Performance-Maximizing Contests

Wojciech Olszewski and Ron Siegel*

November 2017

Abstract

Many sales, sports, and research contests are put in place to maximize contestants' performance. We investigate and provide a complete characterization of the prize structures that achieve this objective in settings with many contestants. The contestants may be ex-ante asymmetric in their abilities and prize valuations, and there may be complete or incomplete information about these parameters. The contestants may be risk neutral, risk averse, or risk seeking, and their performance cost may be linear, concave, or convex. A main takeaway is that awarding numerous prizes whose values gradually decline with contestants' ranking is optimal in the typical case of risk averse contestants that have a convex cost of performance.

*Department of Economics, Northwestern University, Evanston, IL 60208 (e-mail: wo@northwestern.edu) and The Pennsylvania State University, University Park, IL 16802 (e-mail: rus41@psu.edu). We thank Bruno Strulovici and the audiences at Bocconi University, Boston College, the University of British Columbia, Cambridge University, Carnegie Mellon University, Duke University, Johns Hopkins University, Kyoto University, Leicester University, the University of Melbourne, the University of Montreal, the University of New South Wales, Oxford University, the Pennsylvania State University, the University of Pittsburgh, the university of Rochester, the University of Queensland, the University of Southampton, the University of Wisconsin (Madison), the Conference in Honor of Sergiu Hart at Hebrew University, the Econometric Society World Congress in Montreal, the Midwest Economic Theory Conference at the Pennsylvania State University, the University of Hong Kong, the Conference on Economic Design in Istanbul, the Canadian Economic Theory Conference in London, Ontario, the Summer North American Econometric Society Meeting in Philadelphia, the Current Frontiers in the Theory of Contests Workshop in Münich, the 8th Shanghai Microeconomics Workshop, and the Cowles Summer Economic Theory Conference for very helpful comments and suggestions. Financial support from the NSF (grant SES-1325968) is gratefully acknowledged.

1 Introduction

Contests are used in a variety of settings to motivate people and increase their performance. In the context of sales, contests are employed for this purpose by many large firms. For example, Cisco Systems, one of the largest technology companies, regularly runs sales contests among its thousands of partners to boost sales. It publishes the precise criteria used to rank the contestants, and awards prizes to those with the highest performance.[1] Two other examples are HubSpot, a three-billion dollar software company, and Clayton Homes, the largest builder of manufactured housing and modular homes in the United States, whose use of sales contests was reviewed by the Harvard Business Review.[2] In the context of entertainment, sporting contests play an important role, generating an annual attendance that according to the U.S. Census Bureau exceeded 200 million in 2010. Szymanski (2003) discusses the design of sporting contests from an economic perspective, and notes that because organizers generate a profit by selling tickets and "spectators will be attracted by the quality of the field entering the race and the effort the entrants contribute," a reasonable objective of a sporting contest is "to maximize the effort contribution of the selected entrants." In the context of academia, many agencies that fund basic research, such as the European Research Council and the National Science Foundation, administer large contests that motivate researchers to generate high-quality research proposals.

Inspired by the wide range of contests put in place to maximize contestants' performance, we study the prize structures that achieve this objective.[3] That is, we study settings in which a planner awards prizes based on the rank order of contestants' performance and has substan-

[1] A recent example is the 2017 "Cisco Commercial Champs Sales Competition - Win a trip to Taipei, Taiwan," in which 79 round-trip vacation packages were awarded to the highest-performing partners. See https://www.cisco-commercialxcelerate.com/AppFiles/pdf/tnc/sc/CCX_SC_FY17Q3_TnC.pdf

[2] "The Right Way to Use Compensation," Harvard Business Review, April 2015 and "The Sales Director Who Turned Work into a Fantasy Sports Competition," Harvard Business Review, March 2015.

[3] Contests with other objectives give rise to different contest design questions. Many of these questions can be investigated by variants of our analysis, as we discuss in Section 10. One important contest design question for which our methodology is not appropriate is how to design a contest to generate the single best product or innovation, which is relevant in many procurement settings.

tial discretion in dividing the prize budget across prizes, as is the case in many sales, sporting, and research contests.[4] Should a small number of high-value prizes be awarded, or a larger number of lower-value prizes? Or perhaps awarding prizes of different values is optimal? And if so, how should the prize values change with their rank order? This classic contest design question has proven challenging so far, because the equilibria of even simple contest models with a fixed prize structure are often difficult or impossible to derive, and even when they can be derived, it is often only by employing algorithms, which are not conducive to further analysis. Thus, the reasonable approach to finding the performance-maximizing prize structures of solving a family of contest models, one for every prize structure, and identifying the optimal one in the family is impractical, except in severely restricted environments, and even then it typically yields partial results.[5]

We take a different approach, which allows us to provide a complete characterization of the performance-maximizing prize structures in a relatively general environment. The approach, which we describe below, applies to contests with many contestants.[6] The contestants may be ex-ante asymmetric in their abilities and prize valuations, and there may be complete or incomplete information about these parameters. The contestants may be risk neutral, risk averse, or risk seeking, and their performance cost may be linear, concave, or convex. The approach allows us to consider identical prizes, heterogeneous prizes, and a combination of identical and heterogeneous prizes. This is important, because restricting the prize structures a-priori may rule out the optimal ones.

A main takeaway from our analysis is the effect of contestants' risk attitude on the optimal prize structure. This is easiest to see when contestants' performance cost is linear. With

[4]We take the budget as exogenous for most of our analysis. Section 7.2 discusses how to optimally determine the budget when it is not given exogenously.

[5]See Section 9 for some of the existing results.

[6]Settings with a large number of contestants include sales competitions in large firms (Cisco Systems has more than 15,000 partners in the US and regularly runs sales competitions among its partners), sports competitions (between 2010 and 2012, Tokyo, London, New York, Chicago, and Sydney each hosted a marathon with more than 30,000 participants), and research grant competitions (in each of the last several years, the National Science Foundation (NSF) received more than 40,000 grant applications and awarded more than 10,000 grants).

risk aversion, the optimal prize structure consists of a range of positive prizes, all different, with gradually decreasing values. There may in addition be multiple prizes of the highest allowed value, but this does not happen when the highest allowed value is sufficiently high. In contrast, with risk seeking the optimal prize structure consists of as many prizes of the highest allowed value as the budget permits.[7]

Another takeaway from our analysis is the effect of the curvature of contestants' performance cost on the optimal prize structure. The effect of convex costs resembles that of risk aversion, and the effect of concave costs resembles that of risk seeking. In particular, in the most relevant case of risk aversion and convex costs, the optimal prize structure consists of a range of prizes of different values, as described above for risk aversion.[8] While similar, the effects of contestants' cost curvature and risk attitude on the optimal prize structure are not identical. For example, with risk aversion and linear costs the number of prizes is optimally restricted, even when the marginal prize valuation at 0 is infinite, whereas with risk neutrality and convex costs, if the marginal cost at 0 is 0, then almost every contestant is optimally awarded a positive prize.

These results suggest that large sales and workplace competitions aimed at maximizing workers' performance should award many prizes of various values. A similar implication may hold for grant competitions aimed at increasing the overall quality of research in disciplines such as economics, where funding is often not critical for projects' realization.[9] We provide an illustrative example in Section 8.

We also use our characterization of the optimal prize structure to derive some comparative statics. For example, with risk aversion and linear costs, a better pool of contestants (in a sense that implies first-order stochastic dominance) optimally leads to a more homogeneous set of prizes (in the sense of second-order stochastic dominance), and an increase in the

[7]Both of these results hold when the budget is not "too large" in a sense made precise in Section 5.2.

[8]Similarly, for risk seeking and concave costs the optimal prize structure is as descibed above for risk seeking.

[9]In other disciplines, where projects typically require costly equipment, a minimum prize size may be required. A reasonable objective in these settings would be to maximize the quality of the projects that are awarded a prize. Our approach can also be used to study such settings.

prize budget optimally leads to more valuable prizes (in the sense of first-order stochastic dominance). A better pool of contestants (in the sense of first-order stochastic dominance) also leads to higher aggregate performance, even when the prize structure is not optimal. While this last result may seem intuitive, Section 7.2 shows that it does not always hold for small contests.

The idea underlying our approach is to identify and solve a manageable optimization problem whose solution approximates the optimal prize structure. This requires three steps. First, we refer to Olszewski and Siegel (2016), who showed that in a large contest with a fixed prize structure players' equilibrium behavior is approximated by the unique single-agent mechanism that assortatively allocates a continuum of prizes to a continuum of agent types and gives the lowest type a utility of 0. This result is summarized by Theorem 1. The intuition is that with many players the law of large numbers implies that any performance level a player chooses (roughly) leads to a deterministic prize. From the resulting inverse tariff, which maps performance levels to prizes, higher types choose higher performance levels and obtain higher prizes. This leads to the assortative allocation. Second, we show that it suffices to solve for the prize structure that maximizes the performance in the limit single-agent setting. We show this by proving that the optimal prize structure in the limit setting approximates the optimal prize structures in large contests in an upper- and lower-hemicontinuity sense. This is done in Section 4.2. Third, we solve the optimization problem in the limit setting. In the appendix we show that this problem can be formulated as an optimal control problem, but its specific structure prevents us from using off-the-shelf tools to describe the solution. Instead, we solve the problem from first principles by using some ideas from the theories of optimal control and calculus of variations. We first do this for players with linear costs, and then consider the more difficult case of non-linear costs.

The intuition for the optimal prize structure with linear costs follows from a connection to Myerson's (1981) optimal auction with a single buyer. Myerson's optimal auction and the mechanism that approximates the optimal contest both implement monotone allocations that maximize the "virtual surplus" for each type, i.e., the allocation value minus the information rents accrued to higher types. The intuition why the approximating mechanism maximizes the virtual surplus is as follows. An increase in the value of a prize has only two, clear-cut effects: (1) It increases local competition for this prize by types just below the one allocated the prize in equilibrium, and the increased competition increases performance to precisely exhaust the entire benefit from the increase in prize value; (2) it reduces the performance of all higher types, since they can now slack off and obtain a slightly higher prize than they previously could. These two effect are captured by a "virtual performance" expression identical to Myerson's virtual surplus.

Now, suppose that we start with prizes of 0 assigned to all types, and we begin to increase the prizes with the objective of maximizing aggregate performance. We first increase the prize awarded to the highest type, because that type values the prize most, so the effect described in (1) is the strongest. In addition, the effect described in (2) is non-existent, because there are no higher types. If the marginal utility of prizes is increasing, then any previous increase in the prize only magnifies the effect described in (1). So, we keep increasing the prize of the highest type until we reach the highest possible prize. We next increase the prize awarded to the "second-highest" type, and continue in this way until we exhaust the budget. If the marginal utility of prizes is decreasing, then any previous increase in the prize awarded to the highest type reduces the effect described in (1). This makes increasing the prize awarded to the second-highest type more attractive from the perspective of maximizing aggregate performance. Thus, at some point we begin to increase the prize awarded to the second-highest type, which then makes increasing the prize awarded to the third-highest type more attractive, and so on.

The rest of the paper is organized as follows. Section 2 describes the contest environment. Section 3 describes the mechanism design approach to studying large contests. Section 4 formulates the contest design problem and the corresponding mechanism design problem, and shows that it is enough to solve the latter. Section 5 analyzes the optimal prize structure when players have linear costs, and Section 6 extends the analysis to more general costs. This order is motivated by the simplicity and intuition that are gained with linear costs. Nevertheless, the curvature of the cost function has an important effect on the optimal prize structure. Section 5 makes the standard mechanism design assumption of monotone virtual values. In the online appendix, we show that most of the qualitative results (at least for linear costs) continue to hold without this assumption. Section 7 discusses comparative statics, and Section 8 describes an illustrative example of a research grant competition. Section 9 discusses two closely related papers. Section 10 concludes by briefly discussing a few additional contest design questions that can be addressed by our approach. The appendix contains the proofs omitted from the main text.

2 Asymmetric contests

A contest is a game in which n players compete for n prizes. Each player is characterized by a type $x \in [0,1]$, and each prize is characterized by a number $y \in [0,m]$. For concreteness, we will assume that prizes are monetary, so y is simply an amount of money.[10] Prize m is the highest possible prize. This bound on the prize sizes is required by our methods, since the approximation results obtained in Olszewski and Siegel (2016) (henceforth: OS), which are a fundamental tool for our analysis, rely the compactness of the spaces of types and prizes.[11] Beyond being necessary for the analysis, such a prize bound arises naturally in some settings.[12] To apply our results to settings in which such a bound is not imposed, we will find the optimal prize structures for all m, and then take the limit of the optimal structures as m diverges to infinity. The prize values $y_1^n \leq y_2^n \leq \cdots \leq y_n^n$ are commonly known. Some of the values may be 0, so it is without loss of generality to have the same number of prizes as players. Player i's privately known type x_i^n is distributed according to a *cdf* F_i^n, and these distributions, which need not be identical, are commonly known and independent across players. In the special case of complete information, each *cdf* corresponds to a Dirac (degenerate) distribution.

