(IC_{\theta',\theta})$ we obtain

$$(\theta - \theta')\left(S_u(e,\theta) - S_u(e,\theta')\right) \geq 0 \quad \forall \theta, \theta' \in \Theta.$$

Thus, $S_u(e,\theta)$ must be non-decreasing in θ, which is condition (ii). The envelope formula in (iii) follows directly from Theorem 2 of Milgrom and Segal (2002), where absolute continuity of $S(e,\theta)$ holds because the set of transfer profiles is bounded. ■

A.1.2 Proof of Lemma 2

By Lemma 1, credibility implies that, $\forall e, e' \in E$ and $\forall \theta \in \Theta$,

$$\delta(e,e',\theta) = S(e,\theta) - S(e',\theta) = \int_{\underline{\theta}}^{\theta} \left(S_u(e,s) - S_u(e',s)\right) ds = 0.$$

This implies that, for any fixed $e, e' \in E$, $S_u(e,\theta) = S_u(e',\theta)$ for almost every $\theta \in \Theta$. It then also immediately follows that $S_t(e,\theta) = S_t(e',\theta)$ for almost every $\theta \in \Theta$. Now choose an arbitrary $e' \in E$ and define the functions x and $\hat{x}$ by

$$x(\theta) = S_t(e',\theta), \quad \hat{x}(\theta) = S_u(e',\theta) \quad \forall \theta \in \Theta.$$

It follows that, for any $e \in E$, $S_t(e,\theta) = x(\theta)$ and $S_u(e,\theta) = \hat{x}(\theta)$ for almost all $\theta \in \Theta$. ■

A.1.3 Proof of Lemma 3

Suppose $\Phi = (\mu^{e,\theta})_{(e,\theta)\in E\times\Theta}$ implements σ. In particular, Φ is credible, so by Lemma 2 there exists a pair of function x and $\hat{x}$ such that, $\forall e \in E$,

$$\mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^{n} t_i\right] = x(\theta), \quad \mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^{n} u(t_i)\right] = \hat{x}(\theta),$$

for almost every $\theta \in \Theta$. Since Φ implements σ, we also have $\forall i \in I$ and $\forall \sigma_i' \in \Delta\mathbb{R}_+$,

$$\mathbb{E}_{\sigma}\left[\mathbb{E}_{\tau}\left[\mathbb{E}_{\mu^{e,\theta}}[u(t_i)]\right] - c(e_i)\right] \geq \mathbb{E}_{(\sigma_i',\sigma_{-i})}\left[\mathbb{E}_{\tau}\left[\mathbb{E}_{\mu^{e,\theta}}[u(t_i)]\right] - c(e_i)\right].$$

The expected payoff of the principal with (σ,Φ) is given by

$$\Pi_P(\sigma,\Phi) = \mathbb{E}_{\sigma}\left[\sum_{i=1}^{n} e_i - \mathbb{E}_{\tau}\left[\mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^{n} t_i\right]\right]\right] = \mathbb{E}_{\sigma}\left[\sum_{i=1}^{n} e_i\right] - \mathbb{E}_{\tau}\left[x(\theta)\right].$$

For every $e \in E$, define a probability measure $\mu^e \in \Delta T$ such that

$$\mu^e(A) = \mathbb{E}_\tau \left[\mu^{e,\theta}(A)\right]$$

for all measurable subsets $A \subseteq T$.[29] Now construct an alternative contract $\hat{\Phi}$ by setting $\hat{\mu}^{e,\theta} = \mu^e$ for all $(e, \theta) \in E \times \Theta$. This contract satisfies the property of θ-independence stated in the lemma. Since, $\forall (e, \theta) \in E \times \Theta$,

$$\hat{S}_t(e,\theta) = \mathbb{E}_{\hat{\mu}^{e,\theta}}\left[\sum_{i=1}^n t_i\right] = \mathbb{E}_{\mu^e}\left[\sum_{i=1}^n t_i\right] = \mathbb{E}_\tau\left[\mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^n t_i\right]\right] = \mathbb{E}_\tau\left[x(\theta)\right],$$
$$\hat{S}_u(e,\theta) = \mathbb{E}_{\hat{\mu}^{e,\theta}}\left[\sum_{i=1}^n u(t_i)\right] = \mathbb{E}_{\mu^e}\left[\sum_{i=1}^n u(t_i)\right] = \mathbb{E}_\tau\left[\mathbb{E}_{\mu^{e,\theta}}\left[\sum_{i=1}^n u(t_i)\right]\right] = \mathbb{E}_\tau\left[\hat{x}(\theta)\right],$$

by Lemma 1 it is straightforward to check that $\hat{\Phi}$ is credible. Furthermore, note that

$$\begin{aligned}
\Pi_i(\sigma', \hat{\Phi}) &= \mathbb{E}_{\sigma'}\left[\mathbb{E}_\tau\left[\mathbb{E}_{\hat{\mu}^{e,\theta}}\left[u(t_i)\right]\right] - c(e_i)\right] \\
&= \mathbb{E}_{\sigma'}\left[\mathbb{E}_\tau\left[\mathbb{E}_{\mu^e}\left[u(t_i)\right]\right] - c(e_i)\right] \\
&= \mathbb{E}_{\sigma'}\left[\mathbb{E}_\tau\left[\mathbb{E}_{\mu^{e,\theta}}\left[u(t_i)\right]\right] - c(e_i)\right] \\
&= \Pi_i(\sigma', \Phi)
\end{aligned}$$

for all σ' and $i \in I$, which implies that $\hat{\Phi}$ implements σ because Φ implements σ.

Finally, from the above arguments we also obtain that the principal's expected payoff is $\mathbb{E}_\sigma\left[\sum_{i=1}^n e_i\right] - \mathbb{E}_\tau\left[x(\theta)\right]$ with both (σ, Φ) and $(\sigma, \hat{\Phi})$. ■

A.1.4 Proof of Lemma 4

Suppose $\Phi = (\mu^e)_{e \in E}$ implements σ. We first construct a probability measure $\eta \in \Delta T$ by

$$\eta(A) = \mathbb{E}_\sigma\left[\mu^e(A)\right]$$

for all measurable subsets $A \subseteq T$. Furthermore, for each $i \in I$ we construct a probability measure $\eta^{(i)} \in \Delta T$ by setting

$$\eta^{(i)}(A) = \mathbb{E}_\sigma\left[\mu^{(0, e_{-i})}(A)\right]$$

for all measurable subsets $A \subseteq T$. We now construct an alternative contract $\hat{\Phi} = (\hat{\mu}^e)_{e \in E}$ as follows. For $e = \bar{e}$, we let $\hat{\mu}^e = \eta$. For any $e = (e_i, \bar{e}_{-i})$ with $e_i \neq \bar{e}_i$, we let $\hat{\mu}^e = \eta^{(i)}$. For all remaining e, we let $\hat{\mu}^e = \mu^e$.

[29] The assumption discussed in footnote 15 ensures that the expectation (as well as the ones in the proof of the next lemma) is well-defined. It is also easy to show that μ^e is indeed a probability measure.

We first show that $\hat{\Phi}$ is credible. Since Φ is credible and its transfers are independent of θ, by Lemma 2 there exist $x, \hat{x} \in \mathbb{R}_+$ such that $\mathbb{E}_{\mu^e}\left[\sum_{i=1}^n t_i\right] = x$ and $\mathbb{E}_{\mu^e}\left[\sum_{i=1}^n u(t_i)\right] = \hat{x}$ for all $e \in E$. First consider $\hat{\mu}^e$ for $e = \bar{e}$. We obtain

$$\mathbb{E}_{\hat{\mu}^{\bar{e}}}\left[\sum_{i=1}^n t_i\right] = \mathbb{E}_\eta\left[\sum_{i=1}^n t_i\right] = \mathbb{E}_\sigma\left[\mathbb{E}_{\mu^e}\left[\sum_{i=1}^n t_i\right]\right] = \mathbb{E}_\sigma[x] = x$$

and, by the analogous argument, $\mathbb{E}_{\hat{\mu}^{\bar{e}}}[\sum_{i=1}^n u(t_i)] = \hat{x}$. Now consider $\hat{\mu}^e$ for $e = (e_i, \bar{e}_{-i})$ with $e_i \neq \bar{e}_i$. We obtain

$$\mathbb{E}_{\hat{\mu}^{(e_i,\bar{e}_{-i})}}\left[\sum_{i=1}^n t_i\right] = \mathbb{E}_{\eta^{(i)}}\left[\sum_{i=1}^n t_i\right] = \mathbb{E}_\sigma\left[\mathbb{E}_{\mu^{(0,e_{-i})}}\left[\sum_{i=1}^n t_i\right]\right] = \mathbb{E}_\sigma[x] = x$$

and, by the analogous argument, $\mathbb{E}_{\hat{\mu}^{(e_i,\bar{e}_{-i})}}[\sum_{i=1}^n u(t_i)] = \hat{x}$. Since $\hat{\Phi}$ and Φ are identical for all other e, we can conclude that $\mathbb{E}_{\hat{\mu}^e}\left[\sum_{i=1}^n t_i\right] = x$ and $\mathbb{E}_{\hat{\mu}^e}\left[\sum_{i=1}^n u(t_i)\right] = \hat{x}$ for all $e \in E$. It is then straightforward to check that $\hat{\Phi}$ is credible by using Lemma 1.

We next show that, in $\hat{\Phi}$, for each agent $i \in I$ it is a best response to play $\bar{e}_i$ when the remaining agents are playing $\bar{e}_{-i}$, which implies that $\hat{\Phi}$ implements $\bar{e}$. This claim holds because, $\forall i \in I$ and $\forall e_i' \neq \bar{e}_i$,

$$\begin{aligned}
\Pi_i(\bar{e}, \hat{\Phi}) &= \mathbb{E}_\eta[u(t_i)] - c(\bar{e}_i) \\
&= \mathbb{E}_\sigma[\mathbb{E}_{\mu^e}[u(t_i)]] - c\left(\mathbb{E}_\sigma[e_i]\right) \\
&\geq \mathbb{E}_\sigma[\mathbb{E}_{\mu^e}[u(t_i)]] - \mathbb{E}_\sigma[c(e_i)] \\
&\geq \mathbb{E}_\sigma\left[\mathbb{E}_{\mu^{(0,e_{-i})}}[u(t_i)]\right] \\
&\geq \mathbb{E}_\sigma\left[\mathbb{E}_{\mu^{(0,e_{-i})}}[u(t_i)]\right] - c(e_i') \\
&= \mathbb{E}_{\eta^{(i)}}[u(t_i)] - c(e_i') = \Pi_i((e_i', \bar{e}_{-i}), \hat{\Phi}),
\end{aligned} \tag{2}$$

where the first inequality follows the convexity of c and the second inequality follows from the fact that Φ implements σ.

Finally, from the above arguments we also obtain that the principal's expected payoff is $\sum_{i=1}^n \bar{e}_i - x$ with both (σ, Φ) and $(\bar{e}, \hat{\Phi})$. ■

A.1.5 Proof of Lemma 5

Suppose $\Phi = (\mu^e)_{e \in E}$ implements $\bar{e}$. We now construct an alternative contract $\hat{\Phi} = (\hat{\mu}^e)_{e \in E}$ as follows. For $e = \hat{e}$, we define $\hat{\mu}^e$ by generating a profile of prizes $t = (t_1, ..., t_n)$ according to $\mu^{\bar{e}}$ and then allocating these prizes randomly and uniformly among the agents. For any $e = (e_i, \hat{e}_{-i})$ with $e_i \neq \hat{e}_i$, we let $\hat{\mu}^e$ by given as follows. A number j is drawn uniformly from I and then a profile of prizes $t = (t_1, ..., t_n)$ is generated according to $\mu^{(0,\bar{e}_{-j})}$. The deviating agent i gets the prize t_j and the remaining $n-1$ prizes are allocated randomly and uniformly among the non-deviating agents. Note that, by construction, this punishment rule for unilateral deviations does not depend on the identity of the agent being punished. For all remaining e, we let $\hat{\mu}^e = \mu^e$.

We first show that $\hat{\Phi}$ is credible. By Lemma 2, credibility and θ-independence of Φ imply that there exists $x, \hat{x} \in \mathbb{R}_+$ such that $\mathbb{E}_{\mu^e}\left[\sum_{i=1}^n t_i\right] = x$ and $\mathbb{E}_{\mu^e}\left[\sum_{i=1}^n u(t_i)\right] = \hat{x}$ for all $e \in E$. Now first consider $\hat{\mu}^e$ for $e = \hat{e}$. We obtain

$$\mathbb{E}_{\hat{\mu}^{\hat{e}}}\left[\sum_{i=1}^n t_i\right] = \mathbb{E}_{\mu^{\bar{e}}}\left[\sum_{i=1}^n t_i\right] = x,$$
$$\mathbb{E}_{\hat{\mu}^{\hat{e}}}\left[\sum_{i=1}^n u(t_i)\right] = \mathbb{E}_{\mu^{\bar{e}}}\left[\sum_{i=1}^n u(t_i)\right] = \hat{x}.$$

Now consider $\hat{\mu}^e$ for any $e = (e_i, \hat{e}_{-i})$ with $e_i \neq \hat{e}_i$. We obtain

$$\mathbb{E}_{\hat{\mu}^{(e_i,\hat{e}_{-i})}}\left[\sum_{i=1}^n t_i\right] = \sum_{j=1}^n \frac{1}{n}\mathbb{E}_{\mu^{(0,\bar{e}_{-j})}}\left[\sum_{i=1}^n t_i\right] = \frac{1}{n}\sum_{j=1}^n x = x,$$
$$\mathbb{E}_{\hat{\mu}^{(e_i,\hat{e}_{-i})}}\left[\sum_{i=1}^n u(t_i)\right] = \sum_{j=1}^n \frac{1}{n}\mathbb{E}_{\mu^{(0,\bar{e}_{-j})}}\left[\sum_{i=1}^n u(t_i)\right] = \frac{1}{n}\sum_{j=1}^n \hat{x} = \hat{x}.$$

Since $\hat{\Phi}$ and Φ are identical for all other e, we can conclude that $\mathbb{E}_{\hat{\mu}^e}\left[\sum_{i=1}^n t_i\right] = x$ and $\mathbb{E}_{\hat{\mu}^e}\left[\sum_{i=1}^n u(t_i)\right] = \hat{x}$ for all $e \in E$. It is then straightforward to check that $\hat{\Phi}$ is credible by using Lemma 1.