In the contest, each player chooses her performance t, the player with the highest performance obtains the highest prize, the player with the second-highest performance obtains the second-highest prize, and so on. Ties are resolved by a fair lottery. The utility of a player of

[10] More generally, the number y represents the prize's cost, so prize y costs y.

[11] The necessity of assuming compactness will also be apparent in the present analysis. In some cases, it will be optimal to spend the entire budget on the highest possible prizes. With no bound on the prize sizes, an optimal prize structure would not exist.

[12] This can be the result of policy, fairness considerations, or technological limitations. For example, several NSF categories have maximal awards.

type x from exerting performance $t \geq 0$ and obtaining prize y is

$$U(x, y, t) = xh(y) - c(t), \tag{1}$$

where $h(0) = c(0) = 0$, and prize valuation h and performance cost c are continuously differentiable and strictly increasing. Notice that the game is strategically equivalent to one in which players have private information about their performance cost, as in Spence's (1973) signalling model, by dividing the utility by x to obtain $h(y) - c(t)/x$. This has no effect on the results. We assume that sufficiently high performance levels are prohibitively costly, that is, $h(m) < c(t)$ for large enough t, so no player chooses performance higher than $c^{-1}(h(m))$. The functional form (1) and special cases thereof have been assumed in numerous existing papers (see, for example, Clark and Riis (1998), henceforth: CR, Bulow and Levin (2006), henceforth: BL, and Moldovanu and Sela (2001), henceforth: MS).

Our analysis will focus on large contests, that is, contests with a large n. We will consider sequences of contests, and refer to a contest with n players and n prizes as the "n-th contest" in the sequence. Every contest has at least one (mixed-strategy) Bayesian Nash equilibrium.[13]

3 Using mechanism design to study the equilibria of large contests

The optimal design of asymmetric contests of the kind described in Section 2 is difficult or impossible, because no method currently exists for characterizing their equilibria for most type and prize distributions. And even in the few cases for which a characterization exists, the equilibria have a complicated form, or can be derived only by means of algorithms (see, for example , BL, Siegel (2010), and Xiao (2016)), so they are not amenable to further analysis. Our approach to contest design builds on the technique for studying the equilibria of large contests that was developed in OS. We now describe this technique, which allows us to approximate the equilibrium outcomes of large contests by considering the mechanism

[13] This follows, for example, from Corollary 5.2 in Reny (1999), because the mixed extension is better-reply secure.

that implements a particular allocation of a continuum of prizes to a continuum of agent types.

3.1 Limit distributions

The equilibrium approximation technique requires the contests in the sequence to become increasingly similar in some sense as n increases. To formalize this requirement, let $F^n = \left(\sum_{i=1}^{n} F_i^n\right)/n$, so $F^n(x)$ is the expected percentile ranking of type x in the n-th contest given the random vector of players' types. Denote by G^n the empirical prize distribution, which assigns a mass of $1/n$ to each prize y_j^n (recall that there is no uncertainty about the prizes). We require that F^n converge in weak*-topology to a distribution F that has a continuous, strictly positive density f, and that G^n converge to some (not necessarily continuous) distribution G.[14] We then say that the sequence of contests converges.

The assumption that distribution F has a continuous and strictly positive density is required by the approximation results in OS. Notice that this assumption does not imply a similar restriction on distributions F_i^n of players' types, so these distributions may have gaps and atoms.[15] It is important that we do not restrict G, since we will optimize over prize distributions, and any exogenous restriction on the optimal prize distribution would restrict the scope of our analysis.[16]

To get a sense for this convergence of F^n and G^n to limit distributions F and G, consider two extreme cases: asymmetric contests with complete information, and symmetric contests with incomplete information. A simple way to construct a converging sequence of contests with complete information is first to choose the desired limit distributions F and G, and then to set player i's deterministic type in the n-th contest to be $x_i^n = F^{-1}(i/n)$ (so F_i^n is a Dirac distribution concentrated on x_i^n) and prize j in the n-th contest to be $y_j^n = G^{-1}(j/n)$,

[14] Convergence in weak*-topology can be defined as convergence of *cdf*s at points at which the limit *cdf* is continuous (see Billingsley (1995)).

[15] The restriction on F does rule out, for example, a sequence of contests with complete information that have a non-vanishing fraction of identical players.

[16] Section 8 illustrates that our methods can also be used to optimize over a restricted set of prize distributions, by considering an example with identical prizes.

where

$$G^{-1}(z) = \inf\{y : G(y) \geq z\} \text{ for } 0 \leq z \leq 1.$$

Then, the n-th contest is one of complete information, F^n converges to F, and G^n converges to G.

One example is contests with identical prizes and players who differ in their valuations for a prize. For this, consider $h(y) = y$, F uniform, and G that has $G(y) = 1-p$ for all $y \in [0,1)$ and $G(1) = 1$, where $p \in (0,1)$ is the limit ratio of the number of identical (non-zero) prizes to the number of players. Then $x_i^n = i/n$, $y_j^n = 0$ if $j/n \leq 1-p$, and $y_j^n = 1$ if $j/n > 1-p$. The n-th contest is an all-pay auction with n players and $\ulcorner pn \urcorner$ identical (non-zero) prizes, and the value of a prize to player i is i/n. Such contests were studied by CR, who considered competitions for promotions, rent seeking, and rationing by waiting in line.

Another example with complete information is contests with heterogeneous prizes and players who differ in their constant marginal valuation for a prize. For this, consider $h(y) = y$ and F and G uniform. Then $x_i^n = i/n$ and $y_j^n = j/n$. The n-th contest is an all-pay auction with n players and n heterogeneous prizes, and the value of prize j to player i is ij/n^2. Such contests were studied by BL, who considered hospitals that have a common ranking for residents and compete for them by posting wages.[17]

Many other asymmetric contests with complete information can be accommodated, including contests for which no equilibrium characterization exists. One such class of examples, for which no equilibrium characterization exists, is contests with a combination of heterogeneous and identical prizes.

At the other extreme we have symmetric contests with incomplete information, in which players have the same iid type distributions $F_i^n = F^n$ that converge to distribution F. This case includes the setting of MS. Beyond these extreme cases, our model also accommodates large contests with incomplete information and ex-ante asymmetric players.[18] No equilibrium

[17] Xiao (2016) presented another model with complete information and heterogenous prizes, in which players' marginal utility of prizes is increasing. He considered quadratic and exponential specifications, which are obtained in our model by setting $h(y) = y^2$ and $h(y) = e^y$, respectively, and F and G uniform.

[18] A reader interested in studying a specific n-player contest may also take $F = F^n = \left(\sum_{i=1}^n F_i^n\right)/n$. This of course requires that the sample of players in the contest approximately represent the entire population,

characterization exists for such settings.

3.2 Assortative allocation and transfers

As will be stated in the next subsection, the mechanism that approximates the equilibrium outcomes of large contests implements the *assortative allocation*, which assigns to each type x prize $y^A(x) = G^{-1}(F(x))$. That is, the quantile in the prize distribution of the prize assigned to type x is the same as the quantile of type x in the type distribution. It is well known (see, for example, Myerson (1981)) that the unique incentive-compatible mechanism that implements the assortative allocation and gives type $x = 0$ a utility of 0 specifies for every type x performance

$$t^A(x) = c^{-1}\left(xh\left(y^A(x)\right) - \int_0^x h\left(y^A(\widetilde{x})\right) d\widetilde{x}\right). \tag{2}$$

For example, in the setting corresponding to CR the assortative allocation assigns prize 0 to types $x \leq 1-p$ and assigns one of the identical positive prizes to each type $x > 1-p$. The associated performance is $t^A(x) = 0$ for $x \leq 1-p$ and $t^A(x) = 1-p$ for $x > 1-p$. In the setting corresponding to BL, the assortative allocation assigns prize x to type x, and the associated performance is $t^A(x) = x^2/2$.

3.3 The approximation result

Corollary 2 in OS, which we state as Theorem 1 below, shows that the equilibrium outcomes of large contests are approximated by the unique mechanism that implements the assortative allocation and gives type $x = 0$ a utility of 0.[19]

Theorem 1 *(OS) For any $\varepsilon > 0$ there is an N such that for all $n \geq N$, in any equilibrium of the n-th contest each of a fraction of at least $1-\varepsilon$ of the players i obtains with probability at least $1-\varepsilon$ a prize that differs by at most ε from $y^A(x_i^n)$, and chooses performance that is with probability at least $1-\varepsilon$ within ε of $t^A(x_i^n)$.*

which implies that adding additional players would roughly replicate the sample.

[19] OS also provide a result on the rates of convergence, which roughly says that it suffices for $1/N$ to be smaller than an expression of order $-\varepsilon^2/\ln\varepsilon$. We refer the reader to their Section 6 for the precise statement of the result. We will not discuss the rates of convergence in the present paper.

4 Performance maximization

4.1 Contest design problem

We will be interested in maximizing players' expected aggregate performance. This classical contest design goal is natural in some settings, as discussed in the introduction and in Section 8. In other settings the designer may value the performance of some types more than that of other types, or value a player's performance based on other statistics of performance. Such settings can be captured by slight variants of our analysis, as discussed in Section 10.

More precisely, when $t_i^n(x)$ is the performance of player i of type x, the expected average performance is

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^1 t_i^n(x)\,dF_i^n(x)\,.$$

We will maximize this quantity across the equilibria of contests in which the average budget, that is, the budget per capita, is Y:

$$\frac{1}{n}\sum_{j=1}^{n} y_j^n \leq Y.$$

The reason that we work with averages is to avoid the quantities becoming infinitely large as n tends to infinity.

For most of the analysis we will treat the budget constraint as exogenous and optimize over all possible prize structures. This fits settings in which the contest designer does not determine the budget but can determine how it is allocated across prizes.[20] When the designer also determines the budget, our results on the optimal prize structure for exogenous budgets can be used as an intermediate step to determine the optimal budget. This is illustrated in Section 7.2. Section 8 illustrates how our approach can be applied when the designer is restricted in the prize structures he can use, and the extra performance that is generated when such restrictions are removed.

[20] Departments in many organizations do not determine their own budget but control how the budget is used.

4.2 The design problem in the limit setting

Our first result shows that in order to maximize the expected average performance in large contests it is enough to solve the corresponding design problem in the limit setting. To obtain the result, we first observe that given a converging sequence of contests, Theorem 1 implies that the expected average performance for large n is approximated by the average performance in the mechanism that implements the assortative allocation in the limit setting:

$$\int_0^1 t^A(x) f(x)\, dx. \tag{3}$$

Corollary 1 *For any $\varepsilon > 0$ there is an N such that for all $n \geq N$, in any equilibrium of the n-th contest the expected average performance is within ε of (3).*

Corollary 1 applies to a given limit distribution G. Our aim is to characterize the limit distribution (and show that it exists) when the prizes in the sequence are the ones that maximize the expected average performance of the contests in the sequence. We do this by showing that the limit of the sequence of maximizing prizes coincides with the distribution of prizes that maximizes (3).

To show this, consider a sequence of type distributions that converges to distribution F with a continuous, strictly positive density f, and denote by $G^n_{\max}$ the empirical distribution of prizes that maximizes the equilibrium expected average performance in the n-th contest over all equilibria and all sets of prizes $y^n_1 \leq \cdots \leq y^n_n$ whose average is no greater than Y.[21] We denote by $M^n_{\max}$ the maximal expected average performance attained by $G^n_{\max}$. For the limit setting, we denote by $\mathcal{M}$ the set of prize distributions that maximize (3) subject to the budget constraint $\int_0^m y dG(y) \leq Y$. An upper hemi-continuity argument, given in the appendix, shows that a maximizing distribution exists.

Claim 1 *The set $\mathcal{M}$ is not empty.*

[21] That a maximizing set of prizes exists can be shown by a straightforward upper hemi-continuity argument of the kind used, for example, to prove Corollary 2 in Siegel (2009). We note, however, that our results do not depend on the existence of such a maximizing set of prizes. For example, none of the analysis changes if $G^n_{\max}$ is instead chosen to correspond to a set of n prizes that lead to some equilibrium with an expected average performance that is within $1/n$ of the supremum of the expected average performance over all sets of n prizes that meet the budget constraint and all equilibria for any given set of prizes.

Denote by M the corresponding maximal value of (3) subject to the budget constraint. Finally, consider any metrization of the weak*-topology on the space of prize distributions.

Proposition 1 *1. For any $\varepsilon > 0$, there is an N such that for every $n \geq N$, $G^n_{\max}$ is within ε (in the metrization) of some distribution in $\mathcal{M}$. In particular, if there is a unique prize distribution $G_{\max}$ that maximizes (3) subject to the budget constraint, then $G^n_{\max}$ converges to $G_{\max}$ in weak*-topology. 2. $M^n_{\max}$ converges to M. 3. For any $\varepsilon > 0$, there are an N and a $\delta > 0$ such that for any $n \geq N$ and any empirical prize distribution G^n of n prizes that is within δ of some G in M, the expected average performance in any equilibrium of the n-th contest with empirical prize distribution G^n is within ε of $M^n_{\max}$.*

Part 1 of Proposition 1 shows that the optimal prize distributions in large contests are approximated by the prize distributions that maximize (3) subject to the budget constraint. Part 2 shows that the maximal expected average performance is approximated by the maximal value of (3) subject to the budget constraint. Part 3 shows that any prize distribution that is close to a prize distribution that maximizes (3) subject to the budget constraint generates an expected average performance (in any equilibrium) that is close to maximal. For example, given a prize distribution G that maximizes (3) subject to the budget constraint, the set of n prizes defined by $y^n_j = G^{-1}(j/n)$ for $j = 1, ..., n$ generates, for large n, an expected average performance that is close to maximal; moreover, the average prize Y^n for the so defined distributions G^n converges to the average prize Y for the distribution G.[22]

By Proposition 1, we can focus on solving the following problem:

$$\begin{array}{l} \max_G \int_0^1 t^A(x) f(x)\, dx \\ \text{s.t. } \int_0^m y dG(y) \leq Y. \end{array} \tag{4}$$

We will now transform the problem (4) to obtain a more manageable form. We first transform the budget constraint to an equivalent constraint as a function of G^{-1}. Since G is a probability distribution on $[0, m]$, we have $\int_0^m y dG(y) = m - \int_0^m G(y)\, dy$ (by integrating by parts); we also have $\int_0^m G(y)\, dy + \int_0^1 G^{-1}(z)\, dz = m$ (by looking at the areas below

[22]It is easy to see that for any distribution G, such distributions G^n that converge to G can also be chosen so that Y^n does not exceed Y.

the graphs of G and G^{-1} in the square $[0,m] \times [0,1]$). Thus, the budget constraint can be rewritten as

$$\int_0^1 G^{-1}(z)\,dz \leq Y. \tag{5}$$

The interpretation of $G^{-1}(z)$, for each quantile $z \in [0,1]$ in the type distribution, is the prize allocated to type x in quantile $z = F(x)$. This is because the allocation is assortative.

We similarly transform the objective function. By substituting (2) into (3) and denoting by $L(x) = h\left(y^A(x)\right)$ the valuation of the prize allocated to type x in the assortative allocation, we obtain the following expression for the average performance in the mechanism that implements the assortative allocation:

$$\int_0^1 \left(c^{-1}\left(xL(x) - \int_0^x L(\widetilde{x})\,d\widetilde{x}\right)\right) f(x)\,dx. \tag{6}$$

In Appendix 11 we show that maximizing (6) subject to (5) can be written as an optimal control problem in variable G^{-1}.[23] When cost function c is linear, the problem reduces to a calculus of variations problem, and in this case the analysis becomes much simpler. We begin with the analysis of this simpler case.

5 Linear cost functions

The case of linear costs is particularly instructive, both because it turns out to be easier to solve than the general case, and because the first-order conditions that characterize the optimal inverse prize distribution G^{-1} and the conditions that guarantee its monotonicity are relatively easy to interpret.

With linear costs $c(t) = t$ the expression (6) for the average performance in the mechanism that implements the assortative allocation becomes

$$\int_0^1 \left(xL(x) - \int_0^x L(\widetilde{x})\,d\widetilde{x}\right) f(x)\,dx,$$

[23] We cannot easily apply standard tools from optimal control theory to solve the problem, because G^{-1} is required to be the inverse of a cdf G. Instead, we characterize the solution by using some ideas from optimal control theory in a way that may also provide some intuition for the solution.

and integrating by parts we obtain

$$\int_0^1 L(x)\left(x - \frac{1-F(x)}{f(x)}\right) f(x)\,dx. \tag{7}$$

Observe that (7) coincides with the expected revenue from a bidder in a single-object independent private-value auction if we let $L(x)$ be the probability that the bidder wins the object when his type is x (Myerson (1981)). This provides some intuition for why (7) approximates the expected average performance in large contests. In the auction setting, increasing the probability that type x obtains the object along with the price the type is charged allows the auctioneer to capture the entire increase in surplus for this type, but requires a decrease in the price that higher types are charged to maintain incentive compatibility. This net increase in revenue, or "virtual value," also coincides with a monopolist's marginal revenue (Bulow and Roberts (1989)). In a large contest, increasing the prize that type x obtains also allows the designer to capture the entire surplus increase for this type, because the higher prize increases this type's competition with slightly lower types until the surplus increase from the higher prize is exhausted. But the prize increase also decreases the competition of higher types for their prizes, since the prize of type x becomes more attractive to them.

We rewrite (7) by noting that $L(F^{-1}(z)) = h(G^{-1}(z))$ is the value of the prize in quantile z and substituting $z = F(x)$ to obtain

$$\int_0^1 h\left(G^{-1}(z)\right)\left(F^{-1}(z) - \frac{1-z}{f(F^{-1}(z))}\right) dz = \int_0^1 h\left(G^{-1}(z)\right) k(z)\,dz, \tag{8}$$

where

$$k(z) = F^{-1}(z) - (1-z)/f\left(F^{-1}(z)\right).$$

Maximizing (8) subject to (5) is a calculus of variations problem in variable G^{-1}. The value $k(z)$ can be interpreted as the marginal "virtual performance." That is, the additional average performance that can be induced by a marginal increase in the prize valuation h resulting from an increase in the prize assigned to the type in quantile z. This additional performance is the combination of the increase in performance of the type in quantile z and the decrease in performance by all higher types.

Before presenting a rigorous analysis of the calculus of variations problem, we provide some intuition for our approach. Consider first the relaxed problem of maximizing (8) subject to (5) without imposing a monotonicity constraint on G^{-1} (implied by G^{-1} being an inverse *cdf*). We derive two conditions. The first is a condition on distribution F that guarantees that the maximizer of the relaxed problem is nondecreasing, and therefore solves the original problem. The second is a version of the Euler-Lagrange equation, which is a necessary condition for a solution of the relaxed problem. To derive both conditions, notice that a slight increase of the prize $G^{-1}(z)$ allocated to the type in quantile z increases (8) by $h'(G^{-1}(z))\, k(z)\, dz$.[24] Thus, a sufficient condition for monotonicity of the maximizer of the relaxed problem is that $k(z)$ strictly increases in z. This is because whenever $z < z'$ (so $k(z) < k(z')$) but $G^{-1}(z) > G^{-1}(z')$, these values of G^{-1} can be exchanged, which increases the sum $h(G^{-1}(z))\, k(z) + h(G^{-1}(z'))\, k(z')$ without violating the budget constraint. Strict monotonicity of $k(z)$ is equivalent to the following assumption, which we maintain in the main text:[25]

Assumption 1. $P(x) = x - (1 - F(x)) / f(x)$ strictly increases in $x \in [0, 1]$.

This assumption is standard in the mechanism design literature, and corresponds to Myerson's (1981) "regular case."[26] The assumption is implied, for example, by a monotone hazard rate of distribution F. When players are ex-ante symmetric, Assumption 1 is satisfied when it holds for the distributions $F^n = F_i^n$ and the densities f^n pointwise converge to the limiting density f. In complete-information settings, an assumption on the limit distribution is particularly natural, because we often first choose the desired limit distribution F and then set player i's deterministic type in the n-th contest to be $x_i^n = F^{-1}(i/n)$ (such settings are studied in BL and CR).

We now heuristically derive conditions on the maximizer of the relaxed problem that rule out small improvements.

[24] This is just a heuristic illustration. The proof of Lemma 1 does not involve the relaxed problem, but the intuition using the relaxed problem is clearer.

[25] In the online appendix we show that many of the results for linear costs also hold without Assumption 1.

[26] In Myerson's setting the assumption guarantees that choosing $L(x)$ to maximize the integrand of (7) type by type leads to a non-decreasing function L. The argument for monotonicity in our case is slightly different, because of the budget constraint.

5.1 Conditions describing the solution

Consider a maximizer G^{-1} of the relaxed problem. It cannot be that there are $z \neq z'$ with $G^{-1}(z), G^{-1}(z') \in (0, m)$ and $h'(G^{-1}(z))\,k(z) < h'(G^{-1}(z'))\,k(z')$, because slightly decreasing $G^{-1}(z)$ and increasing $G^{-1}(z')$ by the same amount increases the sum $h(G^{-1}(z))\,k(z) + h(G^{-1}(z'))\,k(z')$ without violating the budget constraint. Similarly, if $G^{-1}(z) = 0$ and $G^{-1}(z') > 0$, then $h'(G^{-1}(z))\,k(z) \leq h'(G^{-1}(z'))\,k(z')$, and if $G^{-1}(z) = 1$ and $G^{-1}(z') < 1$, then $h'(G^{-1}(z))\,k(z) \geq h'(G^{-1}(z'))\,k(z')$.

Now, for any prize distribution G (optimal or not), in the assortative allocation there are quantiles $z_{\min} \leq z_{\max}$ in $[0,1]$ such that types in quantiles $z \leq z_{\min}$ in the limit distribution F are each allocated the prize $G^{-1}(z) = 0$ (no prize),[27] types in quantiles $z > z_{\max}$ are each allocated the highest possible prize $G^{-1}(z) = m$, and types in intermediate quantiles $z_{\min} < z < z_{\max}$ are allocated positive, non-maximal prizes $G^{-1}(z) \in (0, m)$. Since under Assumption 1 the maximizer of the relaxed problem solves the original problem, we obtain the following result:[28]

Lemma 1 *Given a prize distribution G, let $z_{\min} \leq z_{\max}$ in $[0,1]$ be such that $G^{-1}(z) = 0$ for $z \leq z_{\min}$, $G^{-1}(z) = m$ for $z > z_{\max}$, and $G^{-1}(z) \in (0, m)$ for $z \in (z_{\min}, z_{\max})$. If G is an optimal prize distribution, then it satisfies the following conditions:*

If $z_{\min} < z_{\max}$ (Case 1): Then, there exists a $\lambda \geq 0$ such that $h'(G^{-1}(z))\,k(z) = \lambda$ for $z \in (z_{\min}, z_{\max}]$; in addition, $h'(0)\,k(z_{\min}) \leq \lambda$, and $h'(m)k(z_{\max}) \geq \lambda$ if $z_{\max} < 1$.

If $z_{\min} = z_{\max}$ (Case 2): Then, $h'(0)\,k(z_{\min}) \leq h'(m)\,k(z_{\max})$.

In the special case of $h'(0) = \infty$, it is understood that $k(z_{\min}) = 0$ and $h'(0)\,k(z_{\min}) = 0$.

The parameter λ in Case 1 is the shadow price of the budget constraint, that is, by how much the average performance increases if the budget is increased slightly. This shadow price can be used to determine the optimal budget when the budget is endogenous, as discussed in Section 7.2.

[27] The inequality $z \leq z_{\min}$ is weak because G^{-1} is left-continuous as the inverse of a probability distribution.

[28] The proof of Lemma 1 shows that under Assumption 1 any inverse *cdf* G^{-1} that satisfies the budget constraint but fails the conditions in Lemma 1 can be improved upon by another inverse *cdf* that satisfies the budget constraint.

5.2 Risk averse, risk neutral, and risk loving players

We now use Lemma 1 to characterize the optimal prize distribution for risk averse, risk natural, and risk loving players, who have concave, linear, and convex prize valuation functions h, respectively. We first identify the maximal amount that will ever be allocated to prizes, and show that when the budget exceeds this amount players' risk attitude does not affect the optimal prize distribution. We then show how players' risk attitude affects the optimal prize distribution when the budget is smaller than this amount.