We next show that, in $\hat{\Phi}$, for each agent $i \in I$ it is a best response to play $\hat{e}_i$ when the remaining agents are playing $\hat{e}_{-i}$, which implies that $\hat{\Phi}$ implements $\hat{e}$. To prove this claim, note that

$$\mathbb{E}_{\mu^{\bar{e}}}[u(t_i)] - c(\bar{e}_i) \geq \mathbb{E}_{\mu^{(0,\bar{e}_{-i})}}[u(t_i)]$$

holds for all $i \in I$ because Φ implements $\bar{e}$. Summing over all $i \in I$ and dividing by n yields

$$\mathbb{E}_{\mu^{\bar{e}}}\left[\frac{1}{n}\sum_{k=1}^n u(t_k)\right] - \frac{1}{n}\sum_{k=1}^n c(\bar{e}_k) \geq \frac{1}{n}\sum_{k=1}^n \mathbb{E}_{\mu^{(0,\bar{e}_{-k})}}\left[u(t_k)\right].$$

We now obtain, $\forall i \in I$ and $\forall e_i \neq \hat{e}_i$,

$$\Pi_i(\hat{e}, \hat{\Phi}) = \mathbb{E}_{\hat{\mu}^{\hat{e}}}\left[u(t_i)\right] - c\left(\hat{e}_i\right)$$

$$
\begin{aligned}
&= \mathbb{E}_{\mu^{\bar{e}}}\left[\frac{1}{n}\sum_{k=1}^{n} u(t_k)\right] - c(\hat{e}_i) \\
&\geq \mathbb{E}_{\mu^{\bar{e}}}\left[\frac{1}{n}\sum_{k=1}^{n} u(t_k)\right] - \frac{1}{n}\sum_{k=1}^{n} c(\bar{e}_k) \\
&\geq \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_{\mu^{(0,\bar{e}_{-k})}}\left[u(t_k)\right] - c(e_i) \\
&= \mathbb{E}_{\hat{\mu}^{(e_i,\hat{e}_{-i})}}[u(t_i)] - c(e_i) = \Pi_i((e_i,\hat{e}_{-i}),\hat{\Phi}),
\end{aligned}
$$

where the first inequality follows from convexity of c. Hence the claim follows.

Finally, from the above arguments we also obtain that the principal's expected payoff is $\sum_{i=1}^{n} \bar{e}_i - x$ with both $(\bar{e},\Phi)$ and $(\hat{e},\hat{\Phi})$. ■

A.1.6 Proof of Lemma 6

Suppose $\Phi = (\mu^e)_{e\in E}$ implements the symmetric profile $\hat{e}$. From the proof of Lemma 5 we know that it is without loss of generality to assume that Φ has the following form. If $e = \hat{e}$, a profile of prizes $t = (t_1, ..., t_n)$ is generated according to some probability measure π and these prizes are randomly and uniformly allocated to the agents. If $e = (e_i, \hat{e}_{-i})$ with $e_i \neq \hat{e}_i$ for some $i \in I$, a profile of prizes $t^d = (t^d_1, ..., t^d_n)$ is generated according to some (i-independent) probability measure ρ and agent i gets t^d_n, while the remaining $n-1$ prizes are randomly and uniformly allocated among the other agents. For all other effort profiles e, the transfer rule can be chosen as for $\hat{e}$. Thus, we have

$$
\mathbb{E}_{\mu^{\hat{e}}}[u(t_i)] = \mathbb{E}_{\pi}\left[\frac{1}{n}\sum_{k=1}^{n} u(t_k)\right], \quad \mathbb{E}_{\mu^{(e_i,\hat{e}_{-i})}}[u(t_i)] = \mathbb{E}_{\rho}[u(t^d_n)].
$$

Furthermore, by Lemma 2, credibility and θ-independence of Φ imply that there exist $x, \hat{x} \in \mathbb{R}_+$ such that $\mathbb{E}_{\mu^e}[\sum_{i=1}^{n} t_i] = x$ and $\mathbb{E}_{\mu^e}[\sum_{i=1}^{n} u(t_i)] = \hat{x}$ for all $e \in E$.

Now construct a contest C_y with prize profile y as follows. Define $\underline{t}^d$ as the certainty equivalent of a deviating agent's random transfers in contract Φ, i.e., $u(\underline{t}^d) = \mathbb{E}_{\rho}[u(t^d_n)]$. Note that $\underline{t}^d \leq \mathbb{E}_{\rho}[t^d_n]$ by concavity of u. Then define the prize profile

$$
y = \left(\frac{x-\underline{t}^d}{n-1}, \ldots, \frac{x-\underline{t}^d}{n-1}, \underline{t}^d\right).
$$

The allocation rule of C_y is as follows. If $e = \hat{e}$, the prizes are randomly and uniformly allocated among all agents. If $e = (e_i, \hat{e}_{-i})$ with $e_i \neq \hat{e}_i$ for some $i \in I$, the deviating agent i obtains $\underline{t}^d$ and all other agents obtain $(x - \underline{t}^d)/(n-1)$. For all other effort profiles e, the prizes are again randomly and uniformly allocated among all agents.

Since C_y is a contest, it is credible. Furthermore, $\forall i \in I$ and $\forall e_i \neq \hat{e}_i$,

$$
\begin{aligned}
\Pi_i(\hat{e}, C_y) &= \left(\frac{n-1}{n}\right) u\left(\frac{x-\underline{t}^d}{n-1}\right) + \frac{1}{n}u(\underline{t}^d) - c(\hat{e}_i) \\
&\geq \left(\frac{n-1}{n}\right) u\left(\frac{\mathbb{E}_\rho[\sum_{k=1}^n t_k^d)] - \mathbb{E}_\rho[t_n^d)]}{n-1}\right) + \frac{1}{n}\mathbb{E}_\rho[u(t_n^d)] - c(\hat{e}_i) \\
&= \left(\frac{n-1}{n}\right) u\left(\mathbb{E}_\rho\left[\sum_{k=1}^{n-1} \frac{1}{n-1} t_k^d\right]\right) + \frac{1}{n}\mathbb{E}_\rho[u(t_n^d)] - c(\hat{e}_i) \\
&\geq \left(\frac{n-1}{n}\right) \mathbb{E}_\rho\left[u\left(\sum_{k=1}^{n-1} \frac{1}{n-1} t_k^d\right)\right] + \frac{1}{n}\mathbb{E}_\rho[u(t_n^d)] - c(\hat{e}_i) \\
&\geq \left(\frac{n-1}{n}\right) \mathbb{E}_\rho\left[\sum_{k=1}^{n-1} \frac{1}{n-1} u\left(t_k^d\right)\right] + \frac{1}{n}\mathbb{E}_\rho[u(t_n^d)] - c(\hat{e}_i) \\
&= \mathbb{E}_\rho\left[\frac{1}{n}\sum_{k=1}^{n-1} u\left(t_k^d\right)\right] + \mathbb{E}_\rho\left[\frac{1}{n} u\left(t_n^d\right)\right] - c(\hat{e}_i) \\
&= \mathbb{E}_\rho\left[\frac{1}{n}\sum_{k=1}^{n} u\left(t_k^d\right)\right] - c(\hat{e}_i) \\
&= \mathbb{E}_\pi\left[\frac{1}{n}\sum_{k=1}^{n} u(t_k)\right] - c(\hat{e}_i) \\
&\geq \mathbb{E}_\rho\left[u(t_n^d)\right] - c(e_i) \\
&= u(\underline{t}^d) - c(e_i) = \Pi_i((e_i, \hat{e}_{-i}), C_y),
\end{aligned}
$$

where the first inequality follows from $\underline{t}^d \leq \mathbb{E}_\rho[t_n^d]$, the second and third inequalities follow from concavity of u, and the last inequality follows from the fact that Φ implements $\hat{e}$. We can thus conclude that C_y also implements $\hat{e}$.

Finally, from the above arguments we also obtain that the principal's expected payoff is $\sum_{i=1}^n \hat{e}_i - x$ with both $(\hat{e}, \Phi)$ and $(\hat{e}, C_y)$. ■

A.2 Proof of Theorem 2

Only-if-statement. Suppose (σ^*, C_y^*) solves (P). We first claim that σ^* must be a pure-strategy effort profile. By contradiction, suppose there exists $j \in I$ such that σ_j^* is not a Dirac measure. We can now proceed exactly as in the proof of Lemma 4 to construct a contract $\hat{\Phi}$ (in fact, a contest) that implements a pure-strategy profile $\bar{e}$. The only difference to the proof of Lemma 4 is that we let $\bar{e}_j = \mathbb{E}_{\sigma_j^*}[e_j] + \epsilon$ for some $\epsilon > 0$ (but still $\bar{e}_i = \mathbb{E}_{\sigma_i^*}[e_i]$ for all $i \neq j$). Credibility of $\hat{\Phi}$ and the fact that $\bar{e}_i$ is a best response to $\bar{e}_{-i}$ for all $i \neq j$ follow exactly as in the proof of Lemma 4. Since c is strictly convex and σ_j^* is not a Dirac measure, the first inequality in (2) is strict for j when $\epsilon = 0$. It then follows that $\bar{e}_j$ is a best response to $\bar{e}_{-j}$ for sufficiently small $\epsilon > 0$. Since the principal's payoff with $(\bar{e}, \hat{\Phi})$ is increased by ϵ, (σ^*, C_y^*) cannot have been a solution to (P).

Now suppose $(\bar{e}, C_y^*)$ solves (P), where $\bar{e}$ may still be asymmetric. Denote $x = \sum_{k=1}^n y_k$. We next show that whenever $y_n > 0$, there exists another contest $C_{\tilde{y}}$ with $\sum_{k=1}^n \tilde{y}_k = x$ that implements an effort profile $\tilde{e}$ with $\sum_{i=1}^n \tilde{e}_i > \sum_{i=1}^n \bar{e}_i$, and hence $(\bar{e}, C_y^*)$ cannot have been a solution to (P). Denote by $p_i^k(e)$ the probability that agent i receives prize y_k in C_y^* when the effort profile is e. Note that

$$\Pi_i(\bar{e}, C_y^*) = \sum_{k=1}^n p_i^k(\bar{e})u(y_k) - c(\bar{e}_i) \geq \sum_{k=1}^n p_i^k(0, \bar{e}_{-i})u(y_k) \geq u(y_n),$$

because C_y^* implements $\bar{e}$. Now consider an agent $j \in I$ for which $p_j^n(\bar{e}) < 1$. Construct a contest $C_{\tilde{y}}$ with a profile of prizes $\tilde{y}$ given by $\tilde{y}_1 = y_1 + \delta$, $\tilde{y}_n = y_n - \delta$, and $\tilde{y}_k = y_k$ for all $k \neq 1, n$, where $\delta \in (0, y_n]$. Note that $\sum_{k=1}^n \tilde{y}_k = x$. Let effort profile $\tilde{e}$ be such that $\tilde{e}_j = \bar{e}_j + \epsilon$ and $\tilde{e}_i = \bar{e}_i$ for all $i \neq j$, where $\epsilon > 0$. Note that $\sum_{i=1}^n \tilde{e}_i > \sum_{i=1}^n \bar{e}_i$. The rule of contest $C_{\tilde{y}}$ is the following. If the effort profile is $\tilde{e}$, then the prizes $\tilde{y}$ are allocated such that each agent i receives prize $\tilde{y}_k$ with probability $\tilde{p}_i^k(\tilde{e}) = p_i^k(\bar{e})$. If some agent i unilaterally deviates from $\tilde{e}$, then agent i receives the prize $\tilde{y}_n$, while the prizes $\tilde{y}_1, \ldots, \tilde{y}_{n-1}$ are allocated randomly and uniformly among the remaining agents. Otherwise, the allocation of prizes can be chosen arbitrarily. For sufficiently small $\epsilon > 0$ we then have, $\forall i \in I$ and $\forall e_i \in \mathbb{R}_+$,

$$\begin{aligned}
\Pi_i(\tilde{e}, C_{\tilde{y}}) &= \sum_{k=1}^n \tilde{p}_i^k(\tilde{e})u(\tilde{y}_k) - c(\tilde{e}_i) \\
&= \Pi_i(\bar{e}, C_y^*) + p_i^1(\bar{e})(u(y_1 + \delta) - u(y_1)) \\
&\quad + p_i^n(\bar{e})(u(y_n - \delta) - u(y_n)) + c(\bar{e}_i) - c(\tilde{e}_i) \\
&\geq u(y_n) + p_i^1(\bar{e})(u(y_1 + \delta) - u(y_1)) \\
&\quad + p_i^n(\bar{e})(u(y_n - \delta) - u(y_n)) + c(\bar{e}_i) - c(\tilde{e}_i) \\
&\geq u(y_n - \delta) = u(\tilde{y}_n) \geq \Pi_i((e_i, \tilde{e}_{-i}), C_{\tilde{y}}),
\end{aligned}$$

where the second inequality holds because

$$u(y_n) + p_i^n(\bar{e})(u(y_n - \delta) - u(y_n)) \geq u(y_n - \delta)$$

for all $i \in I$, with strict inequality for j. Hence $C_{\tilde{y}}$ implements $\tilde{e}$.

When studying the set of all contest solutions to (P), we thus need to consider only pure-strategy effort profiles $\bar{e}$ and contests C_y with $y_n = 0$. Fix a sum of prizes $x \in [0, \bar{T}]$.

Let e^x be the (unique) effort level that solves

$$\frac{n-1}{n}u\left(\frac{x}{n-1}\right) - c(e^x) = 0.$$

Note that, by the assumptions on u and c, the solution e^x is differentiable, strictly increasing and strictly concave in x. We now claim that ne^x is an upper bound on the sum of efforts implementable with a contest C_y that has $\sum_{k=1}^n y_k = x$ and $y_n = 0$, and it can be reached only by implementing the symmetric effort profile $(e^x, ..., e^x)$. Suppose first that C_y implements an effort profile $\bar{e}$ with $\sum_{i=1}^n \bar{e}_i \geq ne^x$ but $\bar{e} \neq (e^x, \ldots, e^x)$. Note that

$$\Pi_i(\bar{e}, C_y) = \sum_{k=1}^{n} p_i^k(\bar{e})u(y_k) - c(\bar{e}_i) \geq u(y_n) - c(0) = 0,$$

because C_y implements $\bar{e}$. Summing these inequalities over all agents we obtain

$$\sum_{i=1}^{n}\sum_{k=1}^{n} p_i^k(\bar{e})u(y_k) - \sum_{i=1}^{n} c(\bar{e}_i) = \sum_{k=1}^{n-1} u(y_k) - \sum_{i=1}^{n} c(\bar{e}_i) \geq 0.$$

However, due to weak concavity of u and strict convexity of c we also have

$$\sum_{k=1}^{n-1} u(y_k) - \sum_{i=1}^{n} c(\bar{e}_i) < (n-1)u\left(\frac{x}{n-1}\right) - nc(e^x) = 0,$$

a contradiction. Observe next that $(e^x, \ldots, e^x)$ can indeed be implemented. For instance, let $y = (x/(n-1), \ldots, x/(n-1), 0)$ and choose the rules of C_y as follows. If the effort profile is $(e^x, \ldots, e^x)$, then the prizes are allocated randomly and uniformly across the agents. If some agent i unilaterally deviates from $(e^x, \ldots, e^x)$, then agent i receives the prize 0, while each other agent receives $x/(n-1)$. Otherwise, the allocation of prizes can be chosen arbitrarily. It follows immediately from the definition of e^x that this contest indeed implements $(e^x, \ldots, e^x)$.