Denote by $x^* \in (0,1)$ the unique type that satisfies $P(x^*) = 0$, and by $z^* = F(x^*) \in (0,1)$ the quantile of type x^* in the type distribution, so $k(z^*) = 0$.[29] That is, type x^* is the type for whom the marginal virtual performance is 0. Types $x < x^*$ have negative marginal virtual performance, so the value of the integrand in (8) for them is negative, and types $x > x^*$ have positive marginal virtual performance, so the value for them is positive. Since the marginal prize utility is positive regardless of the curvature of h, optimizing the integrand in (8) separately for each $z \in [0,1]$ leads to assigning the lowest possible prize $G^{-1}(z) = 0$ to types in quantiles $z \leq z^*$, and assigning the highest possible prize m to types in quantiles $z > z^*$.[30] This G^{-1} is left-continuous and nondecreasing, so the corresponding G is a prize distribution. Its cost is $m(1 - F(x^*))$, so this distribution is the optimal one when the budget Y is at least $m(1 - F(x^*))$. We thus obtain the following result.

Proposition 2 *If $Y \geq m(1 - F(x^*))$, then for any function h the optimal prize distribution consists of a mass $1 - F(x^*) \in (0,1)$ of the highest possible prize, m, and a mass $F(x^*)$ of prize 0.*

Proposition 2 shows that with a sufficiently large budget it is optimal to award a set of identical prizes, as in the all-pay auctions studied by CR, rather than heterogeneous prizes, as in, for example, the all-pay auctions studied by BL, or a combination of identical and heterogeneous prizes. Notice that the optimal mass of prizes, $1 - F(x^*)$, is independent of the size of the highest possible prize, m. Another implication is that if increasing the budget

[29] Such a type exists and is unique because by Assumption 1, $P(x)$ strictly increases in x, and since f is continuous and strictly positive on $[0,1]$, P is also continuous on $[0,1]$.

[30] This corresponds to Case 2 of Lemma 1, with $z_{\min} = z_{\max} = z^*$.

is costly (see Section 7.2), then the budget will optimally not exceed $m(1-F(x^*))$, since awarding additional prizes would reduce the average performance. This is analogous to a monopolist limiting the quantity sold.

When the budget is lower than $m(1-F(x^*))$, players' risk attitude affects the optimal prize distribution. We first present the simpler result for convex functions h.

Proposition 3 *If $Y < m(1-F(x^*))$ and h is weakly convex, so players are risk neutral or risk loving, then the optimal prize distribution consists of a mass Y/m of the highest possible prize, m, and a mass $1-Y/m$ of prize 0.*

Proof: Weak convexity implies that $z_{\min} = z_{\max}$, so only prizes 0 and m are awarded. Otherwise, since h' and G^{-1} are weakly increasing and k is strictly increasing, for any $z' < z''$ in $(z_{\min}, z_{\max})$ we would have $h'(G^{-1}(z'))\,k(z') < h'(G^{-1}(z''))\,k(z'')$, which would violate the condition $h'(G^{-1}(z'))\,k(z') = h'(G^{-1}(z''))\,k(z'') = \lambda$ in Case 1 of Lemma 1.

Proposition 3 shows that awarding identical prizes remains optimal when the budget is low, provided that agents' marginal prize utility is nondecreasing. If the highest possible prize is increased, fewer maximal prizes are optimally awarded. The limit as m grows arbitrarily large corresponds to a single grand prize.

Propositions 2 and 3 fully characterize the optimal prize distribution when the budget is large (regardless of the curvature of h) and when the marginal prize utility is increasing. In these cases, the optimal prize distribution does not depend on the precise functional form of h. With a small budget and decreasing marginal prize utility, however, the optimal prize distribution depends more heavily on h. We first provide a qualitative characterization of the optimal prize distribution in this case, and then a full characterization for strictly concave functions h.

Proposition 4 *1. If $Y < m(1-F(x^*))$ and h is weakly concave (but not linear on $[0,m]$), so players are weakly risk averse (but not risk neutral), then any optimal prize distribution assigns a positive mass to the set of intermediate prizes $(0,m)$. In addition, any optimal prize distribution may have atoms only at prize 0 and prize m. 2. If h is strictly concave, then any optimal prize distribution awards all prizes up to the highest prize awarded. That is, the optimal G strictly increases on $[0, G^{-1}(1)]$.*

Proof: Observe that $z_{\min} < z_{\max}$. Indeed, since $h'(0) > h'(m)$, we cannot have that $z_{\min} = z_{\max}$ and $h'(0)\,k(z_{\min}) \leq h'(m)\,k(z_{\max})$, unless $k(z_{\min}) = k(z_{\max}) \leq 0$. But $k(z_{\max}) \leq 0$ implies that $z_{\max} \leq z^*$. Since $G^{-1}(z) = m$ for $z > z_{\max}$, we obtain that $\int_0^1 G^{-1}(z)\,dz \geq m(1 - F(x^*)) > Y$ violates the budget constraint (5). This yields the first part of 1. For the second part, notice that $G^{-1}(z)$ strictly increases in z on interval $(z_{\min}, z_{\max})$, so G does not have atoms there. This follows from the fact that $h'(G^{-1}(z))\,k(z) = \lambda$ on $(z_{\min}, z_{\max}]$ and the fact that $k(z)$ strictly increases in z.

To see 2, note that h' is strictly decreasing and, by assumption, continuous. Thus, $h'(G^{-1}(z))\,k(z) = \lambda$ also implies that G^{-1} is continuous on $(z_{\min}, z_{\max}]$. If G^{-1} were not right-continuous at $z_{\min}$, then the fact that $h'(0)\,k(z_{\min}) \leq \lambda$ and the assumption that h' is strictly decreasing would violate the condition $h'(G^{-1}(z))\,k(z) = \lambda$ for z slightly higher than $z_{\min}$.[31] Thus, G^{-1} is continuous on $[z_{\min}, z_{\max}]$, which means that G strictly increases on $[0, G^{-1}(z_{\max})]$. If $G^{-1}(z_{\max}) = G^{-1}(1)$, this completes the proof. If $G^{-1}(z_{\max}) < G^{-1}(1)$, which can happen when $z_{\max} < 1$, then $G^{-1}(z_{\max}) < G^{-1}(1) = m$, which also completes the proof, as otherwise the fact that $h'(m)k(z_{\max}) \geq \lambda$ and the assumption that h' is strictly decreasing would violate the condition $h'(G^{-1}(z_{\max}))\,k(z_{\max}) = \lambda$.

Proposition 4 shows that decreasing marginal prize utility optimally leads to awarding intermediate prizes, whose values gradually decrease with players' performance ranking. Among the (positive) prizes, only the highest possible prize, m, may optimally be awarded to multiple players. This generally does not occur when m is sufficiently large, however, as the the following result shows.

Proposition 5 *Suppose that $Y < m(1 - F(x^*))$. Let $G^m_{\max}$ be an optimal prize distribution when m is the highest possible prize. If h is weakly concave (but not linear on $[0, m]$), and $h'(y) \to 0$ as $y \to \infty$, then there exists an m such that $G^{m'}_{\max} = G^m_{\max}$ for any $m' \geq m$, and this $G^m_{\max}$ may have an atom only at prize 0.*

[31] More precisely, the argument delivering the right-continuity at $z_{\min}$ applies only to cases in which $h'(0) < \infty$. The case when $h'(0) = \infty$ requires a somewhat special treatment.

If $h'(0) = \infty$, then $k(z_{\min}) = 0$, so if G^{-1} were not right-continuous at $z_{\min}$ the product $h'\left(G^{-1}(z)\right)k(z)$ would be strictly positive for any $z \in (z_{\min}, z_{\max}]$, but would approach 0 as $z \downarrow z_{\min}$, so could not be constant on $(z_{\min}, z_{\max}]$.

It may be tempting to attribute the qualitative difference between the optimal prize distributions with convex and concave prize valuations to the difference in players' risk attitudes as follows: lotteries between no prize and the highest possible prize are riskier than ones over a range of intermediate prizes, so the former can elicit more performance when players are risk loving, and the latter when they are risk averse. This intuition is misleading, however, because in large contests almost all player types are nearly certain of the prize they receive in equilibrium (Theorem 1).[32] Instead, what drives the qualitative difference is how the marginal prize utility changes as the prize increases. Because the marginal prize utility is always positive, absent the budget constraint it is optimal to award the lowest possible prize to types with negative marginal virtual performance, and the highest possible prize to types with positive marginal virtual performance. The budget constraint introduces a tradeoff between the prizes allocated to different types. This tradeoff is optimally resolved by comparing the product of the marginal prize utility and the marginal virtual performance across types. Since the marginal virtual performance increases in type, what determines the comparison is whether the marginal prize utility increases or decreases in the prize, which correspond to convex and concave prize valuations. In the former case, increasing the prize increases the product, so it is optimal to allocate the highest possible prize to the highest types. In the latter case, increasing the prize decreases the product, so continuity of the marginal virtual performance implies that as we increase the prizes awarded to some types, it becomes increasingly attractive to award prizes to slightly lower types. The optimal prize distribution equates the product across all types allocated intermediate prizes. Such types exist, because $Y < m\left(1 - F\left(x^*\right)\right)$ implies that not all types can be awarded the highest possible prize.

We now provide a full characterization of the optimal prize distribution when $Y < m\left(1 - F\left(x^*\right)\right)$ and h is strictly concave. Since the optimal G is strictly increasing (part 2 of Proposition 4), G^{-1} is continuous, so we have $h'\left(0\right) k\left(z_{\min}\right) = \lambda$. Thus,

$$z_{\min} = k^{-1}(\lambda / h'\left(0\right)). \tag{9}$$

[32] In addition, if the intuition were correct, we would expect the optimal prize distribution to vary with players' risk attitudes also when the budget exceeds $m\left(1 - F\left(x^*\right)\right)$, in contrast to the statement of Proposition 2.

Since $h'\left(G^{-1}\left(z_{\max}\right)\right) k\left(z_{\max}\right)=\lambda$ and h' is decreasing, $h'(m) k\left(z_{\max}\right) \leq \lambda$. If $z_{\max}<1$, then we also have $h'(m) k\left(z_{\max}\right) \geq \lambda$ (because we are in Case 1 of Lemma 1), so we obtain $h'(m) k\left(z_{\max}\right)=\lambda$. Thus,

$$z_{\max}=1 \text{ or } k^{-1}(\lambda / h'(m)). \tag{10}$$

In addition,

$$G^{-1}(z)=\left(h'\right)^{-1}(\lambda / k(z)) \text{ for } z \in\left(z_{\min}, z_{\max}\right] \tag{11}$$

and

$$G^{-1}(z)=\begin{cases}0 & z \leq z_{\min} \\ m & z>z_{\max}\end{cases}.$$

Thus, G^{-1} is pinned down by λ. The value of λ is determined by the fact that (5) holds as an equality (because $Y<m\left(1-F\left(x^*\right)\right)$).

To demonstrate the usefulness of this characterization, we now derive the optimal G^{-1} for contests with prize valuations $h(y)=y^{1/j}$ for $j>1$ (and any type distribution F). This will be useful in Section 7. We assume that the maximal prize m is large enough that $z_{\max}=1$ (see Proposition 5), which also implies that $Y<m\left(1-F\left(x^*\right)\right)$, so the entire budget is used. Since $h'(0)=\infty$, we have $z_{\min}=z^*$. Since $\left(h'\right)^{-1}(r)=(jr)^{j/(1-j)}$, by (11) we have

$$G^{-1}(z)=\left(h'\right)^{-1}(\lambda / k(z))=\frac{1}{\lambda^{j/(j-1)} j^{j/(j-1)}} k(z)^{j/(j-1)} \tag{12}$$

for $z \in\left(z^*, 1\right]$. Thus,

$$Y=\int_{z^*}^1 G^{-1}(z)\, d z=\frac{1}{\lambda^{j/(j-1)} j^{j/(j-1)}} \int_{z^*}^1 k(z)^{j/(j-1)} d z,$$

so

$$\lambda^{j/(j-1)}=\frac{1}{Y j^{j/(j-1)}} \int_{z^*}^1 k(\widetilde{z})^{j/(j-1)} d \widetilde{z}.$$

Substituting this expression for $\lambda^{j/(j-1)}$ into (12) we obtain

$$G^{-1}(z)=Y \frac{k(z)^{j/(j-1)}}{\int_{z^*}^1 k(\widetilde{z})^{j/(j-1)} d \widetilde{z}} \text{ for } z \in\left(z^*, 1\right] \text{ and } G^{-1}(z)=0 \text{ for } z \leq z^*. \tag{13}$$

6 More general cost functions

In this section, we develop the conditions satisfied by an optimal prize distribution for a more general class of cost functions. For this it is useful to substitute $x = F^{-1}(z)$ into (6) to express the average performance as

$$\int_0^1 c^{-1}\left(\widetilde{L}(z)\right) dz, \tag{14}$$

where

$$\widetilde{L}(z) = F^{-1}(z)\, L\left(F^{-1}(z)\right) - \int_0^{F^{-1}(z)} L(\widetilde{x})\, d\widetilde{x} \tag{15}$$

is the cost of the performance of type x in quantile $z = F(x)$. Notice that $\widetilde{L}(z)$ is well defined even when function $G^{-1}(z)$ is not monotone. We will consider such functions in some of our proofs.