Given any sum of prizes x, the highest payoff that the principal can achieve is thus given by $\Pi_P(x) = ne^x - x$, and the problem is reduced to a choice of $x \in [0, \bar{T}]$. Since Π_P is continuous in x, it follows that a solution exists. Furthermore, since Π_P is differentiable and strictly concave, the first-order condition $\partial \Pi_P/\partial x = 0$ that is stated in part (i) of the theorem uniquely characterizes a value $\bar{x} > 0$ (given the assumptions on u and c), and the optimal value of x is given by $x^* = \min\{\bar{x}, \bar{T}\}$. The resulting implemented optimal effort level is then given by $e^* = e^{x^*}$.

We complete the proof of the only-if-statement by showing that any optimal contest has the profile of prizes $y = (x^*/(n-1), \ldots, x^*/(n-1), 0)$ whenever u is strictly concave. By contradiction, let C_y be a contest that implements $(e^*, ..., e^*)$ with $\sum_{k=1}^n y_k = x^*$ and $y_n = 0$

but $y_1 \neq y_{n-1}$. Proceeding as before, summing the inequalities $\Pi_i((e^*, \ldots, e^*), C_y) \geq 0$ over all agents yields $\sum_{k=1}^{n-1} u(y_k) - nc(e^*) \geq 0$. Strict concavity of u, however, implies that $\sum_{k=1}^{n-1} u(y_k) - nc(e^*) < (n-1)u(x^*/(n-1)) - nc(e^*) = 0$, a contradiction.

If-statement. We showed above that the upper bound on the principal's payoff is given by $ne^* - x^*$. Thus, any contest which implements $(e^*, ..., e^*)$ with the prize sum x^* attains the upper bound. ∎

A.3 Proof of Corollary 1

Each optimal contest induces individual efforts of e^* and pays a sum of x^*, as characterized in Theorem 2. Now consider the principal's first-best problem. If the agents' efforts were directly observable and verifiable, then the principal could ask for individual efforts of e and would have to compensate the agents with a transfer sum x such that $u(x/n) - c(e) = 0$. Put differently, for a given transfer sum x the maximal achievable individual effort is

$$e^x = c^{-1}\left(u\left(\frac{x}{n}\right)\right),$$

and the first-best problem is to maximize $ne^x - x$ by choice of $x \in [0, \bar{T}]$. With the same arguments as in the proof of Theorem 2, this yields $x^{FB} = \min\{\tilde{x}, \bar{T}\}$, where $\tilde{x}$ is given by

$$u'\left(\frac{\tilde{x}}{n}\right) = c'\left(c^{-1}\left(u\left(\frac{\tilde{x}}{n}\right)\right)\right).$$

The resulting optimal effort level is

$$e^{FB} = c^{-1}\left(u\left(\frac{x^{FB}}{n}\right)\right).$$

Now suppose that the agents are risk-neutral, i.e., the function u is linear. The conditions characterizing (e^*, x^*) in Theorem 2 then coincide with those characterizing (e^{FB}, x^{FB}) above, which implies $(e^*, x^*) = (e^{FB}, x^{FB})$. Then suppose that the agents are risk-averse, i.e., the function u is strictly concave. If $x^* \neq x^{FB}$ there is nothing to prove. Hence assume $x^* = x^{FB}$. Inspection of the conditions that define e^* and e^{FB} then immediately reveals that $e^* < e^{FB}$. ∎

B Appendix – Not For Print

B.1 Proof of Theorem 3

The proof proceeds in two steps. Step 1 shows that $(e^*, \ldots, e^*)$ is an equilibrium of the contest C_y^* described in the theorem. Step 2 shows that no other equilibria exist. The structure of the arguments in Step 2 is reminiscent of equilibrium characterization proofs in all-pay auctions with or without caps (e.g. Baye, Kovenock, and De Vries, 1996, or Che and Gale, 1998). Throughout the proof, we adopt the interpretation of a non-physical cap, so efforts $e_i > e^*$ are possible but are not differentiated by the reviewer. The result then also follows for the case where e^* is a physical bound on efforts.

Step 1. Consider deviations e_i' of agent i from $(e^*, \ldots, e^*)$. If $e_i' > e^*$, we obtain

$$\begin{aligned}
\Pi_i((e^*, \ldots, e^*), C_y^*) &= \frac{n-1}{n} u\left(\frac{x^*}{n-1}\right) - c(e^*) \\
&> \frac{n-1}{n} u\left(\frac{x^*}{n-1}\right) - c(e_i') \\
&= \Pi_i((e^*, \ldots, e_i', \ldots, e^*), C_y^*).
\end{aligned}$$

If $e_i' < e^*$, we obtain

$$\begin{aligned}
\Pi_i((e^*, \ldots, e^*), C_y^*) &= \frac{n-1}{n} u\left(\frac{x^*}{n-1}\right) - c(e^*) \\
&= 0 \geq -c(e_i') = \Pi_i((e^*, \ldots, e_i', \ldots, e^*), C_y^*).
\end{aligned}$$

Thus, the contest C_y^* implements the effort profile $(e^*, ..., e^*)$.

Step 2. By contradiction, suppose C_y^* also implements some other profile $\sigma \neq (e^*, \ldots, e^*)$. Denote the support of σ_i by L_i, so $e_i \in L_i$ if and only if every open neighbourhood N of e_i satisfies $\sigma_i(N) > 0$. We first show that it must be that $L_i \subseteq [0, e^*]$ for all $i \in I$. Suppose not, so there exists an agent i and an effort level $e_i > e^*$ such that $\sigma_i((e_i - \epsilon, e_i + \epsilon)) > 0 \ \ \forall \epsilon > 0$. Fix $\bar{\epsilon} > 0$ such that $e_i - \bar{\epsilon} > e^*$. Note that the expected payoff of agent i playing $e_i' \geq e^*$ with probability one, while the other agents play σ_{-i}, is

$$\Pi_i(e_i', \sigma_{-i}) = \left[1 - \prod_{j \neq i} \sigma_j([e^*, \infty)) + \prod_{j \neq i} \sigma_j([e^*, \infty)) \frac{n-1}{n}\right] u\left(\frac{x^*}{n-1}\right) - c(e_i').$$

where we omit the dependence on C_y^* to simplify notation. Since c is strictly increasing, we have $\Pi_i(e^*, \sigma_{-i}) > \Pi_i(e_i', \sigma_{-i})$ for all $e_i' > e^*$. Hence $\Pi_i(e^*, \sigma_{-i}) > \Pi_i(\bar{e}_i, \sigma_{-i})$ for all $\bar{e}_i \in (e_i - \bar{\epsilon}, e_i + \bar{\epsilon})$. Since $\sigma_i((e_i - \bar{\epsilon}, e_i + \bar{\epsilon})) > 0$, agent i could strictly increase his expected payoff by shifting the mass from this interval to e^*. Thus, σ is not an equilibrium. From now on, we only consider the cases where $L_i \subseteq [0, e^*] \ \forall i \in I$. Let $\underline{e}_i = \min L_i$. Since the proposed profile σ is different from $(e^*, ..., e^*)$, it must be that $\underline{e} = \min_{i\in I} \underline{e}_i < e^*$.

First, suppose that $\underline{e} > 0$. Furthermore suppose that $\sigma_j(\{\underline{e}\}) > 0$ for exactly one agent $j \in I$, or that $\sigma_i(\{\underline{e}\}) = 0$ for all $i \in I$. In the latter case let j be such that $\underline{e}_j = \underline{e}$. Then there exists some $\bar{\epsilon} > 0$ such that

$$\Pi_j(\underline{e}+\epsilon, \sigma_{-j}) \leq \left[1 - \prod_{i\neq j} \sigma_i((\underline{e}+\epsilon, \infty))\right] u\left(\frac{x^*}{n-1}\right) - c(\underline{e}+\epsilon) < 0$$

for all $\epsilon < \bar{\epsilon}$. Intuitively, the probability that agent j wins a positive prize approaches zero as ϵ approaches zero (by right continuity of $\sigma_i((\underline{e}+\epsilon,\infty))$ in ϵ and $\sigma_i((\underline{e},\infty)) = 1$), while the cost of effort at $\underline{e}$ is strictly positive. Hence agent j could strictly increase his expected payoff by shifting the mass $\sigma_j([\underline{e}, \underline{e}+\bar{\epsilon})) > 0$ from $[\underline{e}, \underline{e}+\bar{\epsilon})$ to 0. Next suppose that $\sigma_i(\{\underline{e}\}) > 0$ for at least two agents $i = j, k$. Then there exists a small $\epsilon > 0$ such that

$$\begin{aligned}
\Pi_j(\underline{e}, \sigma_{-j}) &\leq \left[1 - \left(1 - \frac{1}{2}\sigma_k(\{\underline{e}\})\right) \prod_{i\neq j,k} \sigma_i((\underline{e},\infty))\right] u\left(\frac{x^*}{n-1}\right) - c(\underline{e}) \\
&< \left[1 - \prod_{i\neq j} \sigma_i((\underline{e},\infty))\right] u\left(\frac{x^*}{n-1}\right) - c(\underline{e}+\epsilon) \\
&\leq \Pi_j(\underline{e}+\epsilon, \sigma_{-j}).
\end{aligned}$$

The intuition is that a small upward deviation from $\underline{e}$ increases the probability of winning discretely, while marginally increasing the effort costs. Hence agent j could strictly increase his expected payoff by shifting the mass $\sigma_j(\{\underline{e}\}) > 0$ from $\underline{e}$ to $\underline{e}+\epsilon$. We conclude that there does not exist an equilibrium $\sigma \neq (e^*, ..., e^*)$ with $\underline{e} > 0$.

Second, suppose that $\underline{e} = 0$. Consider first the case where $\sigma_i(\{0\}) = 0$ for all $i \in I$, that is, no agent places an atom on 0. If there is an agent j such that $\underline{e}_k > 0$ for all $k \neq j$, then there exists some $\bar{\epsilon} > 0$ such that $\Pi_j(\epsilon, \sigma_{-j}) = -c(\epsilon)$ for all $\epsilon < \bar{\epsilon}$. Agent j could then strictly increase his expected payoff by shifting the mass $\sigma_j((0, \bar{\epsilon})) > 0$ from $(0, \bar{\epsilon})$ to 0. Thus, there have to be at least two agents j and k with $\underline{e}_j = \underline{e}_k = 0$. But in this case, observe that

$$\Pi_j(\epsilon, \sigma_{-j}) \leq \left[\left(1 - \prod_{i\neq j} \sigma_i((\epsilon,\infty))\right) u\left(\frac{x^*}{n-1}\right) - c(\epsilon)\right]$$

and

$$\lim_{\epsilon\to 0} \left[\left(1 - \prod_{i\neq j} \sigma_i((\epsilon,\infty))\right) u\left(\frac{x^*}{n-1}\right) - c(\epsilon)\right] = 0.$$

Thus for every $\bar{\Pi} > 0$ there exists $\bar{\epsilon} > 0$ such that $\Pi_j(\epsilon, \sigma_{-j}) < \bar{\Pi}$ for all $\epsilon < \bar{\epsilon}$. Intuitively, both the probability of winning and the costs approach zero as $\epsilon \to 0$. However, it must be that $\Pi_j(e^*, \sigma_{-j}) > 0$ since $\Pi_j(e^*, \ldots, e^*) = 0$ and the probability that j wins a positive prize is strictly greater if the other agents play σ_{-j}, because at least agent k exerts efforts lower than e^* with strictly positive probability. Hence agent j could strictly increase his expected payoff by shifting the mass $\sigma_j((0, \bar{\epsilon})) > 0$ from $(0, \bar{\epsilon})$ to e^*, for some sufficiently small $\bar{\epsilon} > 0$. The only remaining case is $\sigma_j(\{0\}) > 0$ for at least one agent $j \in I$. Observe that there can only be one such agent, since otherwise a small upward deviation from 0 would lead to a discrete increase in the probability of winning a positive prize, analogous to the argument above. Then it must be that $\Pi_j(\sigma_j, \sigma_{-j}) = 0$ since $\Pi_j(0, \sigma_{-j}) = 0$. This can only be the maximum payoff of agent j if all other agents exert deterministic efforts equal to e^*, since otherwise $\Pi_j(e^*, \sigma_{-j}) > 0$. In this case, agent j is indifferent between playing 0 or e^*, and all other effort levels yield strictly lower payoffs. This implies $\sigma_j(\{0\}) + \sigma_j(\{e^*\}) = 1$. Now consider an agent $k \neq j$. Observe that a deviation by agent k to some ϵ with $0 < \epsilon < e^*$ leads to payoffs

$$\Pi_k(\epsilon, \sigma_{-k}) = \sigma_j(\{0\}) u\left(\frac{x^*}{n-1}\right) - c(\epsilon).$$

Thus a sufficiently small $\epsilon > 0$ will be a profitable deviation whenever

$$\sigma_j(\{0\}) u\left(\frac{x^*}{n-1}\right) > \sigma_j(\{0\}) u\left(\frac{x^*}{n-1}\right) + (1 - \sigma_j(\{0\}))\frac{n-1}{n} u\left(\frac{x^*}{n-1}\right) - c(e^*).$$

This can be reformulated to

$$0 > -\sigma_j(\{0\})\frac{n-1}{n} u\left(\frac{x^*}{n-1}\right),$$

which always holds because $\sigma_j(\{0\}) > 0$. ■

B.2 Proof of Theorem 4

Consider a nested contest with prize profile $y = (x^*/(n-1), ..., x^*/(n-1), 0)$ and the general success function (1). We will show that, for an appropriate choice of f, the effort profile $(e^*, \ldots, e^*)$ is an equilibrium. The proof proceeds in three steps. In Step 1, we derive the agents' payoff function in the nested contest. Step 2 introduces the specific value $r^*(n)$ stated in the theorem. In Step 3, we then complete the proof that the resulting contest indeed implements the desired effort profile.

Step 1. Let $p(e_i)$ denote the probability that agent i wins none of the $n-1$ positive prizes, given that all other agents exert effort e^*. Furthermore, let u^* be the utility derived from a positive prize. Then, the expected payoff of agent i, when all other agents exert e^*, is given by

$$
\begin{aligned}
\Pi_i(e_i) &= [1 - p(e_i)]\, u^* - c(e_i) \\
&= \left[1 - \frac{(n-1)! f(e^*)^{n-1}}{\prod_{k=1}^{n-1} [f(e_i) + (n-k) f(e^*))]}\right] u^* - c(e_i) \\
&= \left[1 - \prod_{k=1}^{n-1} \frac{(n-k) f(e^*)}{[f(e_i) + (n-k) f(e^*))]}\right] u^* - c(e_i) \\
&= \left[1 - \prod_{k=1}^{n-1} \frac{(n-k) f(e^*)}{[f(e_i) + (n-k) f(e^*))]}\right] \left(\frac{n}{n-1}\right) c(e^*) - c(e_i).
\end{aligned}
$$

Now suppose $f(e_i) = c(e_i)^r$ for some $r \geq 0$. It is easy to see that $\Pi_i(0) = \Pi_i(e^*) = 0$ for any r. We will show in the next two steps that $\Pi_i(e_i) \leq 0$ for all e_i when $r = r^*(n) = (n-1)/(H_n - 1)$, where $H_n = \sum_{k=1}^n 1/k$ is the n-th harmonic number. This implies that $(e^*, \ldots, e^*)$ is an equilibrium.

Step 2. Consider any $e_i > 0$ (we already know the value of Π_i for $e_i = 0$). To determine the sign of $\Pi_i(e_i)$, we can equivalently examine the sign of

$$
\Pi_i(e_i) \left[\frac{n-1}{nc(e^*)}\right] = \left[1 - \prod_{k=1}^{n-1} \frac{(n-k) c(e^*)^r}{[c(e_i)^r + (n-k) c(e^*)^r)]}\right] - \left(\frac{n-1}{n}\right) \frac{c(e_i)}{c(e^*)}.
$$

Make the change of variables $y^* = c(e^*)^r$ and $y = c(e_i)^r$ to obtain

$$
F(y|r) = \left[1 - \prod_{k=1}^{n-1} \frac{(n-k) y^*}{[y + (n-k) y^*]}\right] - \frac{n-1}{n} \left(\frac{y}{y^*}\right)^{\frac{1}{r}}.
$$

After the additional variable substitution $x = y^*/y$ we obtain

$$
F(x|r) = \left[1 - \prod_{k=1}^{n-1} \frac{(n-k) x}{[1 + (n-k) x]}\right] - \frac{n-1}{n} \left(\frac{1}{x}\right)^{\frac{1}{r}}.
$$

Showing that $F(x|r) \leq 0$ for all $x > 0$, $x \neq 1$, is then sufficient to ensure that the contest with parameter r implements the optimum.

Fix any x and let us look for $r(x)$ such that $F(x|r(x)) = 0$. Since F is strictly increasing in r whenever $x \in (0,1)$, we obtain that $F(x|r) \leq 0$ for any fixed $x \in (0,1)$ whenever $r \leq r(x)$, so $r(x)$ gives an upper bound on the possible values of r. Similarly, since F is strictly decreasing in r whenever $x \in (1,\infty)$, we obtain that $F(x|r) \leq 0$ for any fixed $x \in (1,\infty)$ whenever $r \geq r(x)$, so $r(x)$ gives a lower bound on the possible values of r. Thus it is sufficient to find a value r^* such that $r(x) \geq r^*$ for all $x \in (0,1)$ and $r(x) \leq r^*$

for all $x \in (1, \infty)$.

Rewriting the equation $F(x|r(x)) = 0$, we have

$$\begin{aligned}
\left[1 - \prod_{k=1}^{n-1} \frac{(n-k)x}{[1+(n-k)x]}\right] &= \frac{n-1}{n}\left(\frac{1}{x}\right)^{\frac{1}{r(x)}} \\
\log\left[1 - \prod_{k=1}^{n-1} \frac{(n-k)x}{[1+(n-k)x]}\right] &= \log\left(\frac{n-1}{n}\right) - \frac{1}{r(x)}\log(x) \\
\frac{1}{r(x)}\log(x) &= \log\left(\frac{n-1}{n}\right) - \log\left[1 - \frac{(n-1)!x^{n-1}}{\prod_{k=1}^{n-1}[1+(n-k)x]}\right] \\
\frac{1}{r(x)}\log(x) &= \log\left[\frac{n-1}{n}\frac{\prod_{k=1}^{n-1}[1+(n-k)x]}{\prod_{k=1}^{n-1}[1+(n-k)x] - (n-1)!x^{n-1}}\right] \\
r(x) &= \frac{\log(x)}{\log\left[\frac{n-1}{n}\frac{\prod_{k=1}^{n-1}[1+(n-k)x]}{\prod_{k=1}^{n-1}[1+(n-k)x] - (n-1)!x^{n-1}}\right]}.
\end{aligned}$$

Denote

$$g(x) = \frac{n-1}{n}\frac{\prod_{k=1}^{n-1}[1+(n-k)x]}{\prod_{k=1}^{n-1}[1+(n-k)x] - (n-1)!x^{n-1}}$$

so that

$$r(x) = \frac{\log(x)}{\log(g(x))}.$$

Note that $g(x) > 0$ for any $x > 0$. We will first show that $\lim_{x\nearrow 1} r(x) = \lim_{x\searrow 1} r(x) = r^*(n) = (n-1)/(H_n - 1)$. Note that for $x = 1$ both the denominator and the numerator of $r(x)$ equal zero. Hence we use l'Hôpital's rule. Observe that

$$\begin{aligned}
(\log(g(x)))' &= \frac{g'(x)}{g(x)} \\
&= \frac{\left(\frac{\partial}{\partial x}\prod_{k=1}^{n-1}[1+(n-k)x]\right)\left(\prod_{k=1}^{n-1}[1+(n-k)x] - (n-1)!x^{n-1}\right)}{\left(\prod_{k=1}^{n-1}[1+(n-k)x] - (n-1)!x^{n-1}\right)\prod_{k=1}^{n-1}[1+(n-k)x]} \\
&\quad - \frac{\left(\prod_{k=1}^{n-1}[1+(n-k)x]\right)\frac{\partial}{\partial x}\left(\prod_{k=1}^{n-1}[1+(n-k)x] - (n-1)!x^{n-1}\right)}{\left(\prod_{k=1}^{n-1}[1+(n-k)x] - (n-1)!x^{n-1}\right)\prod_{k=1}^{n-1}[1+(n-k)x]} \\
&= \frac{\left(\prod_{k=1}^{n-1}[1+(n-k)x]\right)\frac{\partial}{\partial x}\left((n-1)!x^{n-1}\right)}{\left(\prod_{k=1}^{n-1}[1+(n-k)x] - (n-1)!x^{n-1}\right)\prod_{k=1}^{n-1}[1+(n-k)x]}
\end{aligned}$$

$$
\begin{aligned}
&- \frac{\left((n-1)!x^{n-1}\right)\left(\dfrac{\partial}{\partial x}\prod_{k=1}^{n-1}\left[1+(n-k)x\right]\right)}{\left(\prod_{k=1}^{n-1}\left[1+(n-k)x\right]-(n-1)!x^{n-1}\right)\prod_{k=1}^{n-1}\left[1+(n-k)x\right]} \\
&= \frac{\left(\prod_{k=1}^{n-1}\left[1+(n-k)x\right]\right)(n-1)\left((n-1)!x^{n-2}\right)}{\left(\prod_{k=1}^{n-1}\left[1+(n-k)x\right]-(n-1)!x^{n-1}\right)\prod_{k=1}^{n-1}\left[1+(n-k)x\right]} \\
&- \frac{\left((n-1)!x^{n-1}\right)\left(\sum_{k=1}^{n-1}(n-k)\prod_{j\neq k}\left[1+(n-j)x\right]\right)}{\left(\prod_{k=1}^{n-1}\left[1+(n-k)x\right]-(n-1)!x^{n-1}\right)\prod_{k=1}^{n-1}\left[1+(n-k)x\right]}.
\end{aligned}
$$

We evaluate this at $x=1$, that is,

$$
\begin{aligned}
(\log(g(x)))'|_{x=1} &= \frac{\left(\prod_{k=1}^{n-1}\left[1+(n-k)\right]\right)(n-1)(n-1)!}{\left(\prod_{k=1}^{n-1}\left[1+(n-k)\right]-(n-1)!\right)\prod_{k=1}^{n-1}\left[1+(n-k)\right]} \\
&\quad - \frac{(n-1)!\left(\sum_{k=1}^{n-1}(n-k)\prod_{j\neq k}\left[1+(n-j)\right]\right)}{\left(\prod_{k=1}^{n-1}\left[1+(n-k)\right]-(n-1)!\right)\prod_{k=1}^{n-1}\left[1+(n-k)\right]} \\
&= \frac{n!(n-1)(n-1)!}{\left(n!-(n-1)!\right)n!} \\
&\quad - \frac{(n-1)!\left(\sum_{k=1}^{n-1}(n-k)\prod_{j\neq k}\left[1+(n-j)\right]\right)}{\left(n!-(n-1)!\right)n!} \\
&= 1 - \frac{\left(\sum_{k=1}^{n-1}(n-k)\prod_{j\neq k}\left[1+(n-j)\right]\right)}{(n-1)\,n!} \\
&= 1 - \frac{n!\left(\sum_{k=1}^{n-1}\dfrac{n-k}{n-k+1}\right)}{(n-1)\,n!} \\
&= \frac{n-1-\left(\sum_{k=1}^{n-1}\dfrac{n-k}{n-k+1}\right)}{n-1} \\
&= \frac{1+\sum_{k=1}^{n-1}\dfrac{n-k+1}{n-k+1}-\sum_{k=1}^{n-1}\dfrac{n-k}{n-k+1}-1}{n-1} \\
&= \frac{1+\sum_{k=1}^{n-1}\dfrac{1}{n-k+1}-1}{n-1} \\
&= \frac{H_n-1}{n-1}.
\end{aligned}
$$

Thus we have

$$
\lim_{x\nearrow 1} r(x) = \lim_{x\searrow 1} r(x) = \left.\frac{1/x}{(\log(g(x)))'}\right|_{x=1} = \frac{n-1}{H_n-1}.
$$

To complete the proof of the theorem, it is now sufficient to show that $r(x)$ is weakly monotonically decreasing on $(0,1)$ and on $(1,\infty)$. We will do this in the next step.

Step 3. To show monotonicity of $r(x)$, we will apply a suitable version of the l'Hôpital montone rule. Proposition 1.1 in Pinelis (2002) (together with Corollary 1.2 and Remark 1.3) implies that $r(x) = \log(x)/\log(g(x))$ is weakly decreasing on $(0,1)$ and $(1,\infty)$ if

$$\frac{(\log(x))'}{(\log(g(x)))'} = \frac{g(x)}{xg'(x)}$$

is weakly decreasing.[30] We will thus show that

$$\left(\frac{g(x)}{xg'(x)}\right)' = \frac{[g'(x)x - g(x)]g'(x) - xg(x)g''(x)}{(xg'(x))^2} \leq 0.$$

For this, it is sufficient to show the following three conditions:

(*a*) $g'(x) > 0$,

(*b*) $g''(x) \geq 0$,

(*c*) $g'(x)x - g(x) \leq 0$.

We will verify these conditions in the following three lemmas. To do this, consider the function g. We can write

$$\begin{aligned}\prod_{k=1}^{n-1}[1+(n-k)x] &= (n-1)!x^{n-1} + a_{n-2}x^{n-2} + a_{n-3}x^{n-3} + \cdots + a_1x + 1\\ &= (n-1)!x^{n-1} + \gamma(x),\end{aligned}$$

where $a_1, \ldots, a_{n-2}$ are strictly positive coefficients (that depend on n), so that γ is a polynomial of degree $n-2$ which is strictly positive for all $x > 0$.[31] We can then rewrite

$$g(x) = \frac{n-1}{n}\frac{(n-1)!x^{n-1} + \gamma(x)}{\gamma(x)}.$$

Lemma 7 *Condition $g'(x) > 0$ is satisfied.*

Proof. Observe that

$$\begin{aligned}g'(x) &= \frac{n-1}{n}\frac{(n-1)(n-1)!x^{n-2}\gamma(x) - (n-1)!x^{n-1}\gamma'(x)}{\gamma(x)^2}\\ &= \frac{n-1}{n}\frac{(n-1)!x^{n-2}[(n-1)\gamma(x) - x\gamma'(x)]}{\gamma(x)^2},\end{aligned}$$

[30] Proposition 1.1 is applicable because $\log(x)$ and $\log(g(x))$ are differentiable on the respective intervals and $\lim_{x\to 1}\log(x) = \lim_{x\to 1}\log(g(x)) = 0$ holds. The remaining prerequisite $(\log(g(x)))' = g'(x)/g(x) > 0$ also holds, because $g(x) > 0$ and $g'(x) > 0$ according to Lemma 7 below.

[31] To avoid confusion, the formula should be read as $\gamma(x) = 1$ for $n = 2$ and as $\gamma(x) = a_1x$ for $n = 3$.

and, since

$$
\begin{aligned}
(n-1)\gamma(x) &= (n-1)a_{n-2}x^{n-2} + (n-1)a_{n-3}x^{n-3} + \ldots + (n-1)a_1 x + n - 1 \text{ and} \\
x\gamma'(x) &= (n-2)a_{n-2}x^{n-2} + (n-3)a_{n-3}x^{n-3} + \ldots + a_1 x,
\end{aligned}
$$

it follows that $(n-1)\gamma(x) - x\gamma'(x) > 0$, which implies that $g'(x) > 0$. □

Lemma 8 *Condition $g''(x) \geq 0$ is satisfied.*

Proof. Observe that

$$
g''(x) = \frac{(n-1)(n-1)!}{n}\left[\frac{(n-1)x^{n-2}\gamma(x) - x^{n-1}\gamma'(x)}{\gamma(x)^2}\right]',
$$

so that $g''(x) \geq 0$ is equivalent to

$$
\begin{aligned}
0 &\leq \left[\frac{(n-1)x^{n-2}\gamma(x) - x^{n-1}\gamma'(x)}{\gamma(x)^2}\right]' \\
&= \frac{[(n-2)(n-1)x^{n-3}\gamma(x) + (n-1)x^{n-2}\gamma'(x) - (n-1)x^{n-2}\gamma'(x) - x^{n-1}\gamma''(x)]\gamma(x)^2}{\gamma(x)^4} \\
&\quad - \frac{[(n-1)x^{n-2}\gamma(x) - x^{n-1}\gamma'(x)]2\gamma(x)\gamma'(x)}{\gamma(x)^4} \\
&= \frac{[(n-2)(n-1)x^{n-3}\gamma(x) - x^{n-1}\gamma''(x)]\gamma(x)^2}{\gamma(x)^4} \\
&\quad - \frac{[(n-1)x^{n-2}\gamma(x) - x^{n-1}\gamma'(x)]2\gamma(x)\gamma'(x)}{\gamma(x)^4} \\
&= \frac{\gamma(x)x^{n-3}}{\gamma(x)^4}\left[(n-2)(n-1)\gamma(x)^2 - x^2\gamma''(x)\gamma(x) - 2(n-1)x\gamma(x)\gamma'(x) + 2x^2\gamma'(x)^2\right].
\end{aligned}
$$

The expression in the square bracket is a polynomial of degree $(2n-4)$. We will show that all coefficients of this polynomial are positive, which implies that the polynomial, and hence also $g''(x)$, is non-negative.

Using the auxiliary definitions $a_0 = 1$ and $a_\kappa = 0$ for $\kappa < 0$, the coefficient multiplying x^{2n-j} in this polynomial, for any $4 \leq j \leq 2n$, is given by

$$
\begin{aligned}
&\sum_{k=2}^{j-2}(n-2)(n-1)a_{n-k}a_{n-j+k} - \sum_{k=2}^{j-2}(n-k)(n-k-1)a_{n-k}a_{n-j+k} \\
&\qquad - \sum_{k=2}^{j-2}2(n-1)(n-k)a_{n-k}a_{n-j+k} + \sum_{k=2}^{j-2}2(n-k)(n-j+k)a_{n-k}a_{n-j+k} \\
&= \sum_{k=2}^{j-2}(n^2-3n+2)a_{n-k}a_{n-j+k} - \sum_{k=2}^{j-2}(n^2-2nk-n+k^2+k)a_{n-k}a_{n-j+k}
\end{aligned}
$$

$$-\sum_{k=2}^{j-2} 2(n^2 - nk - n + k)a_{n-k}a_{n-j+k} + \sum_{k=2}^{j-2} 2(n^2 - nj + jk - k^2)a_{n-k}a_{n-j+k}$$
$$= \sum_{k=2}^{j-2}(2 + 4nk - 3k^2 - 3k - 2nj + 2jk)a_{n-k}a_{n-j+k}.$$

Let $\varphi(n, j, k) = 2 + 4nk - 3k^2 - 3k - 2nj + 2jk$. We will show in several steps that $\sum_{k=2}^{j-2} \varphi(n, j, k)a_{n-k}a_{n-j+k} \geq 0$. For $n = 2$ and $n = 3$, this condition can easily be verified directly. Hence we suppose that $n > 3$ from now on.