Similarly to Section 5, to derive the conditions for optimality it is useful to consider the effect of a slight increase Δ in the value of G^{-1} at quantile z on the average performance (14). In the case of linear cost c, the effect was to generate an increase of $h'\left(G^{-1}(z)\right) k(z)\,\Delta$, where $k(z)$ was the marginal virtual performance. With non-linear costs, the marginal virtual performance $K(z)$ (given by (17) below) in the corresponding expression for the increase will involve G^{-1} and the derivative of c^{-1}. The expression will be instrumental in formulating conditions that characterize the optimal G^{-1} and generalize the conditions in Lemma 1. But because the expression involves G^{-1}, it cannot be used directly in formulating Assumption 2 below, which guarantees that any optimizer of the relaxed problem is nondecreasing and generalizes Assumption 1, because such an assumption must refer only to the primitives of the model, that is, only to functions h, c, and F.

To estimate the effect of a slight increase in $G^{-1}(z)$ on (14), consider a function G^{-1} that takes values only in the set $\{0, 1/2^n, 2/2^n, ..., (2^n-1)/2^n, 1\}$, and is constant on each interval $(0, 1/2^n], (1/2^n, 2/2^n], ..., ((2^n-1)/2^n, 1]$. Suppose that we increase the value of G^{-1} on an interval $(l/2^n, (l+1)/2^n]$ by $\Delta = 1/2^n$. Since $L\left(F^{-1}(z)\right) = h\left(G^{-1}(z)\right)$, this change increases the value of L on $(F^{-1}(l/2^n), F^{-1}((l+1)/2^n)]$ by $h'(G^{-1}((l+1)/2^n))\Delta$, to a first-order approximation. In Figure 1 this corresponds to shifting the graph of L on $(F^{-1}(l/2^n), F^{-1}((l+1)/2^n)]$ to the right by the width of the shaded square. This change does not affect $\tilde{L}$, and thus the integrand in (14), on intervals $(k/2^n, (k+1)/2^n]$ for $k < l$. It increases the integrand for $z \in (l/2^n, (l+1)/2^n]$, to a first-order approximation, by

$$(c^{-1})'(\widetilde{L}((l+1)/2^n))F^{-1}((l+1)/2^n)h'((G)^{-1}((l+1)/2^n))\Delta$$

(the union of the shaded and darkened rectangles in Figure 1). For any $k > l$, it decreases the integrand for $z \in (k/2^n, (k+1)/2^n]$, to a second-order approximation, by

$$(c^{-1})'(\widetilde{L}((k+1)/2^n))h'((G)^{-1}((l+1)/2^n))\Delta[F^{-1}((l+1)/2^n) - F^{-1}(l/2^n)]$$

(the shaded square in Figure 1).

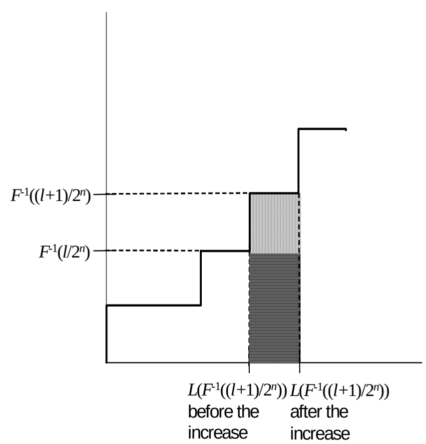

Figure 1: Increasing G^{-1}

Since $F^{-1}((l+1)/2^n) - F^{-1}(l/2^n) = \Delta/f(F^{-1}((l+1)/2^n))$, to a first-order approximation, letting $z = (l+1)/2^n$, we express the total increase in (14) as

$$h'\left(G^{-1}(z)\right)\left(F^{-1}(z)\left(c^{-1}\right)'\left(\widetilde{L}(z)\right) - \frac{\int_z^1 \left(c^{-1}\right)'\left(\widetilde{L}(\widetilde{z})\right)d\widetilde{z}}{f\left(F^{-1}(z)\right)}\right)\Delta^2. \tag{16}$$

Recalling that $\widetilde{L}(z)$ is the cost of the performance of type x in quantile $z = F(x)$, we can interpret

$$K(z) = \left(F^{-1}(z)\left(c^{-1}\right)'\left(\widetilde{L}(z)\right) - \frac{\int_z^1 \left(c^{-1}\right)'\left(\widetilde{L}(\widetilde{z})\right)d\widetilde{z}}{f\left(F^{-1}(z)\right)}\right) \tag{17}$$

as the marginal virtual performance of the type in quantile z: a marginal increase in the prize allocated to this type intensifies competition for this prize and exhausts the corresponding increase in allocation utility (the first term on the right-hand side of (17)), but reduces competition by all higher types (the second term on the right-hand side of (17)). With non-linear costs these effects depend on the prizes allocated to lower types (through $\widetilde{L}(z)$), because they determine the current performance, which affects the marginal cost of performance. This dependency disappears with linear cost $c(t) = t$, in which case $K(z)$ coincides with $k(z)$.

With linear costs, Assumption 1 guarantees that the maximizer G^{-1} of the relaxed problem is nondecreasing. Indeed, if $G^{-1}(z)$ were lower on an interval $(l/2^n, (l+1)/2^n]$ than on an interval $(j/2^n, (j+1)/2^n]$ for some $j < l$, we could exchange the two values, generating a higher increase on $(l/2^n, (l+1)/2^n]$ than a decrease on $(j/2^n, (j+1)/2^n]$. The assumption that $K(z)$ is strictly increasing would be a natural counterpart of Assumption 1 in the more general setting. Unfortunately, $K(z)$ involves the endogenous variable G^{-1} (through $\widetilde{L}(z)$), which would make the assumption unattractive; moreover, it would no longer serve its purpose, because exchanging the values of G^{-1} on $(l/2^n, (l+1)/2^n]$ and $(j/2^n, (j+1)/2^n]$ would affect the value of $K(z)$. We instead make the following assumption, expressed only in terms of the primitives of the model, which guarantees directly that exchanging the values of G^{-1} on the two intervals is beneficial, and therefore guarantees that the maximizer of the relaxed problem is nondecreasing. For the assumption, recall that no player chooses performance higher than $c^{-1}(h(m))$, and let $\underline{c} = \min\{c'(t) : t \in [0, c^{-1}(h(m))]\}$ and $\overline{c} = \max\{c'(t) : t \in [0, c^{-1}(h(m))]\}$. We restrict attention to continuously differentiable density functions f.

Assumption 2. For all z in $[0,1]$,

$$\frac{2}{\overline{c} f(F^{-1}(z))} + \frac{f'(F^{-1}(z))(1-z)}{\underline{c} f^3(F^{-1}(z))} > 0, \tag{18}$$

where if $\underline{c} = 0$ the second fraction is equal to ∞, $-\infty$, or 0 when its numerator is positive, negative, or 0, respectively.

Assumption 2 generalizes Assumption 1, because when $\underline{c} = \overline{c} = 1$ the left-hand side of (18) is equal to $k'(z)$. Assumption 2 is satisfied, for example, whenever f is nondecreasing (for any cost function c).[33] Note also that Assumption 2 imposes no conditions on valuation function h.[34]

6.1 Conditions describing the solution

Equipped with (17), we obtain the following analogue of Lemma 1.

Lemma 2 *Given a prize distribution G, let $z_{\min} \leq z_{\max}$ in $[0,1]$ be such that $G^{-1}(z) = 0$ for $z \leq z_{\min}$, $G^{-1}(z) = m$ for $z > z_{\max}$, and $G^{-1}(z) \in (0, m)$ for $z \in (z_{\min}, z_{\max})$. If G is an optimal prize distribution, then it satisfies the following conditions:*

If $z_{\min} < z_{\max}$ (Case 1): Then, there exists a $\lambda \geq 0$ such that

$$h'\left(G^{-1}(z)\right) K(z) = \lambda \tag{19}$$

for $z \in (z_{\min}, z_{\max}]$; in addition,

$$h'(0) K(z_{\min}) \leq \lambda, \tag{20}$$

and

$$h'(m) K(z_{\max}) \geq \lambda \tag{21}$$

if $z_{\max} < 1$.

If $z_{\min} = z_{\max}$ (Case 2): Then,

$$h'(0) K(z_{\min}) \leq \lim_{z \downarrow z_{\max}} h'(m) K(z). \tag{22}$$

The difference between Case 2 in Lemma 2 and Case 2 in Lemma 1 arises because $k(z)$ is continuous at every z, whereas $K(z)$ is left-continuous at every z but changes discontinuously at quantiles z at which $G^{-1}(z)$ increases discontinuously. In particular, if $z_{\min} = z_{\max}$, then $\widetilde{L}(z_{\min}) = 0$ (type $z_{\min}$ obtains prize 0 and chooses performance 0) but $\widetilde{L}(z) = G(0)\, h(m)$

[33] In addition, the set of primitives for which Assumption 2 holds is generic in the sense that if it holds for some pair of a continuous derivative of a cost function and a continuous derivative of a density function, then it holds for all such pairs that are sufficiently close to it in the sup norm.

[34] We strongly conjecture that, similarly to the case of linear costs, many of the results for general costs also hold without Assumption 2.

for all $z > z_{\max}$ (types above $z_{\max}$ obtain prize m and choose the performance with cost $G(0)\, h(m)$, which makes type $z_{\min} = z_{\max} = G(0)$ indifferent between choosing this performance and obtaining prize m and choosing performance 0 and obtaining prize 0).

A more subtle difference from Lemma 1 relates to (20). The intuition for (20) is that if the inequality were reversed, then in the relaxed problem increasing $G^{-1}(z)$ for z slightly below $z_{\min}$ by decreasing $G^{-1}(z)$ for z in $(z_{\min}, z_{\max})$ would increase the average performance. This relies on $z_{\min} > 0$, which is always the case with linear costs (because $k(0) < 0$). More generally, however, it can be that $z_{\min} = 0$ (see part 3 of Proposition 6 below). But in this case (20) follows from (19) directly, because $G^{-1}(z)$, and therefore $K(z)$, are continuous at $z = 0$. Otherwise $y_{\min} = \lim_{z \downarrow 0} G^{-1}(z) > 0$, so G^{-1} could be "shifted down" to reduce $h(G^{-1}(z))$ by $h(y_{\min})$,[35] which would reduce the cost of providing the prizes without changing each type's performance. The prizes $G^{-1}(z)$ for z close to 1 could then be increased, which would increase the average performance.

6.2 Risk averse and risk neutral players with convex costs

We now use Lemma 2 to characterize the optimal prize distribution for risk averse and risk neutral players with convex costs. This will generalize Proposition 4 and highlight additional features of the optimal prize distribution implied by convex costs. As we will see, the effects of risk aversion and convex costs on the optimal prize distribution are qualitatively similar, but not identical. It is also possible to generalize the results from Section 5.2 for risk loving players to concave costs, but this case seems less relevant for economic applications. Lemma 2 can also be used to study the optimal prize distribution for risk loving players with convex costs, but no general results exist in this case, because the effects of convex costs and convex prize valuations go in opposite directions.[36]

Proposition 6 *Suppose that h is weakly concave. 1. If c is weakly convex but not linear on any interval with lower bound 0, then any optimal prize distribution assigns a positive mass*

[35] The new inverse prize distribution would assign prize $G^{-1}(z) - h^{-1}\left(h\left(G^{-1}(z)\right) - h(y_{\min})\right)$ to quantile z. This maintains the same value of (2) for every x.