Observe that for any k there is $k' = j - k$ such that $a_{n-k}a_{n-j+k} = a_{n-k'}a_{n-j+k'}$. Hence we first consider the case where j is odd, so that we can write

$$\sum_{k=2}^{j-2} \varphi(n, j, k)a_{n-k}a_{n-j+k} = \sum_{k=2}^{\frac{j-1}{2}}[\varphi(n, j, k) + \varphi(n, j, j - k)]a_{n-k}a_{n-j+k}.$$

Since $\varphi(n, j, k) + \varphi(n, j, j - k)$ is an integer, we can think of this expression as a long sum where each of the terms $a_{n-k}a_{n-j+k}$ appears exactly $|\varphi(n, j, k) + \varphi(n, j, j - k)|$ times, added or subtracted depending on the sign of $\varphi(n, j, k) + \varphi(n, j, j - k)$. Now note that $\sum_{k=2}^{(j-1)/2}[\varphi(n, j, k) + \varphi(n, j, j - k)] = 0$ holds. This follows because we can write

$$\begin{aligned}
&\sum_{k=2}^{\frac{j-1}{2}}[\varphi(n, j, k) + \varphi(n, j, j - k)] \\
&= \sum_{k=2}^{j-2} \varphi(n, j, k) \\
&= \sum_{k=2}^{j-2}(2 - 2nj) + (4n - 3 + 2j)\sum_{k=2}^{j-2} k - 3\sum_{k=2}^{j-2} k^2 \\
&= (j - 3)(2 - 2nj) + (4n - 3 + 2j)\frac{j(j - 3)}{2} - 3\frac{(j - 3)(2j^2 - 3j + 4)}{6} \\
&= (j - 3)\left(2 - 2nj + 2nj - \frac{3j}{2} + j^2 - j^2 + \frac{3j}{2} - 2\right) \\
&= 0.
\end{aligned}$$

Thus, for each instance where a term $a_{n-k'}a_{n-j+k'}$ is subtracted in the long sum, we can find a term $a_{n-k''}a_{n-j+k''}$ which is added. We claim that the respective terms which are added are weakly larger than the terms which are subtracted. This claim follows once we show that both $\varphi(n, j, k) + \varphi(n, j, j - k)$ and $a_{n-k}a_{n-j+k}$ are weakly increasing in k within the range of the sum. In that case, the terms which are subtracted are those for small k and the terms which are added are those for large k, and the latter are weakly larger. The same argument in fact applies when j is even, so that we can write

$$\sum_{k=2}^{j-2} \varphi(n,j,k) a_{n-k} a_{n-j+k}$$
$$= \sum_{k=2}^{\frac{j-2}{2}} [\varphi(n,j,k) + \varphi(n,j,j-k)] a_{n-k} a_{n-j+k} + \varphi(n,j,j/2) a_{n-j/2}^2.$$

Importantly, for the last term we have

$$\begin{aligned}
\varphi(n,j,j/2) &= 2 - 2nj - 3\left(\frac{j}{2}\right)^2 + \frac{j}{2}(4n - 3 + 2j) \\
&= 2 - j^2\frac{3}{4} - j\frac{3}{2} + j^2 \\
&= 2 + j\left(\frac{j}{4} - \frac{3}{2}\right) \\
&> 0,
\end{aligned}$$

so that the last and largest term $a_{n-j/2}^2 = a_{n-j/2} a_{n-j/2}$ is indeed also added.

We first show that $\varphi(n,j,k) + \varphi(n,j,j-k)$ is weakly increasing in k in the relevant range. We have

$$\begin{aligned}
&\varphi(n,j,k) + \varphi(n,j,j-k) \\
&= (2 - 2nj - 3k^2 + k(4n - 3 + 2j)) + (2 - 2nj - 3(j-k)^2 + (j-k)(4n - 3 + 2j)) \\
&= 4 - 4nj - 3(2k^2 + j^2 - 2jk) + j(4n - 3 + 2j).
\end{aligned}$$

Treating k as a real variable, we obtain

$$\begin{aligned}
\frac{\partial}{\partial k}\left[\varphi(n,j,k) + \varphi(n,j,j-k)\right] &= -3(4k - 2j) \\
&= -6(2k - j) > 0
\end{aligned}$$

for all $k < j/2$, so the claim follows.

We now show that $a_{n-k} a_{n-j+k}$ is weakly increasing in k in the relevant range. Formally, we show that $a_{n-k} a_{n-j+k} \leq a_{n-k-1} a_{n-j+k+1}$ for any $k < j/2$. Observe that we can write

$$\begin{aligned}
a_1 &= \sum_{k_1=1}^{n-1} (n - k_1), \\
a_2 &= \sum_{k_2=1}^{n-2} \sum_{k_1=k_2+1}^{n-1} (n - k_2)(n - k_1),
\end{aligned}$$

$$\vdots$$

$$a_j = \sum_{k_j=1}^{n-j} \sum_{k_{j-1}=k_j+1}^{n-j+1} \cdots \sum_{k_1=k_2+1}^{n-1} (n-k_j)(n-k_{j-1})\dots(n-k_1).$$

Intuitively, each summand in the definition of a_j is the product of j different elements chosen from the set $\{(n-1),(n-2),\dots,1\}$, and the nested summation goes over all the different possibilities in which these j elements can be chosen. Using simplified notation for the nested summation, we can thus write (where α, β, λ, and η take the role of the indices of summation, like k in the expression above):

$$\begin{aligned}
a_{n-k} &= \sum (n-\alpha_{n-k})(n-\alpha_{n-k-1})\dots(n-\alpha_1),\\
a_{n-j+k} &= \sum (n-\beta_{n-j+k})(n-\beta_{n-j+k-1})\dots(n-\beta_1),\\
a_{n-k-1} &= \sum (n-\lambda_{n-k-1})(n-\lambda_{n-k-2})\dots(n-\lambda_1),\\
a_{n-j+k+1} &= \sum (n-\eta_{n-j+k+1})(n-\eta_{n-j+k})\dots(n-\eta_1).
\end{aligned}$$

Rewriting the inequality $a_{n-k}a_{n-j+k} \leq a_{n-k-1}a_{n-j+k+1}$ using this notation, we obtain

$$\begin{aligned}
&\sum (n-\alpha_{n-k})(n-\alpha_{n-k-1})\dots(n-\alpha_1)(n-\beta_{n-j+k})(n-\beta_{n-j+k-1})\dots(n-\beta_1)\\
&\leq \sum (n-\lambda_{n-k-1})(n-\lambda_{n-k-2})\dots(n-\lambda_1)(n-\eta_{n-j+k+1})(n-\eta_{n-j+k})\dots(n-\eta_1).
\end{aligned}$$

Observe that each summand of the LHS sum is the product of $(n-k)+(n-j+k) = 2n-j$ elements, all of them chosen from the set $\{(n-1),(n-2),\dots,1\}$. The first $n-k$ elements are all different from each other, and the last $n-j+k$ elements are all different from each other. Thus, since $n-k > n-j+k$ when $k < j/2$, in each summand at most $n-j+k$ elements can appear twice. Furthermore, the LHS sum goes over all the different combinations that satisfy this property. Similarly, each summand of the RHS sum is the product of $(n-k-1)+(n-j+k+1) = 2n-j$ elements, all of them chosen from the same set $\{(n-1),(n-2),\dots,1\}$. The first $n-k-1$ elements are all different from each other, and the last $n-j+k+1$ elements are all different from each other. Thus, (weakly) more than $n-j+k$ elements can appear twice in these summands.[32] Since the RHS sum goes over all the different combinations that satisfy this property, for each summand on the LHS there exists an equal summand on the RHS. This shows that the inequality indeed holds. $\square$

Lemma 9 *Condition $g'(x)x - g(x) \leq 0$ is satisfied.*

[32] The inequality $n-k-1 \geq n-j+k+1$ can be rearranged to $k \leq j/2-1$, which follows from $k < j/2$, except if j is odd and $k = (j-1)/2$. Thus, typically, up to $n-j+k+1$ elements can appear twice. If j is odd and $k = (j-1)/2$, up to $n-k-1$ elements can appear twice, which is identical to $n-j+k$ in that case.

Proof. We have

$$g'(x)x - g(x) = \frac{n-1}{n}\left[\frac{(n-1)!x^{n-1}[(n-1)\gamma(x) - x\gamma'(x)]}{\gamma(x)^2} - \frac{(n-1)!x^{n-1} + \gamma(x)}{\gamma(x)}\right],$$

and therefore $g'(x)x - g(x) \leq 0$ if and only if

$$\begin{aligned}
0 &\geq (n-1)!x^{n-1}[(n-1)\gamma(x) - x\gamma'(x)] - (n-1)!x^{n-1}\gamma(x) - \gamma(x)^2 \\
&= (n-1)!x^{n-1}(n-2)\gamma(x) - (n-1)!x^n\gamma'(x) - \gamma(x)^2 \\
&= (n-1)![(n-2)a_{n-2}x^{2n-3} + (n-2)a_{n-3}x^{2n-4} + \cdots + (n-2)a_1x^n + (n-2)x^{n-1} \\
&\quad - (n-2)a_{n-2}x^{2n-3} - (n-3)a_{n-3}x^{2n-4} - \cdots - a_1x^n] - \gamma(x)^2 \\
&= (n-1)![a_{n-3}x^{2n-4} + 2a_{n-4}x^{2n-5} + \cdots + (n-3)a_1x^n + (n-2)x^{n-1}] - \gamma(x)^2 \\
&= (n-1)![a_{n-3}x^{2n-4} + 2a_{n-4}x^{2n-5} + \cdots + (n-3)a_1x^n + (n-2)x^{n-1}] \\
&\quad - \sum_{j=4}^{n+1}\sum_{k=2}^{j-2} a_{n-k}a_{n-j+k}x^{2n-j} - \rho,
\end{aligned}$$

where $\rho \geq 0$ is some positive remainder of $\gamma(x)^2$. To show $g'(x)x - g(x) \leq 0$, it is therefore sufficient to ignore ρ and show that the overall coefficient on x^{2n-j} in the last expression is not positive. That is, it is sufficient to show that, for all $j \in \{4, \ldots, n+1\}$,

$$(n-1)!(j-3)a_{n-j+1} - \sum_{k=2}^{j-2} a_{n-k}a_{n-j+k} \leq 0.$$

Observe that the sum has exactly $(j-3)$ elements. Then, it is sufficient to show that, for all $k \in \{2, \ldots, j-2\}$,

$$(n-1)!a_{n-j+1} \leq a_{n-k}a_{n-j+k}. \tag{3}$$

To demonstrate condition (3), we will first write the values of the coefficients a_j in a different way. Instead of summing over all possibilities in which j different elements from the set $\{(n-1), (n-2), \ldots, 1\}$ can be chosen, we can sum over the $n-j-1$ elements not chosen, and divide the factorial $(n-1)!$ by the product of these elements. This yields

$$\begin{aligned}
a_{n-2} &= \sum_{k_1=1}^{n-1} \frac{(n-1)!}{n-k_1}, \\
a_{n-3} &= \sum_{k_2=1}^{n-2}\sum_{k_1=k_2+1}^{n-1} \frac{(n-1)!}{(n-k_2)(n-k_1)}, \\
&\vdots
\end{aligned}$$

$$a_{n-j} = \sum_{k_{j-1}=1}^{n-j+1} \sum_{k_{j-2}=k_{j-1}+1}^{n-j+2} \cdots \sum_{k_1=k_2+1}^{n-1} \frac{(n-1)!}{(n-k_{j-1})(n-k_{j-2})\dots(n-k_1)},$$

$$\vdots$$

$$a_1 = \sum_{k_{n-2}=1}^{2} \sum_{k_{n-3}=k_{n-2}+1}^{3} \cdots \sum_{k_1=k_2+1}^{n-1} \frac{(n-1)!}{(n-k_{n-2})(n-k_{n-3})\dots(n-k_1)}.$$

Rewriting condition (3), we then have

$$\begin{aligned}
&\sum_{\lambda_{j-2}=1}^{n-j+2} \sum_{\lambda_{j-3}=\lambda_{j-2}+1}^{n-j+3} \cdots \sum_{\lambda_1=\lambda_2+1}^{n-1} \frac{((n-1)!)^2}{(n-\lambda_{j-2})(n-\lambda_{j-3})\dots(n-\lambda_1)} \\
&\leq \left[\sum_{\alpha_{k-1}=1}^{n-k+1} \sum_{\alpha_{k-2}=\alpha_{k-1}+1}^{n-k+2} \cdots \sum_{\alpha_1=\alpha_2+1}^{n-1} \frac{(n-1)!}{(n-\alpha_{k-1})(n-\alpha_{k-2})\dots(n-\alpha_1)}\right] \\
&\times \left[\sum_{\beta_{j-k-1}=1}^{n-j+k+1} \sum_{\beta_{j-k-2}=\beta_{j-k-1}+1}^{n-j+k+2} \cdots \sum_{\beta_1=\beta_2+1}^{n-1} \frac{(n-1)!}{(n-\beta_{j-k-1})(n-\beta_{j-k-2})\dots(n-\beta_1)}\right].
\end{aligned}$$

Observe that for each summand on the LHS, the denominator is a product of $j-2$ different elements from the set $\{(n-1), (n-2), \dots, 1\}$. In fact, the LHS sum goes over all the different possibilities in which these $j-2$ elements can be chosen. On the RHS, after multiplication, the denominator of each summand is a product of $(k-1)+(j-k-1) = j-2$ elements from the same set, where replication of some elements may be possible (but is not necessary). Since the RHS sum goes over all these different possibilities, for each summand on the LHS there exists an equal summand on the RHS. This shows that the inequality holds. □ ■

B.3 Proof of Proposition 1

Observe first that $c(e^*)/c'(e^*) < e^*$ holds due to strict convexity of c and $c(0) = 0$. We can therefore write the probability that agent 1 wins the prize in the described contest, holding the effort $e_2 = e^*$ fixed, as a piecewise function

$$p(e_1) = \begin{cases} 1 & \text{if } e_1 > e^* + \frac{c(e^*)}{c'(e^*)}, \\ \frac{1}{2} + \frac{1}{2}\frac{c'(e^*)}{c(e^*)}(e_1 - e^*) & \text{if } e^* - \frac{c(e^*)}{c'(e^*)} \leq e_1 \leq e^* + \frac{c(e^*)}{c'(e^*)}, \\ 0 & \text{if } e_1 < e^* - \frac{c(e^*)}{c'(e^*)}. \end{cases}$$

Then, the expected payoff of agent 1 is given by

$$\Pi_1(e_1) = p(e_1)u^* - c(e_1) = p(e_1)2c(e^*) - c(e_1).$$

It follows that $\Pi_1(e^*) = 0$. We now consider the three types of deviations from e^*.