[36] For details see footnote 31 and the sentence that precedes it.

to the set of intermediate prizes $(0,m)$*. In addition, any optimal prize distribution may have atoms only at* 0 *(no prize) and* m *(the highest possible prize). 2. If* c *is strictly convex, then any optimal prize distribution awards all prizes up to the highest prize awarded. That is,* G *strictly increases on* $[0, G^{-1}(1)]$*. 3. If the marginal cost of the first unit of performance is* 0*, that is,* $c'(0) = 0$*, then* $z_{\min} = 0$*, so almost every type is awarded a positive prize.*

Proof: The first part of 1 is true because $z_{\min} < z_{\max}$. Indeed, if $z_{\min} = z_{\max}$, then $\widetilde{L}(z) = G(0)\,h(m)$ for all $z > z_{\max}$ (as explained immediately after Lemma 2), and since $\widetilde{L}(z_{\min}) = 0$ and $(c^{-1})'(0) > (c^{-1})'(G(0)h(m))$, we obtain that $K(z_{\min}) > \lim_{z \downarrow z_{\max}} K(z)$. Together with $h'(0) \geq h'(m)$, (22) is violated.[37] For the second part of 1, an atom at some intermediate prize would mean that $G^{-1}(\underline{z}) = G^{-1}(\overline{z})$ for some $z_{\min} < \underline{z} < \overline{z} < z_{\max}$. We would then have $h'(G^{-1}(\underline{z})) = h'(G^{-1}(\overline{z}))$ and $\widetilde{L}(z)$ constant on $[\underline{z}, \overline{z}]$. The derivative of $K(z)$ on $[\underline{z}, \overline{z}]$ would then be

$$\frac{2}{f(F^{-1}(z))}\left(c^{-1}\right)'\left(\widetilde{L}(z)\right) + \frac{f'(F^{-1}(z))\int_z^1 (c^{-1})'\left(\widetilde{L}(\widetilde{z})\right)d\widetilde{z}}{f^3(F^{-1}(z))},$$

which is strictly positive if $f'(F^{-1}(z)) \geq 0$, and also if $f'(F^{-1}(z)) < 0$ (by Assumption 2). We could then not have (19) for both $z = \underline{z}$ and $z = \overline{z}$.

For 2, notice that $\widetilde{L}(z)$ increases discontinuously when $G^{-1}(z)$ increases discontinuously. So, if $(c^{-1})'$ is strictly decreasing, a discontinuity in $G^{-1}(z)$ would leads to a discontinuous decrease in the left-hand side of (19). Thus, G^{-1} is continuous on $(z_{\min}, z_{\max}]$. If G^{-1} were not right-continuous at $z_{\min}$, then (19) and (20) could not both be satisfied, because of the discontinuous decrease of $(c^{-1})'$ at $z_{\min}$ (and, if h is strictly concave, also a discontinuous decrease of h'). Thus, G strictly increases on $[0, G^{-1}(z_{\max})]$. If $z_{\max} < 1$, then $G^{-1}(z_{\max}) = G^{-1}(1) = m$. Indeed, if $G^{-1}(z_{\max}) < G^{-1}(1)$, then (19) and (21) could not both be satisfied, because of the discontinuous decrease in the left-hand side of (19) at $z_{\max}$.

For 3, suppose that $z_{\min} > 0$. If G^{-1} is discontinuous at $z_{\min}$, then (20) cannot hold. And if G^{-1} is continuous at $z_{\min} > 0$, then $K(z)$, and so the left-hand side of (19), diverge to ∞ for z that tends to $z_{\min}$ from the right, so (19) is violated for z's close to $z_{\min}$.

[37] If h were strictly convex, we would have $h'(0) < h'(m)$, so we could be in Case 1 or Case 2 of Lemma 2.

Proposition 6 highlights some similarities and differences between the effects on the optimal prize distribution of convex costs and concave prize valuations. When the budget is not large ($Y < m(1 - F(x^*))$), both strictly convex costs with linear prize valuations and linear costs with strictly concave prize valuations optimally lead to awarding intermediate prizes. The prizes' values gradually decrease with players' performance ranking, with only the highest possible prize (among the positive prizes) possibly being awarded to multiple players. But when the budget is large ($Y \geq m(1 - F(x^*))$), convex costs still lead to awarding intermediate prizes (since Proposition 6 holds for any budget), whereas concave prize valuations lead to awarding only the highest possible prize. This is because with convex costs a slight change in the prize a type is awarded induces a higher change in performance when the prize is 0 than when the prize is m, but affects higher types' performance in the same way, both when the prize is 0 and when the prize is m. Thus, it cannot be optimal for some type to be awarded prize 0 and a slightly higher type to be awarded prize m. In addition, if the marginal performance cost at 0 is 0, then almost every type is optimally awarded a positive prize. This is because a marginal cost of 0 implies that a slight increase from 0 in the prize awarded to a positive type leads to an increase in that type's performance that infinitely outweighs the decrease in the performance of higher types. In other words, it is optimal to have almost every type participate in the contest, unlike with linear costs. These two differences can be seen by comparing the optimal prize distributions in Figure 3 in Section 7.1 and Figure 4 in Section 8, which correspond to a contest with concave valuations and linear costs and a contest with linear valuations and convex costs with a marginal cost of 0 at 0. In addition, the derivation of the optimal prize distributions in Figure 4 illustrates how to use Lemma 2 and Proposition 6 to explicitly derive the optimal prize distribution with convex costs.

7 Comparative statics

In this section, we investigate how varying players' ability distribution and increasing the budget affect the optimal prize distribution and the maximal expected average performance. For example, an agency supporting basic research can affect the ability distribution in the pool of contestants by imposing participation eligibility criteria. We conduct the investigation in the limit setting, since Corollary 1 and Proposition 1 imply approximately the same effects for large contests. For the results on the optimal prize distribution, we restrict attention to linear costs and use our characterization of the unique optimal prize distribution for convex and concave prize valuations.

7.1 Optimal prize distribution

We restrict attention to linear costs $c(t) = t$, and first consider how varying the limit type distribution F affects the optimal prize distribution G. For this we denote by $\widetilde{F}$ a second limit type distribution with corresponding optimal prize distribution $\widetilde{G}$ (we will use ˜ for all the relevant variables under $\tilde{F}$). The case of a large budget ($Y \geq m(1 - F(x^*))$) is relatively simple, and less interesting, since by Proposition 2 the optimal prize distribution consists of a mass $1 - F(x^*)$ of the highest possible prize and a mass $F(x^*)$ of prize 0, where x^* satisfies $P(x^*) = 0$. Thus, the effect of a change in F is determined by its effect on $F(x^*)$. If the budget is not large ($Y < m(1 - F(x^*))$) and valuation function h is weakly convex, then Proposition 3 shows that the optimal prize distribution consists of a mass Y/m of the highest possible prize and a mass $1 - Y/m$ of prize 0. In particular, it is independent of the limit type distribution.

The remaining case of $Y < m(1 - F(x^*))$ and strictly concave h is less straightforward and more interesting. Unlike the case of a large budget, as long as $Y < m\left(1 - \widetilde{F}(\widetilde{x}^*)\right)$, the optimal prize distributions G and $\widetilde{G}$ cannot be compared in terms of first-order stochastic dominance (FOSD). Otherwise, the budget constraint (5) would be violated by one of these distributions. The following result provides a sufficient condition for $\widetilde{G}$ to second-order stochastically dominate (SOSD) G, that is, for $\widetilde{G}$ to be less dispersed than G. For the result, denote by $Q(z) = \widetilde{F}^{-1}(z)/F^{-1}(z)$ the ratio of the types in quantile $z > 0$ in the two distributions.

Proposition 7 *Suppose that h is strictly concave and $Q(z)$ is weakly convex and strictly decreasing in z. Then G^{-1} crosses $\widetilde{G}^{-1}$ once, from below, and therefore $\widetilde{G}$ SOSD G.*

The assumption in Proposition 7 that $Q(z)$ is decreasing implies that $\tilde{F}$ FOSD F. Thus, Proposition 7 roughly says that the optimal heterogeneity in prizes is lower when the population of contestants is sufficiently more able. The next result shows that under an additional condition every player type optimally obtains a lower prize when the population of contestants is sufficiently more able.

Proposition 8 *Suppose that h is strictly concave, $Q(z)$ is weakly convex and strictly decreasing in z, and $\widetilde{P}(x) \le P(x)$ for every type x. Then $\widetilde{y}^A(x) \le y^A(x)$ for every type x, that is, every type optimally obtains a lower prize under $\widetilde{F}$ than under F.*

The ranking $\widetilde{P} \le P$ holds, for example, when $\widetilde{F}$ dominates F in the hazard ratio sense, which is implied by domination in the the likelihood ratio sense (so $\widetilde{f}/f$ is a weakly increasing function). For a class of distributions that satisfy the conditions of Propositions 7 and 8, consider the family of *cdf*s x^α for $\alpha > 0$. For any $\tilde{\alpha} > \alpha > 0$, let $\widetilde{F}(x) = x^{\tilde{\alpha}}$ and $F(x) = x^\alpha$. We have

$$Q(z) = \frac{\widetilde{F}^{-1}(z)}{F^{-1}(z)} = \frac{z^{\frac{1}{\tilde{\alpha}}}}{z^{\frac{1}{\alpha}}} = z^{\frac{\alpha-\tilde{\alpha}}{\alpha\tilde{\alpha}}},$$

which is strictly decreasing and convex, since $\alpha - \tilde{\alpha} < 0$. In addition,

$$\frac{\widetilde{f}(x)}{f(x)} = \frac{\tilde{\alpha}x^{\tilde{\alpha}-1}}{\alpha x^{\alpha-1}} = \frac{\tilde{\alpha}}{\alpha}x^{\tilde{\alpha}-\alpha},$$

which is strictly increasing, since $\tilde{\alpha} - \alpha > 0$. The following figure depicts the optimal inverse prize distribution G^{-1} and the associated assortative allocation $y^A(x) = G^{-1}(F(x))$ for $h(y) = \sqrt{y}$, $Y = 1/6$, and $F(x) = x^\alpha$ for various values of α. Functions G^{-1} and y^A were computed by using (13) for $j = 2$ and noting that

$$k(z) = \frac{(\alpha+1)z - 1}{\alpha z^{\frac{\alpha-1}{\alpha}}} \text{ and } z^* = \frac{1}{\alpha+1}.$$

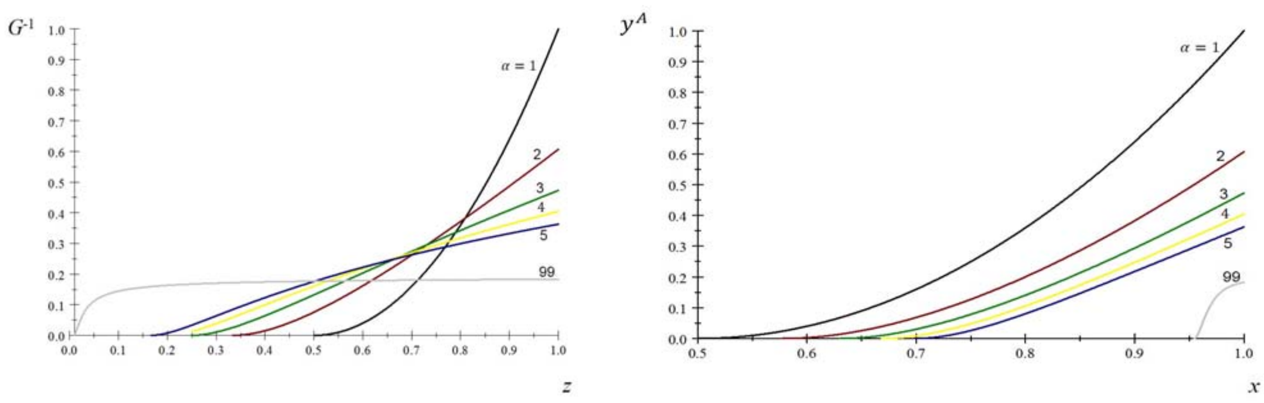

Figure 2: The optimal inverse prize distribution (left) and corresponding assortative allocations (right) for various values of α

Consistent with Proposition 7, the optimal inverse prize distributions for lower values of α cross those for higher values of α once, from below, and are therefore more dispersed. Consistent with Proposition 8, every type optimally obtains a lower prize as α increases. Note that as α increases the mass $1-z^*$ of types that obtain a prize increases (since $z^* = 1/(\alpha+1)$ decreases), but the set of types $[x^*, 1]$ that obtain a prize shrinks (since $x^* = (\alpha+1)^{-1/\alpha}$ increases).

We now consider the effect of increasing the budget Y on the optimal prize distribution. By Proposition 2, such an increase only has an effect when the budget is not large ($Y < m(1 - F(x^*))$). In this case, the increase leads to a FOSD shift in the optimal prize distribution, for any distribution F and regardless of whether h is convex or concave. Every type obtains a weakly higher prize, because the type distribution does not change.