Case 1: $e_1 < e^* - c(e^*)/c'(e^*)$. It follows immediately that $\Pi_1(e_1) \leq 0$ in this range, which implies that these deviations are not profitable.

Case 2: $e^* - c(e^*)/c'(e^*) \leq e_1 \leq e^* + c(e^*)/c'(e^*)$. Observe that $\Pi_1'(e_1) = c'(e^*) - c'(e_1)$ in this range. Hence the first-order condition yields the unique solution $e_1 = e^*$. Since $\Pi_1''(e_1) = -c''(e_1) < 0$, this is indeed the maximum over this range.

Case 3: $e_1 > e^* + c(e^*)/c'(e^*)$. We have $\Pi_1(e_1) < \Pi_1(e^* + c(e^*)/c'(e^*))$ for this range. Hence, by the arguments for the previous case, these deviations are not profitable either.

We conclude that $e_1 = e^*$ is a best response to $e_2 = e^*$. The argument for agent 2 is symmetric, which implies that the contest implements (e^*, e^*). ■

B.4 Proof of Proposition 2

Suppose that the condition $\sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \leq 2/(\pi\beta^2)$ is satisfied. Consider a contest as described in the proposition. We proceed in two steps. Step 1 derives an expression for agent i's expected payoff as a function of the effort profile e. Step 2 shows that $e_i = e^*$ is a best response when agent $j \neq i$ chooses $e_j = e^*$.

Step 1. Given an effort profile e, the probability that agent 1 wins the prize is

$$p(e) = \Pr\left[\frac{\tilde{\eta}\tilde{e}_1}{\tilde{e}_2} \geq 1\right] = \Pr\left[\frac{\tilde{\eta}\tilde{\eta}_1 e_1}{\tilde{\eta}_2 e_2} \geq 1\right] = \Pr\left[\frac{\tilde{\eta}_2}{\tilde{\eta}\tilde{\eta}_1} \leq \frac{e_1}{e_2}\right].$$

Since the variables $\tilde{\eta}_1$, $\tilde{\eta}_2$ and $\tilde{\eta}$ are log-normally distributed, it follows that the compound variable $\tilde{\eta}_2/(\tilde{\eta}\tilde{\eta}_1)$ is also log-normal, with location parameter $\nu = \nu_2 - \nu_1 - \nu_\eta = 0$ and scale parameter $\sigma^2 = \sigma_1^2 + \sigma_2^2 - \sigma_{12} + \sigma_\eta^2 = 2/(\pi\beta^2)$. The cdf of the log-normal distribution is given by $F(x) = \Phi\left((\log x - \nu)/\sigma\right)$, where Φ is the cdf of the standard normal distribution. Thus we can write

$$p(e) = \Phi\left(\log(e_1/e_2)\beta\sqrt{\frac{\pi}{2}}\right).$$

For the probability that agent 2 wins the prize we obtain

$$\begin{aligned}
1 - p(e) &= 1 - \Phi\left(\log(e_1/e_2)\beta\sqrt{\frac{\pi}{2}}\right) \\
&= \Phi\left(-\log(e_1/e_2)\beta\sqrt{\frac{\pi}{2}}\right) \\
&= \Phi\left(\log(e_2/e_1)\beta\sqrt{\frac{\pi}{2}}\right).
\end{aligned}$$

Hence the expected payoff of agent $i = 1, 2$ is

$$\Pi_i(e) = \Phi\left(\log(e_i/e_j)\beta\sqrt{\frac{\pi}{2}}\right) u^* - c(e_i)$$
$$= \Phi\left(\log(e_i/e_j)\beta\sqrt{\frac{\pi}{2}}\right) 2\gamma e^{*\beta} - \gamma e_i^\beta.$$

Step 2. Suppose $e_j = e^*$ and consider the choice of agent $i \neq j$. We immediately obtain $\Pi_i(e^*, e^*) = 0$. We will now show that $\Pi_i(e_i, e^*) \leq 0$ always holds, i.e.,

$$\Phi\left(\log(e_i/e^*)\beta\sqrt{\frac{\pi}{2}}\right) \leq \frac{1}{2}\left(\frac{e_i}{e^*}\right)^\beta$$

for all $e_i \in \mathbb{R}_+$. After the change of variables $x = \log(e_i/e^*)\beta\sqrt{\pi/2}$ this becomes the requirement that

$$\Phi(x) \leq \frac{1}{2}e^{x\sqrt{2/\pi}} \tag{4}$$

for all $x \in \mathbb{R}$. Inequality (4) is satisfied for $x = 0$, where LHS and RHS both take a value of $1/2$. Furthermore, the LHS function and the RHS function are tangent at $x = 0$, because their derivatives are both equal to $1/\sqrt{2\pi}$ at this point. It then follows immediately that inequality (4) is also satisfied for all $x > 0$, because the LHS is strictly concave in x in this range, while the RHS is strictly convex. We now consider the remaining case where $x < 0$. We use the fact that $\Phi(x) = \text{erfc}(-x/\sqrt{2})/2$, where

$$\text{erfc}(y) = \frac{2}{\sqrt{\pi}}\int_y^\infty e^{-t^2} dt$$

is the complementary error function (see e.g. Chang, Cosman, and Milstein, 2011). After the change of variables $y = -x/\sqrt{2}$ we thus need to verify

$$\text{erfc}(y) \leq e^{-2y/\sqrt{\pi}} \tag{5}$$

for all $y > 0$. Inequality (5) is satisfied for $y = 0$, where LHS and RHS both take a value of 1. Now observe that the derivative of the LHS with respect to y is given by $-2e^{-y^2}/\sqrt{\pi}$, while the derivative of the RHS is $-2e^{-2y/\sqrt{\pi}}/\sqrt{\pi}$. The condition that the former is weakly smaller than the latter can be rearranged to $y \leq 2/\sqrt{\pi}$, which implies that (5) is satisfied for $0 < y \leq 2/\sqrt{\pi}$. For larger values of y, we can use a Chernoff bound for the complementary error function. Theorem 1 in Chang et al. (2011) implies that

$$\text{erfc}(y) \leq e^{-y^2}$$

for all $y \geq 0$. The inequality $e^{-y^2} \leq e^{-2y/\sqrt{\pi}}$ can be rearranged to $y \geq 2/\sqrt{\pi}$. This implies that (5) is satisfied also for $y > 2/\sqrt{\pi}$, which completes the proof. ■

B.5 Proof of Proposition 3

We first derive some results which hold under cost heterogeneity for any number of agents. It is easy to see that Lemmas 1 to 4 continue to hold. Hence it is still without loss of generality to consider a contract $\Phi = (\mu^e)_{e \in E}$ that does not screen θ, satisfies

$$\mathbb{E}_{\mu^e}\left[\sum_{i=1}^{n} t_i\right] = x \text{ and } \mathbb{E}_{\mu^e}\left[\sum_{i=1}^{n} u(t_i)\right] = \hat{x}$$

for all $e \in E$, and implements a possibly asymmetric pure effort profile $\bar{e} = (\bar{e}_1, \ldots, \bar{e}_n)$. For any such contract, we obtain the following intermediate result.

Lemma 10 *If a contract Φ implements a pure-strategy effort profile $\bar{e} = (\bar{e}_1, \ldots, \bar{e}_n)$, then*

$$\frac{1}{n-1}\sum_{i=1}^{n} c_i(\bar{e}_i) \leq u\left(\frac{x}{n-1}\right).$$

Proof. Since Φ implements $\bar{e}$, for each $i \in I$ it must hold that

$$c_i(\bar{e}_i) \leq \mathbb{E}_{\mu^{\bar{e}}}\left[u(t_i)\right] - \mathbb{E}_{\mu^{(0,\bar{e}_{-i})}}\left[u(t_i)\right].$$

Summing over all $i \in I$, we obtain

$$\begin{aligned}
\sum_{i=1}^{n} c_i(\bar{e}_i) &\leq \sum_{i=1}^{n} \mathbb{E}_{\mu^{\bar{e}}}\left[u(t_i)\right] - \sum_{i=1}^{n} \mathbb{E}_{\mu^{(0,\bar{e}_{-i})}}\left[u(t_i)\right] \\
&= \hat{x} - \left(\hat{x} - \sum_{i=2}^{n} \mathbb{E}_{\mu^{(0,\bar{e}_{-1})}}[u(t_i)]\right) - \sum_{i=2}^{n} \mathbb{E}_{\mu^{(0,\bar{e}_{-i})}}[u(t_i)] \\
&\leq \sum_{i=2}^{n} \mathbb{E}_{\mu^{(0,\bar{e}_{-1})}}[u(t_i)] \\
&\leq \sum_{i=2}^{n} u\left(\mathbb{E}_{\mu^{(0,\bar{e}_{-1})}}[t_i]\right) \\
&\leq \sum_{i=2}^{n} u\left(\frac{x}{n-1}\right) \\
&= (n-1)u\left(\frac{x}{n-1}\right),
\end{aligned}$$

where the third inequality follows from concavity of u, and the fourth inequality follows from concavity of u together with the fact that $\sum_{i=2}^{n} \mathbb{E}_{\mu^{(0,\bar{e}_{-1})}}\left[t_i\right] \leq x$. □

For the special case of $n = 2$, we can now replace the previous Lemmas 5 and 6 by the following result.

Lemma 11 *Suppose $n = 2$. For every contract Φ that implements a pure-strategy effort profile $\bar{e}$, there exists a contest C_y that also implements $\bar{e}$ and yields the same expected payoff to the principal.*

Proof. Suppose Φ implements $\bar{e}$. By Lemma 10 it holds that $c_1(\bar{e}_1) + c_2(\bar{e}_2) \leq u(x)$, so in particular $c_i(\bar{e}_i) \leq u(x)$ for both $i = 1, 2$. We now construct a contest C_y with prize profile $y = (x, 0)$ that also implements $\bar{e}$. The case where $x = 0$ and thus $\bar{e}_1 = \bar{e}_2 = 0$ is trivial, so we focus on the case where $x > 0$.

The allocation rule of C_y is as follows. If $e = \bar{e}$, the zero prize is given to agent i with probability $p_i \geq 0$, while the other agent obtains the positive prize. Below we will determine the values p_i such that $p_1 + p_2 = 1$. If $e = (e_i, \bar{e}_{-i})$ with $e_i \neq \bar{e}_i$, the deviating agent i obtains the zero prize for sure and the non-deviating agent obtains the positive prize. For all other effort profiles e, the allocation of the prizes can be chosen arbitrarily. Since C_y is a contest, it is credible.

For each agent $i = 1, 2$, first define $\tilde{p}_i$ implicitly by

$$(1 - \tilde{p}_i)u\,(x) = c_i(\bar{e}_i).$$

Since the LHS of this equation describes the expected payoff of agent i who expects to obtain the zero prize with probability $\tilde{p}_i$, it follows that the contest C_y indeed implements $\bar{e}$ if $p_i \leq \tilde{p}_i$ holds for both $i = 1, 2$. The fact that $c_i(\bar{e}_i) \leq u(x)$ guarantees $\tilde{p}_i \geq 0$ for both $i = 1, 2$. Lemma 10 also implies that

$$\sum_{i=1}^{2} c_i(\bar{e}_i) = \sum_{i=1}^{2}(1 - \tilde{p}_i)u(x) = (2 - \tilde{p}_1 - \tilde{p}_2)u(x) \leq u(x),$$

which guarantees that $\tilde{p}_1 + \tilde{p}_2 \geq 1$. It is therefore possible to find equilibrium punishment probabilities p_i such that $0 \leq p_i \leq \tilde{p}_i$ and $p_1 + p_2 = 1$.

Finally, from the above arguments we also obtain that the principal's expected payoff is $\bar{e}_1 + \bar{e}_2 - x$ with both $(\bar{e}, \Phi)$ and $(\bar{e}, C_y)$. □

Since the principal can achieve the same payoff with a contest as with any other contract, it only remains to be shown that a solution to the principal's problem exists in the heterogeneous cost case. This will be established in the proof of Proposition 5 below. ■

B.6 Proof of Proposition 4

The proof proceeds in two steps. First, we state some properties that must be satisfied by the optimal contract for any given profile of effort cost functions. Second, we prove the main statement about the optimality of contests for large m.

Step 1. Fix any profile of cost functions $(c_1, \ldots, c_n)$. As argued in the proof of Proposition 3, we can restrict attention to non-screening contracts with expected sums of transfers (x) and expected sums of utilities $(\hat{x})$ that are fixed across $e \in E$, and pure-strategy effort profiles. The following result provides a lower bound on the principal's maximal profits within that class. Define the bound

$$\underline{\Pi} = \max_{x \in [0,\bar{T}]} \left[c_1^{-1}(u(x)) - x \right],$$

which satisfies $\underline{\Pi} > 0$ due to our assumptions on c_1 and u.

Lemma 12 *There exists a contest C_y that implements some effort profile $e \in E$ such that $\Pi_P(e, C_y) = \underline{\Pi}$.*

Proof. Let $x^* = \arg\max_{x \in [0,\bar{T}]} \left[c_1^{-1}(u(x)) - x \right]$ and $e_1^* = c_1^{-1}(u(x^*))$. Consider a contest C_y with prize profile $y = (x^*, 0, \ldots, 0)$. If the effort profile e is such that $e_1 = e_1^*$, then agent 1 receives the prize x^* while all other agents receive a zero prize. For any other effort profile, agent 2 receives x^* and all other agents receive a zero prize. It follows immediately that C_y implements $(e_1^*, 0, \ldots, 0)$ and yields the payoff $e_1^* - x^* = \underline{\Pi}$ to the principal. □

The next result is a direct generalization of Lemma 5 to the case of heterogeneous cost functions. The proof proceeds exactly like the proof of Lemma 5 and is therefore omitted.

Lemma 13 *For every contract Φ that implements a pure-strategy profile $\bar{e} = (\bar{e}_1, ..., \bar{e}_n)$ such that*

$$\frac{1}{n} \sum_{i=1}^{n} c_i(\bar{e}_i) > c_k \left(\frac{1}{n} \sum_{i=1}^{n} \bar{e}_i \right) \quad \forall k \in I,$$

there exists a contract $\hat{\Phi}$ that implements the pure-strategy profile $\hat{e} = (\hat{e}_1, ..., \hat{e}_n)$, where $\hat{e}_1 = \ldots = \hat{e}_n = \frac{1}{n} \sum_{i=1}^{n} \bar{e}_i$, and yields the same expected payoff to the principal.