Proposition 9 *If h is weakly convex or concave, an increase in the budget Y leads to a FOSD shift in the optimal prize distribution. In particular, every type obtains a weakly higher prize.*

The following figure depicts the optimal prize distributions for different budgets $Y \leq 1/2$, with $m = 1$, $h(y) = \sqrt{y}$, and F uniform. Since $z^* = 1/2$, Proposition 2 applies for $Y \geq 1/2$, at which point the optimal prize distribution is a mass $1/2$ of prize 1 (the highest possible prize) and a mass $1/2$ of prize 0. The prize distributions were computed by using (13) for $Y \leq 1/6$ (which implies that $z_{\max} = 1$ and there is no mass of prize $m = 1$), and by using the more general characterization from Section 5.2 for $Y > 1/6$ (which implies that $z_{\max} < 1$ and there is a mass of prize $m = 1$).[38] Consistent with Proposition 9, the distributions for higher budgets FOSD those for lower budgets.

[38] The optimal distributions are

$$G(y) = \begin{cases} \frac{1}{2} + \sqrt{\frac{y}{24Y}} & y \in [0, 6Y] \\ 1 & y \in [6Y, 1] \end{cases} \quad \text{and } G(y) = \begin{cases} \frac{1}{2} + \sqrt{\frac{y(3-6Y)^2}{16}} & y \in [0, 1) \\ 1 & y = 1 \end{cases}$$

for $0 < Y \leq 1/6$ and $1/6 < Y \leq 1/2$, respectively.

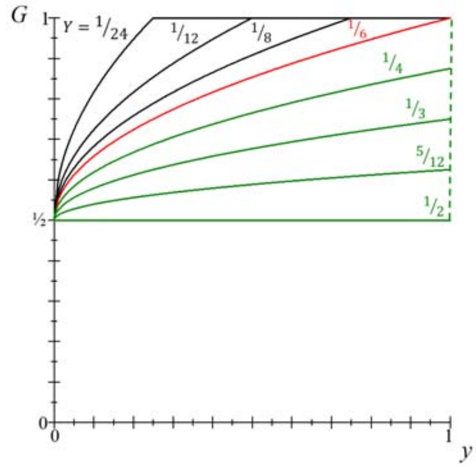

Figure 3: The optimal prize distribution as the budget increases from 0 to 1/2.

7.2 Maximal average performance

We now consider general cost functions c and valuation functions h. We first consider how varying the limit type distribution affects the maximal average performance, attained by the optimal prize distribution. The following result shows that a FOSD shift in the limit type distribution increases the average performance for any prize distribution.

Proposition 10 *If $\widetilde{F}$ FOSD F, then for any prize distribution G the average performance is higher under $\widetilde{F}$ than under F. In particular, the maximal average performance is higher under $\widetilde{F}$ than under F.*

While it may seem intuitive that a more able pool of players will generate higher equilibrium performance, this is not always the case in contests with a small number of players. To see this, consider a two-player all-pay auction with complete information and one prize. The prize is $y = 1$, the prize valuation function satisfies $h(1) = 1$, and the cost function is $c(t) = t$. Players' publicly observed types satisfy $0 < x_1 < x_2 < 1$. It is well known (Hillman and Riley (1989)) that in the unique equilibrium player 2 chooses a bid by mixing uniformly on the interval $[0, x_1]$ and player 1 bids 0 with probability $1 - x_1/x_2$ and with the remaining probability mixes uniformly on the interval $[0, x_1]$. The resulting expected aggregate bids are $x_1/2 + (x_1)^2 / (2x_2)$, which monotonically increase in x_1 and monotonically decrease in x_2. Thus, an increase in player 2's type, even when accompanied by a small increase in player 1's type, decreases the expected aggregate bids. The intuition is that the increased asymmetry between the players, which discourages competition, outweighs the increase in their types, which encourage higher bids. The intuition for Proposition 10 is that in a large contest competition is "localized" in the sense that players compete against players with similar types.[39] Therefore, any decrease in local competition between some types resulting from a FOSD shift in players' type distribution is more than compensated for by an increase in local competition between some higher types.

We now consider the effect of increasing the budget. This clearly weakly increases the maximal average performance, since the set of feasible prize distributions increases. More interestingly, suppose that the designer can determine the budget Y. Suppose that the budget cost is strictly increasing, continuously differentiable, and takes high enough values for large Y to make the designer never choose such values. To compare the marginal budget cost to the marginal budget benefit, consider the most relevant case of concave h and convex c. Propositions 4 and 6 show that Case 1 of Lemmas 1 and 2 applies. The shadow cost λ is then the marginal budget benefit, so the optimal budget Y can be identified by comparing the marginal budget cost to λ. As an example, consider the contest with $m = 1$, $h(y) = \sqrt{y}$, $c(t) = t$, and F uniform for $Y \leq 1/2$, which was discussed at the end of Section 7.1. The characterization from Section 5.2 shows that $\lambda = 1/\sqrt{24Y}$ for $Y \leq 1/6$, and $\lambda = (3/4 - 3Y/2)$ for $Y > 1/6$.[40] Consistent with Proposition 2, $\lambda = 0$ for $Y = 1/2$, so the budget will never optimally exceed 1/2. With a linear budget cost of Y, for example, the optimal budget is $Y = 1/24$, and with a quadratic cost of Y^2 the optimal budget is 3/14.

[39] A discussion of this phenomenon appears in Bulow and Levin (2006).

[40] The maximal average performance is $\sqrt{Y/6}$ for $Y \leq 1/6$ and $[12Y(1-Y)+1]/16$ for $Y \geq 1/6$. The difference between the functional forms is due to the atom at the highest possible prize, 1, which appears when the budget exceeds 1/6.

8 An illustrative example

To illustrate our results and the potential benefit of a gradual prize structure, we consider a grant competition in which research proposals are ranked according to their quality, and prizes are awarded according to this rank order.[41] Researchers vary in their marginal prize valuations, and the cost of a research proposal is convex and increasing in its quality.[42] The designer wishes to maximize the aggregate quality of the submitted proposals.[43] We consider an example of a contest with quadratic costs, uniform limit type distribution, and risk neutral players, and compare the average performance generated by the optimal prize distribution to the one generated by awarding identical prizes.

Formally, we let $c(t) = t^2$, $h(y) = y$, and F be uniform. Suppose a mass I of identical prizes $Y/I > 0$ is awarded (along with a mass $1 - I$ of prize 0). Then, (2) shows that types $x \leq F^{-1}(1-I)$ bid 0, and types $x > F^{-1}(1-I)$ bid $c^{-1}\left(F^{-1}(1-I)\, h(Y/I)\right)$. The average performance is therefore

$$Ic^{-1}\left(F^{-1}(1-I)\, h\left(\frac{Y}{I}\right)\right) = I\sqrt{(1-I)\frac{Y}{I}} = \sqrt{Y}\sqrt{(1-I)\, I}. \tag{23}$$

For the optimal prize distribution G, recall that Proposition 6 shows that $z_{\min} < z_{\max}$ and G may have atoms only at 0 and m. In the appendix, we use the conditions in Case 1 of Lemma 2 to derive G^{-1}. Suppose first that m, the highest possible prize, is at least $4Y$. Then $G^{-1}(z) = 4z^3Y$ for z in $[0,1]$. This distribution is independent of m, and the associated average performance is $\sqrt{Y/3}$. Consistent with Proposition 6, every positive type obtains a positive prize, the prizes increase gradually from 0 to $4Y$, and there are no atoms (see Figure 4 below). The ratio between $\sqrt{Y/3}$ and $\sqrt{Y}\sqrt{(1-I)\, I}$ from (23) is $1/\sqrt{3(1-I)\, I} \geq 2/\sqrt{3} = 1.1547$ (achieved at $I = 1/2$). This shows that replacing identical

[41] Quality can be a function of various variables, such as conceptual novelty, technical novelty, applied value, and clarity. We assume that these dimensions are aggregated to a single score that is used to rank the proposals.

[42] As discussed immediately after (1), it is immediate to allow the players to vary in their marginal performance cost instead of or in addition to their marginal prize valuations.

[43] This is appropriate for research areas, such as economics, in which obtaining funding is often not crucial for the realization of a project, but where competing for such funding may serve to motivate researchers.

prizes with the optimal prize distribution increases the average performance by at least a 15 percent. For suboptimal values of I the increase can be substantially higher. For $I = 1/4$, for example, which roughly mimics the recent proposal funding rates of the NSF's Directorate for Social, Behavioral, and Economic Science,[44] the increase is 33 percent.

Now suppose that the highest possible prize is restricted relative to the budget, so $m < 4Y$. Then, as long as $m \geq 8Y/5$, we have $G^{-1}(z) = 27z^3m^4/\left(64(m-Y)^3\right)$ for z in $[0, 4(m-Y)/(3m)]$ and $G^{-1}(z) = m$ for z in $[4(m-Y)/(3m), 1]$. The associated average performance is $\sqrt{m-Y}\,(1 - 8(m-Y)/(9m))$. Every positive type still obtains a positive prize, and the prizes increase gradually from 0 to m, but there is also a mass $(4Y - m)/(3m)$ of prize m (see Figure 4 below). If m falls below $8Y/5$, then the budget in excess of $5m/8$ is optimally not used, so the optimal prize distribution coincides with the one for $m = 8Y/5$. Notice that unlike the case of linear costs, and consistent with Proposition 6, even when the budget is large ($Y \geq 5m/8$) the optimal prize distribution still awards all prizes between 0 and m, and every positive type obtains a prize.

Restricting the highest possible prize limits the improvement in average performance provided by the optimal prize distribution. For example, the ratio of the average performances for $I = 1/2$ (the optimal mass of identical prizes) is $2/\sqrt{3}$ for $m = 4Y$ and $m \leq 8Y/5$, just like with $m \geq 4Y$, but decreases in m on $[4Y, 2Y]$, reaches a minimum of $10/9$ at $2Y$, and increases on $[2Y, 8Y/5]$.[45]

[44] According to the NSF's 2017 budget request to congress, the rates were 24 percent in 2015 (1042 out of 4284), 23 percent in 2016 (1000 out of 4300), and estimated to be 26 percent in 2017 (1120 out of 4300).

[45] The ratio is $\left(\sqrt{m-Y}\,(1 - (8(m-Y))/(9m))\right)/\left(\sqrt{\min\{2Y, m\}}/(2\sqrt{2})\right)$.

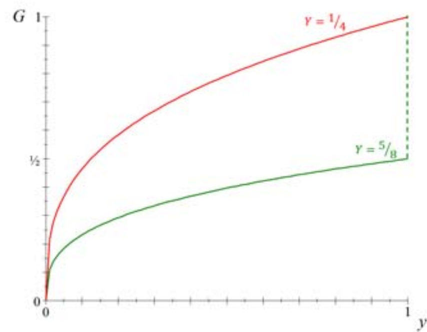

Figure 4: The optimal prize distribution for $m = 1$ and $Y = 1/4$ (red) and $Y = 5/8$ (green)

9 Previous results

Several previous papers consider maximizing the expected aggregate output (or effort) in various contests. Two among them, Glazer and Hassin (1988) and MS (Moldovanu and Sela (2001)) examine this maximization with respect to the prize structure subject to a budget constraint.[46] Both papers study contests in which players' utilities are special cases of (1).

Glazer and Hassin (1988) analyze contests in which contestants are randomly drawn from a population, and use a somewhat specific concept of equilibrium,[47] which facilitates their analysis in a manner similar to that in which our limit approach facilitates the analysis of large contests. They derive an optimal prize structure in two cases. First, when contestants' ability is uniformly distributed in the population, the costs are linear, and prize valuations are weakly concave, they obtain a result that corresponds to our Propositions 3 and 4. Second, when all contestants have identical abilities, they show that the optimal prize structure has $n - 1$ equal prizes and one prize of 0. This result is specific to discrete contests.[48]

[46] Moldovanu, Sela, and Shi (2007) and Immorlica, Stoddard, and Syrgkanis (2015) study this maximization in the context of social status.

[47] They disregard the consistency condition between the distribution of abilities in the population and the equilibrium distribution of output of a randomly chosen contestant.

[48] It can be shown in our setting, and consistent with their result, that when the limit distribution of types F converges to a Dirac distribution, the optimal prize distribution also converges to a Dirac distribution.

MS restrict attention to the symmetric equilibria of discrete contests with ex-ante symmetric contestants, incomplete information, and linear prize valuations. They show that for weakly concave costs, it is optimal to award the entire budget as a single prize.[49] MS also point out that with convex costs awarding the entire budget as a single prize may be inferior to splitting the budget between two prizes.