To summarize, in the search for an optimal contract we can restrict attention to non-screening contracts Φ with an expected sum of transfers x and an expected sum of utilities $\hat{x}$ that are fixed across all $e \in E$, and which implement a pure-strategy effort profile $\bar{e}$ such that $\Pi_P(\bar{e}, \Phi) \geq \underline{\Pi}$ and

$$\frac{1}{n} \sum_{i=1}^{n} c_i(\bar{e}_i) \leq \max_{k \in I} c_k \left(\frac{1}{n} \sum_{i=1}^{n} \bar{e}_i \right). \tag{6}$$

Step 2. Consider any sequence $(c_1^m, \dots, c_n^m)_{m\in\mathbb{N}}$ where, for each agent $i \in I$, the cost functions c_i^m converge uniformly to the common cost function c as $m \to \infty$. Let $(\bar{e}^m, \Phi^m)_{m\in\mathbb{N}}$ be an arbitrary sequence such that contract Φ^m implements effort profile $\bar{e}^m = (\bar{e}_1^m, \dots, \bar{e}_n^m)$ when the cost functions are $(c_1^m, \dots, c_n^m)$. We will write $\hat{e}^m = (1/n)\sum_{i=1}^n \bar{e}_i^m$ for the average effort at step m in the sequence. Without loss of generality, assume that the conditions summarized at the end of Step 1 are satisfied for each $m \in \mathbb{N}$.

Lemma 14 *The sequence*

$$\kappa^m = \max_{k\in I} c_k^m(\bar{e}_k^m) - \frac{1}{n}\sum_{i=1}^n c_i^m(\bar{e}_i^m)$$

converges to zero as $m \to \infty$.

Proof. For every $m \in \mathbb{N}$, let

$$\delta^m = \max_{k\in I} c_k^m(\hat{e}^m) - \frac{1}{n}\sum_{i=1}^n c_i^m(\bar{e}_i^m) \quad \text{and} \quad \psi^m = \max_{i\in I} c_i^m(\bar{e}_i^m) - \max_{k\in I} c_k^m(\hat{e}^m),$$

and hence $\kappa^m = \delta^m + \psi^m$. We will show that $\lim_{m\to\infty}\delta^m = \lim_{m\to\infty}\psi^m = 0$, which immediately implies that $\lim_{m\to\infty}\kappa^m = 0$.

For the sequence δ^m, note that

$$\delta^m = \left[\max_{k\in I} c_k^m(\hat{e}^m) - c(\hat{e}^m)\right] + \left[\frac{1}{n}\sum_{i=1}^n c(\bar{e}_i^m) - \frac{1}{n}\sum_{i=1}^n c_i^m(\bar{e}_i^m)\right] + \left[c(\hat{e}^m) - \frac{1}{n}\sum_{i=1}^n c(\bar{e}_i^m)\right].$$

By uniform convergence of c_i^m to c, $\forall i \in I$, we have

$$\lim_{m\to\infty}(c_i^m(\hat{e}^m) - c(\hat{e}^m)) = 0 \quad \text{and} \quad \lim_{m\to\infty}(c_i^m(\bar{e}_i^m) - c(\bar{e}_i^m)) = 0 \ \ \forall i \in I, \tag{7}$$

and thus

$$\lim_{m\to\infty}\max_{k\in I}(c_k^m(\hat{e}^m) - c(\hat{e}^m)) = 0 \quad \text{and} \quad \lim_{m\to\infty}\left(\frac{1}{n}\sum_{i=1}^n c_i^m(\bar{e}_i^m) - \frac{1}{n}\sum_{i=1}^n c(\bar{e}_i^m)\right) = 0.$$

In addition, by convexity of c we have $c(\hat{e}^m) - \frac{1}{n}\sum_{i=1}^n c(\bar{e}_i^m) \le 0$ for all $m \in \mathbb{N}$, and by condition (6) we have $\delta^m \ge 0$ for all $m \in \mathbb{N}$. Hence, we must also have

$$\lim_{m\to\infty}\left(c(\hat{e}^m) - \frac{1}{n}\sum_{i=1}^n c(\bar{e}_i^m)\right) = 0, \tag{8}$$

as otherwise for some large m we would have $\delta^m < 0$, a contradiction. This concludes that $\lim_{m\to\infty}\delta^m = 0$.

For the sequence ψ^m, we have

$$\psi^m = \max_{k \in I}(c(\hat{e}^m) - c_k^m(\hat{e}^m)) + \max_{i \in I}\left[c_i^m(\bar{e}_i^m) - c(\bar{e}_i^m) + c(\bar{e}_i^m) - c(\hat{e}^m)\right].$$

Hence, by (7), a sufficient condition for $\lim_{m\to\infty} \psi^m = 0$ is

$$\lim_{m\to\infty}(c(\bar{e}_i^m) - c(\hat{e}^m)) = 0 \;\; \forall i \in I. \tag{9}$$

To establish (9), we first claim that there exists $\tilde{e} > 0$ such that $\bar{e}_i^m \in [0, \tilde{e}]$ for all $i \in I$ and all $m \in \mathbb{N}$. The fact that Φ^m implements $\bar{e}^m$ implies $c_i^m(\bar{e}_i^m) \leq u(\bar{T})$ for all $i \in I$. Now fix any $\tilde{u} > u(\bar{T})$. By uniform convergence of each c_i^m to c it follows that there exists $\underline{m}' \in \mathbb{N}$ such that for all $m \geq \underline{m}'$,

$$|c_i^m(\bar{e}_i^m) - c(\bar{e}_i^m)| \leq \tilde{u} - u(\bar{T}) \;\; \forall i \in I,$$

which then implies $c(\bar{e}_i^m) \leq \tilde{u}$ and therefore $\bar{e}_i^m \leq c^{-1}(\tilde{u})$. Now just define $\tilde{e}$ as the maximum among $c^{-1}(\tilde{u})$ and the finite number of values $\bar{e}_i^m$ for all $i \in I$ and $m < \underline{m}'$. We next claim that $\lim_{m\to\infty}(\bar{e}_i^m - \hat{e}^m) = 0$ holds for all $i \in I$. By contradiction, assume there exists $i \in I$ and $\epsilon > 0$ such that for all $\underline{m}' \in \mathbb{N}$ there exists $m \geq \underline{m}'$ so that $|\bar{e}_i^m - \hat{e}^m| \geq \epsilon$. Define $E_i = \{(e_1, \ldots, e_n) \in [0, \tilde{e}]^n \mid |e_i - \frac{1}{n}\sum_{j=1}^n e_j| \geq \epsilon\}$. The set E_i is compact and the function $\chi(e) = \frac{1}{n}\sum_{j=1}^n c(e_j) - c\left(\frac{1}{n}\sum_{j=1}^n e_j\right)$ is continuous on E_i, with $\chi(e) > 0$ due to strict convexity of c and $\epsilon > 0$. Hence $\tilde{\epsilon} = \min_{e \in E_i} \chi(e)$ exists and satisfies $\tilde{\epsilon} > 0$. We have thus shown that there exists $\tilde{\epsilon} > 0$ such that for all $\underline{m}' \in \mathbb{N}$ there exists $m \geq \underline{m}'$ so that $\chi(\bar{e}^m) = -(c(\hat{e}^m) - \frac{1}{n}\sum_{i=1}^n c(\bar{e}_i^m)) \geq \tilde{\epsilon}$, contradicting (8). Finally, (9) now follows immediately because $\bar{e}_i^m \in [0, \tilde{e}]$ and $\hat{e}^m \in [0, \tilde{e}]$ and c is continuous on $[0, \tilde{e}]$. $\square$

Next we show that the sum of effort costs is bounded away from zero for large m.

Lemma 15 *There exist $\underline{m}' \in \mathbb{N}$ and $\underline{c} > 0$ such that $\sum_{i=1}^n c_i^m(\bar{e}_i^m) \geq \underline{c}$ for all $m \geq \underline{m}'$.*

Proof. Let $\underline{\Pi}^m = \max_{x \in [0,\bar{T}]} \underline{\Pi}_1^m(x)$ with $\underline{\Pi}_1^m(x) = (c_1^m)^{-1}(u(x)) - x$ be the lower profit bound for the cost functions $(c_1^m, \ldots, c_n^m)$ as defined in Step 1 of the proof. Hence we know that $\Pi_P(\bar{e}^m, \Phi^m) \geq \underline{\Pi}^m$ holds for all $m \in \mathbb{N}$. Similarly, let $\underline{\Pi}^\infty = \max_{x \in [0,\bar{T}]} \underline{\Pi}_1(x)$ with $\underline{\Pi}_1(x) = c^{-1}(u(x)) - x$ be the bound when the cost functions are $(c, \ldots, c)$. We first claim that $\lim_{m\to\infty} \underline{\Pi}^m = \underline{\Pi}^\infty$. The claim follows immediately once we show that $\underline{\Pi}_1^m$ converges uniformly to $\underline{\Pi}_1$ on $[0, \bar{T}]$. Using Theorem 2 in Barvinek, Daler, and Francu (1991), it can be shown that $(c_1^m)^{-1}$ converges uniformly to c^{-1} on $[0, u(\bar{T})]$.[33] Thus for every $\epsilon > 0$ there

[33] The theorem is directly applicable and implies our claim after we extend the functions c and c_1^m to $\mathbb{R}$ by defining $c_1^m(e) = c(e) = e$ for all $e < 0$.

exists $\underline{m}'' \in \mathbb{N}$ such that for all $m \geq \underline{m}''$,

$$|\underline{\Pi}_1^m(x) - \underline{\Pi}_1(x)| = |(c_1^m)^{-1}(u(x)) - c^{-1}(u(x))| < \epsilon$$

for all $x \in [0, \bar{T}]$, which establishes uniform convergence.

Now fix any ϵ such that $0 < \epsilon < \underline{\Pi}^\infty$ and define $\tilde{\Pi} = \underline{\Pi}^\infty - \epsilon > 0$. Hence there exists $\underline{m}''' \in \mathbb{N}$ such that for all $m \geq \underline{m}'''$,

$$\sum_{i=1}^n \bar{e}_i^m \geq \Pi_P(\bar{e}^m, \Phi^m) \geq \underline{\Pi}^m \geq \tilde{\Pi} > 0.$$

Define

$$\underline{c}^m = \min_{e \in E} \sum_{i=1}^n c_i^m(e_i) \quad \text{s.t.} \quad \sum_{i=1}^n e_i = \tilde{\Pi}.$$

We then obtain that $\sum_{i=1}^n c_i^m(\bar{e}_i^m) \geq \underline{c}^m$ for all $m \geq \underline{m}'''$. Similarly, define

$$\underline{c}^\infty = \min_{e \in E} \sum_{i=1}^n c(e_i) \quad \text{s.t.} \quad \sum_{i=1}^n e_i = \tilde{\Pi},$$

noting that $\underline{c}^\infty > 0$. It again follows from uniform convergence of c_i^m to c for each $i \in I$ that $\lim_{m \to \infty} \underline{c}^m = \underline{c}^\infty$. Fix any ϵ' such that $0 < \epsilon' < \underline{c}^\infty$ and define $\underline{c} = \underline{c}^\infty - \epsilon' > 0$. It follows that there exists $\underline{m}' \in \mathbb{N}$ such that for all $m \geq \underline{m}'$,

$$\sum_{i=1}^n c_i^m(\bar{e}_i^m) \geq \underline{c}^m \geq \underline{c},$$

which completes the proof. □

We can now combine Lemmas 14 and 15 to obtain the following result.

Lemma 16 *There exists $\underline{m} \in \mathbb{N}$ such that for all $m \geq \underline{m}$,*

$$\max_{k \in I} c_k^m(\bar{e}_k^m) \leq \frac{1}{n-1} \sum_{i=1}^n c_i^m(\bar{e}_i^m).$$

Proof. By Lemma 15, there exist $\underline{m}' \in \mathbb{N}$ and $\underline{c} > 0$ such that $\sum_{i=1}^n c_i^m(\bar{e}_i^m) \geq \underline{c}$ for all $m \geq \underline{m}'$. In addition, from the limiting statement about κ^m in Lemma 14 we can conclude that there exists $\underline{m}'' \in \mathbb{N}$ such that for all $m \geq \underline{m}''$,

$$\max_{k \in I} c_k^m(\bar{e}_k^m) - \frac{1}{n} \sum_{i=1}^n c_i^m(\bar{e}_i^m) \leq \frac{\underline{c}}{n(n-1)}.$$

Thus for all $m \geq \underline{m} = \max\{\underline{m}', \underline{m}''\}$ we obtain

$$\begin{aligned}
\max_{k\in I} c_k^m(\bar{e}_k^m) - \frac{1}{n-1}\sum_{i=1}^n c_i^m(\bar{e}_i^m) &= \max_{k\in I} c_k^m(\bar{e}_k^m) - \frac{1}{n}\sum_{i=1}^n c_i^m(\bar{e}_i^m) - \frac{1}{n(n-1)}\sum_{i=1}^n c_i^m(\bar{e}_i^m) \\
&\leq \frac{\underline{c}}{n(n-1)} - \frac{1}{n(n-1)}\sum_{i=1}^n c_i^m(\bar{e}_i^m) \\
&\leq 0. \qquad \square
\end{aligned}$$

Now consider the sequence for any $m \geq \underline{m}$ as given in Lemma 16. Combined with Lemma 10 we can conclude that

$$\max_{k\in I} c_k^m(\bar{e}_k^m) \leq u\left(\frac{x^m}{n-1}\right) \tag{10}$$

holds, where $x^m > 0$ is the expected sum of transfers in contract Φ^m. We now claim that there exists a contest C_y^m which also implements $\bar{e}^m$ and yields the same expected payoff to the principal. Let the prize profile be given by

$$y = \left(\frac{x^m}{n-1}, \ldots, \frac{x^m}{n-1}, 0\right).$$

The allocation rule of C_y^m is as follows. If $e = \bar{e}^m$, the zero prize is given to agent i with probability $p_i^m \geq 0$, while all other agents obtain one of the identical positive prizes. Below we will determine the values p_i^m such that $\sum_{i=1}^n p_i^m = 1$. If $e = (e_i, \bar{e}_{-i}^m)$ with $e_i \neq \bar{e}_i^m$ for some $i \in I$, the deviating agent i obtains the zero prize for sure and all other agents obtain one of the identical positive prizes. For all other effort profiles e, the allocation of the prizes can be chosen arbitrarily. Since C_y^m is a contest, it is credible.

For each agent $i \in I$, first define $\tilde{p}_i^m$ implicitly by

$$(1-\tilde{p}_i^m)u\left(\frac{x^m}{n-1}\right) = c_i^m(\bar{e}_i^m).$$

Since the LHS of this equation describes the expected payoff of agent i who expects to obtain the zero prize with probability $\tilde{p}_i^m$, it follows that the contest C_y^m indeed implements $\bar{e}^m$ if $p_i^m \leq \tilde{p}_i^m$ holds for all $i \in I$. The fact that $c_i^m(\bar{e}_i^m) \leq u(x^m/(n-1))$ for all $i \in I$ due to that (10) guarantees $\tilde{p}_i^m \geq 0$. Lemma 10 also implies that

$$\sum_{i=1}^n c_i^m(\bar{e}_i^m) = \sum_{i=1}^n (1-\tilde{p}_i^m)u\left(\frac{x^m}{n-1}\right) = \left(n - \sum_{i=1}^n \tilde{p}_i^m\right)u\left(\frac{x^m}{n-1}\right) \leq (n-1)u\left(\frac{x^m}{n-1}\right),$$

which guarantees that $\sum_{i=1}^n \tilde{p}_i^m \geq 1$. It is therefore possible to find equilibrium punishment probabilities p_i^m such that $0 \leq p_i^m \leq \tilde{p}_i^m$ $\forall i \in I$ and $\sum_{i=1}^n p_i^m = 1$. The principal's expected payoff is then $\sum_{i=1}^n \bar{e}_i^m - x^m$ with both $(\bar{e}^m, \Phi^m)$ and $(\bar{e}^m, C_y^m)$.

In sum, for any $m \geq \underline{m}$, the principal can achieve a weakly higher payoff with a contest than with any other contract. Hence it only remains to be shown that a solution to the principal's problem exists in the heterogeneous cost case. This will be established in the proof of Proposition 5 below. ■

B.7 Proof of Proposition 5

First, we characterize the set of all pure-strategy effort profiles which can be implemented. Define

$$E^1 = \left\{ e \in E \mid \sum_{i=1}^{n} c_i(e_i) \leq (n-1)u\left(\frac{\bar{T}}{n-1}\right) \right\},$$
$$E^2 = \left\{ e \in E \mid \sum_{i=1}^{n} u^{-1}(c_i(e_i)) \leq \bar{T} \right\}.$$

The following result provides a necessary condition for implementability.

Lemma 17 *A pure-strategy effort profile $\bar{e} \in E$ can be implemented by some contract only if $\bar{e} \in E^1 \cap E^2$.*

Proof. Suppose $\bar{e} \in E$ is implemented by a contract Φ. By contradiction, assume first that $\bar{e} \notin E^1$. Then $\sum_{i=1}^{n} c_i(\bar{e}_i) > (n-1)u(\bar{T}/(n-1))$. By Lemma 10, it also holds that $\sum_{i=1}^{n} c_i(\bar{e}_i) \leq (n-1)u\left(x/(n-1)\right)$, where x is the expected sum of transfers in Φ. Combining the two inequalities, we obtain $x > \bar{T}$. Hence Φ allocates transfer profiles which are not feasible, a contradiction.

Suppose now that $\bar{e} \notin E^2$. From (IC-A) it is clear that $c_i(\bar{e}_i) \leq \mathbb{E}_{\mu^{\bar{e}}}\left[u(t_i)\right] \leq u(\mathbb{E}_{\mu^{\bar{e}}}\left[t_i\right])$ must hold, and hence $u^{-1}(c_i(\bar{e}_i)) \leq \mathbb{E}_{\mu^{\bar{e}}}\left[t_i\right]$, for all $i \in I$. Since $\bar{e} \notin E^2$, we then obtain $x = \sum_{i=1}^{n} \mathbb{E}_{\mu^{\bar{e}}}\left[t_i\right] > \bar{T}$, again a contradiction. □

We now restrict attention to effort profiles $\bar{e} \in E^1 \cap E^2$, assuming without loss of generality that $c_i(\bar{e}_i) \geq c_j(\bar{e}_j)$ if $i < j$. For any such profile, define x' implicitly by

$$u\left(\frac{x'}{n-1}\right) = \frac{1}{n-1}\sum_{i=1}^{n} c_i(\bar{e}_i).$$

The fact that $\bar{e} \in E_1$ implies that x' is well-defined and satisfies $x' \leq \bar{T}$. Furthermore, for each $i \in I$ let $\tilde{x}_i = u^{-1}(c_i(e_i))$ and

$$\tilde{x} = \sum_{i=1}^{n} \tilde{x}_i,$$

which is well-defined and satisfies $\tilde{x} \leq \bar{T}$ because $\bar{e} \in E_2$.

Lemma 18 *If $\bar{e} \in E^1 \cap E^2$, then $\bar{e}$ can be implemented by a generalized contest with an expected sum of transfers* $\max\{x', \tilde{x}\}$.

Proof. Fix any $\bar{e} \in E^1 \cap E^2$. We will construct a generalized contest that implements $\bar{e}$ with an expected sum of transfers $\max\{x', \tilde{x}\} \leq \bar{T}$. We examine two cases separately.

Case 1: $x' \geq \tilde{x}$. Define $y^d = (x'/(n-1), \ldots, x'/(n-1), 0)$. This vector of prizes will be used after any unilateral deviation from $\bar{e}$, so that the deviator gets the zero prize while all other agents get $x'/(n-1)$. Next, we construct the vector of prizes y that will be used to reward the agents in equilibrium. For every $\epsilon \in [0, \tilde{x}_n]$, let

$$y_n(\epsilon) = \tilde{x}_n - \epsilon \quad \text{and} \quad y_i(\epsilon) = \tilde{x}_i + \frac{x' - \tilde{x} + \epsilon}{n-1} \quad \forall i < n.$$

By construction $\sum_{i=1}^n y_i(\epsilon) = x'$. In addition, we have

$$\sum_{i=1}^n u(y_i(0)) \geq \sum_{i=1}^n c_i(\bar{e}_i) = (n-1)u\left(\frac{x'}{n-1}\right) \geq \sum_{i=1}^n u(y_i(\tilde{x}_n)), \tag{11}$$

where the first inequality follows from $y_i(0) \geq \tilde{x}_i$ for all $i \in I$, and the second inequality follows from $y_n(\tilde{x}_n) = 0$, $\sum_{i=1}^n y_i(\tilde{x}_n) = x'$, and the fact that u is concave. Since u is continuous, (11) implies that there exists some $\epsilon^* \in [0, \tilde{x}_n]$ such that

$$\sum_{i=1}^n u(y_i(\epsilon^*)) = \sum_{i=1}^n c_i(\bar{e}_i) = (n-1)u\left(\frac{x'}{n-1}\right). \tag{12}$$

Let $y = y(\epsilon^*) = (y_1(\epsilon^*), ..., y_n(\epsilon^*))$. It follows by construction that any contract which allocates only permutations of the vectors y or y^d is credible.

Since the transfer to a unilateral deviator from $\bar{e}$ is zero, the only remaining step is to show that there is a way to allocate the equilibrium prizes y such that the expected utility of each agent i is at least $c_i(\bar{e}_i)$.

Consider the following n transfer vectors, each of which is a permutation of y. For all $i \in I$, $t^i \in T$ is given by

$$t^i_j = \begin{cases} y_n(\epsilon^*) & \text{if } j = i, \\ y_i(\epsilon^*) & \text{if } j = n, \\ y_j(\epsilon^*) & \text{if } j \neq i, n. \end{cases}$$

In words, in transfer vector t^n each agent j receives the transfer $y_j(\epsilon^*)$. For any $i \neq n$, the transfer vector t^i is the same as t^n except that the transfers of agents i and n are swapped. Next, we will construct the probabilities with which the transfer vectors t^i are allocated in equilibrium.

If $\epsilon^* = 0$, then we have $u(y_i(\epsilon^*)) \geq c_i(\bar{e}_i)$ for all $i \in I$. In this case, we let the transfer vector t^n be allocated with probability one in equilibrium, so that the required expected utility is achieved for each $i \in I$.

If $\epsilon^* > 0$, then we have $u(y_i(\epsilon^*)) > c_i(\bar{e}_i) \geq c_n(\bar{e}_n) > u(y_n(\epsilon^*))$ for all $i < n$. Hence, for all $i < n$ there exists a unique $p^i \in (0, 1)$ such that

$$p^i u(y_n(\epsilon^*)) + (1 - p^i) u(y_i(\epsilon^*)) = c_i(\bar{e}_i). \tag{13}$$

In addition, we define $p^n = 1 - \sum_{i=1}^{n-1} p^i$. To show that $(p^1, ..., p^n)$ is a well-defined probability vector, we only need to establish $p^n \geq 0$. This property indeed holds because

$$\begin{aligned}
\sum_{i=1}^{n-1} p^i &= \sum_{i=1}^{n-1} \frac{u(y_i(\epsilon^*)) - c_i(\bar{e}_i)}{u(y_i(\epsilon^*)) - u(y_n(\epsilon^*))} \\
&\leq \sum_{i=1}^{n-1} \frac{u(y_i(\epsilon^*)) - c_i(\bar{e}_i)}{u(y_{n-1}(\epsilon^*)) - u(y_n(\epsilon^*))} \\
&= \frac{\sum_{i=1}^{n} u(y_i(\epsilon^*)) - u(y_n(\epsilon^*)) - \sum_{i=1}^{n-1} c_i(\bar{e}_i)}{u(y_{n-1}(\epsilon^*)) - u(y_n(\epsilon^*))} \\
&= \frac{\sum_{i=1}^{n} c_i(\bar{e}_i) - u(y_n(\epsilon^*)) - \sum_{i=1}^{n-1} c_i(\bar{e}_i)}{u(y_{n-1}(\epsilon^*)) - u(y_n(\epsilon^*))} \\
&= \frac{c_n(\bar{e}_n) - u(y_n(\epsilon^*))}{u(y_{n-1}(\epsilon^*)) - u(y_n(\epsilon^*))} \\
&< 1.
\end{aligned}$$

Now consider the allocation rule where each vector t^i is allocated with probability p^i. With this rule, each agent $i < n$ receives the prize $y_n(\epsilon^*)$ with probability p^i and the prize $y_i(\epsilon^*)$ with the opposite probability $1 - p^i$. By (13), this yields an expected utility equal exactly to $c_i(\bar{e}_i)$. As for agent n, he receives each prize $y_i(\epsilon^*)$ with probability p^i. Hence, his expected utility is given by

$$\begin{aligned}
\sum_{i=1}^{n} p^i u(y_i(\epsilon^*)) &= \sum_{i=1}^{n-1} p^i u(y_i(\epsilon^*)) + \left(1 - \sum_{i=1}^{n-1} p^i\right) u(y_n(\epsilon^*)) \\
&= \sum_{i=1}^{n-1} p^i \left[u(y_i(\epsilon^*)) - u(y_n(\epsilon^*))\right] + u(y_n(\epsilon^*)) \\
&= \sum_{i=1}^{n-1} \left[u(y_i(\epsilon^*)) - c_i(\bar{e}_i)\right] + u(y_n(\epsilon^*)) \\
&= \sum_{i=1}^{n} u(y_i(\epsilon^*)) - \sum_{i=1}^{n-1} c_i(\bar{e}_i)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} c_i(\bar{e}_i) - \sum_{i=1}^{n-1} c_i(\bar{e}_i) \\
&= c_n(\bar{e}_n),
\end{aligned}
$$

where the third equality follows (13), and the fifth equality follows (12). Thus, agent n also receives an expected utility exactly equal to his effort cost $c_n(\bar{e}_n)$. We can conclude that the generalized contest implements $\bar{e}$ with an expected sum of transfers x'.

Case 2: $x' < \tilde{x}$. Let $y = (\tilde{x}_1, \ldots, \tilde{x}_n)$. Consider a generalized contest where in equilibrium, when $e = \bar{e}$, each agent i receives $\tilde{x}_i$ with probability one, and is thus exactly compensated for the cost of effort.

Next, we construct the vector of prizes y^d that will be used after unilateral deviations. Note that $(n-1)u(\tilde{x}/(n-1)) > (n-1)u(x'/(n-1)) = \sum_{i=1}^{n} c_i(\bar{e}_i)$. In addition, by concavity of u and $u(0) = 0$ we have $u(\tilde{x}) \leq \sum_{i=1}^{n} u_i(\tilde{x}_i) = \sum_{i=1}^{n} c_i(\bar{e}_i)$. But then by continuity of u there must exist $\varepsilon^* \in \left(0, \frac{\tilde{x}}{n-1}\right]$ such that

$$
(n-2)u\left(\frac{\tilde{x}}{n-1} - \varepsilon^*\right) + u\left(\frac{\tilde{x}}{n-1} + (n-2)\varepsilon^*\right) = \sum_{i=1}^{n} c_i(\bar{e}_i) = (n-1)u\left(\frac{x'}{n-1}\right).
$$

Now define

$$
y^d = \left(\frac{\tilde{x}}{n-1} + (n-2)\epsilon^*, \frac{\tilde{x}}{n-1} - \varepsilon^*, \ldots, \frac{\tilde{x}}{n-1} - \epsilon^*, 0\right).
$$

Let the zero prize be allocated to a unilateral deviator from $\bar{e}$, while the other prizes in y^d are allocated randomly among the remaining agents. By construction of y and y^d, this generalized contest is credible, and it follows immediately that it implements $\bar{e}$, with an expected sum of transfers $\tilde{x}$. □

Lemma 19 *If a pure-strategy effort profile $\bar{e}$ is implementable, then the lowest expected sum of transfers at which it can be implemented is* $\max\{x', \tilde{x}\}$.

Proof. First, observe that any contract implementing $\bar{e}$ has to satisfy $c_i(\bar{e}_i) \leq \mathbb{E}_{\mu^{\bar{e}}}[u(t_i)] \leq u(\mathbb{E}_{\mu^{\bar{e}}}[t_i])$ and hence $\tilde{x}_i = u^{-1}(c_i(\bar{e}_i)) \leq \mathbb{E}_{\mu^{\bar{e}}}[t_i]$ for all $i \in I$. Summing up over all $i \in I$ implies that $\tilde{x} \leq \sum_{i=1}^{n} \mathbb{E}_{\mu^{\bar{e}}}[t_i]$, so that $\tilde{x}$ is a lower bound on the expected sum of transfers in any contract implementing $\bar{e}$.

When $x' < \tilde{x}$, the construction given in Case 2 of the proof of Lemma 18 implements $\bar{e}$ at exactly $\tilde{x}$, so the lower bound is reached.

When $x' \geq \tilde{x}$, the construction given in Case 1 of the proof of Lemma 18 implements $\bar{e}$ at exactly x'. We claim that in this case x' is in fact the lowest expected sum of transfers at which $\bar{e}$ can be implemented. Suppose by contradiction that there exists a contract which implements $\bar{e}$ at $x < x'$. By Lemma 10 and the definition of x', we have

$$u\left(\frac{x'}{n-1}\right) = \frac{1}{n-1}\sum_{i=1}^{n} c_i(\bar{e}_i) \le u\left(\frac{x}{n-1}\right),$$

which implies $x \ge x'$, a contradiction. $\square$

To complete the proof of Proposition 5, we finally establish the existence of a solution.

Lemma 20 *An optimal contract exists.*

Proof. First, observe that the set of implementable effort profiles $E^1 \cap E^2$ is compact. By Lemma 19, the minimum expected sum of transfers for implementing an effort profile $e \in E^1 \cap E^2$ is given by $\max\{x'(e), \tilde{x}(e)\}$, where

$$x'(e) = (n-1)u^{-1}\left(\frac{1}{n-1}\sum_{i=1}^{n} c_i(e_i)\right) \quad \text{and} \quad \tilde{x}(e) = \sum_{i=1}^{n} u^{-1}(c_i(e_i))$$

are continuous functions of e. Thus, the principal's problem can be reformulated as

$$\max_{e \in E^1 \cap E^2} \left[\sum_{i=1}^{n} e_i - \max\{x'(e), \tilde{x}(e)\}\right],$$

which amounts to the maximization of a continuous function on a compact set. Hence a solution exists. $\square$ $\blacksquare$