Proposition 3 is an analogue of the result of MS for linear costs. Although Proposition 3 was established under Assumption 1, Corollary 2 in the online appendix shows that this result does not require Assumption 1. Note that Proposition 3 holds for weakly convex (not necessarily linear) prize valuations. In addition, Proposition 3 can be generalized to weakly concave costs by using the conditions in Case 2 of Lemma 2 instead of those in Case 2 of Lemma 1.

Proposition 6 is related the result of MS that shows that with convex costs splitting the budget into two prizes is sometimes better than awarding the entire budget as a single prize. Our results go beyond this, and characterize the optimal prize structure. This facilitates deriving comparative statics and investigating applications. In addition, our results apply to all equilibria of contests with a large, but finite, number of players. The players may be ex-ante symmetric or asymmetric, and may or may not have private information.

10 Concluding remarks

This paper investigates the performance-maximizing prize structures in contests with many contestants. Our key qualitative finding is that risk aversion and convex performance costs call for numerous prizes of different value. This has implications for sales and workplace competitions, and possibly for certain research grant competitions. The analysis facilitates comparative statics, and enables deriving closed-form approximations of the performance-maximizing prize distributions for concrete utility functions and distributions of player types.

[49] Kaplan and Zamir (2016) notice that this result for linear costs is implied by a result from auction theory, which says that if an auction maximizes revenue, the object must be allocated (if it is allocated at all) to the highest bidder. (The auction-theory result also holds when the object must be allocated to some bidder.)

Our approach can be used to investigate many other contest design questions. One example is maximizing the expected aggregate performance when the designer does not have complete discretion in allocating the budget across prizes (as in the case of identical prizes in Section 8), or when the budget is also determined optimally (as discussed in Section 7.2). Another example is maximizing a weighted sum of contestants' performance. For example, the designer may want to maximize the aggregate performance of contestants whose type exceeds a certain cutoff, or the aggregate performance of the top five percent of contestants. Our analysis can be extended to capture both scenarios.[50] As an example, consider mathematical olympiads, and suppose that the goal is to identify and encourage the development of the most mathematically gifted individuals. This would correspond in our setting to maximizing the aggregate performance of a top fraction of the players. In this case, a minor modification of our analysis implies that it is optimal to spend the entire budget on prizes for this top fraction of contestants. Conditional on this top fraction, the optimal type distribution is the one that maximizes the average performance for a distribution of types that is the conditional distribution of the types in the top fraction. More generally, many objectives can be investigated by identifying the corresponding objectives in the limit setting and showing that they approximate the ones in large contests, as we do in Section 4.2, and then solving the optimization problem in the limit setting, using existing tools or from first principles, as we do here.

11 Appendix

Proof of Corollary 1. Theorem 1 shows that for large n, in any equilibrium of the n-th contest the expected average performance is within $\varepsilon/2$ of

$$\frac{\sum_{i=1}^{n}\int_0^1 t^A(x)\,dF_i^n(x)}{n}=\int_0^1 t^A(x)\,dF^n(x),$$

where the equality follows from the definition of F^n. In addition,

$$\int_0^1 t^A(x)\,dF^n(x)\to_n\int_0^1 t^A(x)\,dF(x),$$

[50] In the limit setting there is no distinction between the two scenarios, because higher types choose higher performance.

which follows from the fact that t^A is monotonic and the assumption that F is continuous, because $\int gdF^n \to_n \int gdF$ for any bounded and measurable function g for which distribution F assigns measure 0 to the set of points at which function g is discontinuous. (This fact is established as the first claim of the proof of Theorem 25.8 in Billingsley (1995).) Thus, for large n, $\int_0^1 t^A(x)\, dF^n(x)$ is within $\varepsilon/2$ of $\int_0^1 t^A(x)\, dF(x)$.

Proof of Claim 1. Let $(G^n)_{n=1}^{\infty}$ be a sequence on which (3) converges to its supremum, and which satisfies the budget constraint. By passing to a convergent subsequence (in the weak*-topology) if necessary, assume that G^n converges to some G. We will show below that $(G^n)^{-1}$ converges almost surely to G^{-1}. This will imply that $(y^n)^A(x) = (G^n)^{-1}(F(x))$ converges almost surely to $y^A(x) = G^{-1}(F(x))$, and since functions h and c^{-1} are continuous, also that $(t^n)^A(x)$ given by (2) with G replaced with G^n converges almost surely to $t^A(x)$ given by (2). This will in turn imply that the value of (3) with $(G^n)^{-1}$ instead of G^{-1} converges to the value of (3). Finally, as G^n satisfies the budget constraint, G satisfies the budget constraint as well. Indeed, the budget constraints are integrals of a continuous function (mapping y to y) with respect to distributions G and G^n, respectively, and weak*-topology may be alternatively defined as convergence of integrals of continuous functions.

Thus, it suffices to show that $(G^n)^{-1}$ converges to G^{-1}, except perhaps on the (at most) countable set $R = \{r \in [0,1] : \text{there exist } y' < y'' \text{ such that } G(y) = r \text{ for } y \in (y', y'')\}$.

Suppose first that for some $r \in [0,1]$ and $\delta > 0$ we have that $(G^n)^{-1}(r) \leq G^{-1}(r) - \delta$ for arbitrarily large n. Passing to a subsequence if necessary, assume that the inequality holds for all n, and that $(G^n)^{-1}(r)$ converges to some $y \leq G^{-1}(r) - \delta$. Then, there exists a prize z such that $y < z < G^{-1}(r)$ and G is continuous at z. We cannot have that $G(z) = r$, since this would imply that $G^{-1}(r) \leq z$. Thus, $G(z) < r$. Since $G^n(z)$ converges to $G(z)$, as G is continuous at z, we have that $G^n(z) < r$ for large enough n. This yields $z \leq (G^n)^{-1}(r)$, contradicting the assumption that $(G^n)^{-1}(r)$ converges to $y < z$.

Suppose now that for some $r \in [0,1] - R$ and $\delta > 0$ we have that $(G^n)^{-1}(r) \geq G^{-1}(r) + \delta$ for arbitrarily large n. Passing to a subsequence if necessary, assume that the inequality holds for all n, and that $(G^n)^{-1}(r)$ converges to some $y \geq G^{-1}(r) + \delta$. Then, there exists a prize z such that $G^{-1}(r) < z < y$ and G is continuous at z. We have that $r < G(z)$, as

$r \notin R$. Since $G^n(z)$ converges to $G(z)$, as G is continuous at z, we have that $r \leq G^n(z)$ for large enough n. This yields $(G^n)^{-1}(r) \leq z$, contradicting the assumption that $(G^n)^{-1}(r)$ converges to $y > z$.

Proof of Proposition 1. Since every sequence of distributions has a converging subsequence in weak*-topology, suppose without loss of generality that $G^n_{\max}$ converges to some distribution G. Denote the value of (3) under distribution G by V. If Part 1 is false, then $G \notin \mathcal{M}$, so $V < M$. The distribution G satisfies the budget constraint, since distributions $G^n_{\max}$ satisfy the budget constraint.

Consider a distribution $G_{\max} \in \mathcal{M}$, and for every n consider an empirical distribution G^n of a set of n prizes, such that G^n converges to $G_{\max}$ in weak*-topology. For example, such a set of n prizes is defined by $y^n_j = G^{-1}_{\max}(j/n)$ for $j = 1, ..., n$.

Corollary **1** shows that for large n the expected average performance in any equilibrium of the n-th contest with empirical prize distribution G^n exceeds $(V + M)/2$. On the other hand, Corollary 1 also shows that for large n the expected average performance in any equilibrium of the n-th contest with empirical prize distribution $G^n_{\max}$ falls below $(V + M)/2$. This contradicts the definition of $G^n_{\max}$ for large n.

For Part 2, Corollary 1 applied to the sequence G^n defined above implies that $\liminf M^n_{\max} \geq M$. If $\limsup M^n_{\max} > M$, then there is a corresponding subsequence of $G^n_{\max}$. A converging subsequence of this subsequence has a limit G. For this G, the value of (3) is by Corollary 1 strictly larger than M, a contradiction.

For Part 3, notice that the proof of Claim 1 also implies that the set M is closed in weak*-topology. Thus, if part 3 were false, there would exist a sequence of contests with empirical prize distributions G^n converging to some G in M, such that the expected average performance in an equilibrium of the n-th contest with empirical prize distribution G^n would be lower than $M^n_{\max} - \varepsilon$. This would contradict Part 2 and Corollary 1.

Optimal control formulation. By substituting $\widetilde{z} = F(\widetilde{x})$, we obtain that

$$\int_0^x h\left(G^{-1}(F(\widetilde{x}))\right) d\widetilde{x} = \int_0^{F(x)} \frac{L\left(F^{-1}\left(\widetilde{z}\right)\right)}{f\left(F^{-1}\left(\widetilde{z}\right)\right)} d\widetilde{z},$$

and then by substituting $z = F(x)$ we rewrite (6) as

$$\int_0^1 \left(c^{-1}\left(F^{-1}(z)L\left(F^{-1}(z)\right) - \int_0^z \frac{L\left(F^{-1}\left(\widetilde{z}\right)\right)}{f\left(F^{-1}\left(\widetilde{z}\right)\right)}d\widetilde{z}\right)\right)dz =$$

$$\int_0^1 \left(c^{-1}\left(F^{-1}(z)h\left(G^{-1}(z)\right) - \int_0^z \frac{h\left(G^{-1}(\widetilde{z})\right)}{f(F^{-1}(\widetilde{z}))}d\widetilde{z}\right)\right)dz,$$

where the equality follows from $L(x) = h\left(y^A(x)\right) = h\left(G^{-1}\left(F(x)\right)\right)$. This is an optimal control problem with control variable $U(z) = G^{-1}(z)$ and state variable $X(z) = \int_0^z (h\left(G^{-1}(\widetilde{z})\right)/f(F^{-1}(\widetilde{z})))d\widetilde{z}$, where $X(0) = 0$ and $X'(z) = h\left(U(z)\right)/f(F^{-1}(z))$ (see, for example, Chapter 1.1 in Sethi and Thompson (2000)). The objective functional is

$$\int_0^1 \left(c^{-1}\left(F^{-1}(t)h\left(U(t)\right) - X(t)\right)\right)dt.$$

The budget constraint can be represented as a constraint on the terminal values (in the optimal control terminology) by considering a vector state variable $(X_1(t), X_2(t))$, where $X_1(t) = X(t)$, $X_2'(t) = -U(t)$, and $X_2(0) = Y$. Then, the budget constraint is equivalent to the inequality $X_2(1) \geq 0$.

Proof of Lemma 1. Let G be an optimal distribution, and suppose that $z_{\min} < z_{\max}$. We now show that $h'\left(G^{-1}\left(\underline{z}\right)\right)k\left(\underline{z}\right) = h'\left(G^{-1}\left(\overline{z}\right)\right)k\left(\overline{z}\right)$ for all $\underline{z}, \overline{z} \in (z_{\min}, z_{\max})$. The idea is that if this were not the case, e.g., if we had $<$ instead of $=$, then we could increase G^{-1} around $\overline{z}$ and decrease G^{-1} around $\underline{z}$, thereby increasing the average performance. We must be careful, however, not to violate the budget constraint, and to maintain the monotonicity of G^{-1}. These properties will be easier to control if we first approximate G^{-1} by a piecewise constant function.

To simplify notation, we assume that $m = 1$. We approximate G^{-1} by a sequence of inverse distribution functions $((G^n)^{-1})_{n=1}^{\infty}$. To define $(G^n)^{-1}$, partition interval $[0,1]$ into intervals of size $1/2^n$, and set the value of $(G^n)^{-1}$ on interval $(j/2^n, (j+1)/2^n]$ to be constant and equal to the highest number in the set $\{0, 1/2^n, 2/2^n, ..., (2^n-1)/2^n, 1\}$ that is no higher than $G^{-1}(j/2^n)$. By left-continuity of G^{-1}, $(G^n)^{-1}$ converges pointwise to G^{-1}. By definition of $(G^n)^{-1}$ and monotonicity of G^{-1}, $(G^n)^{-1}$ satisfies the budget constraint (5).

Suppose that $h'\left(G^{-1}\left(\underline{z}\right)\right)k\left(\underline{z}\right) < h'\left(G^{-1}\left(\overline{z}\right)\right)k\left(\overline{z}\right)$ for some $\underline{z}, \overline{z} \in (z_{\min}, z_{\max})$. By left-continuity of G^{-1}, and continuity of h' and k, the previous inequality also holds for points slightly smaller than $\underline{z}$ and $\overline{z}$. Thus, there are $\delta > 0$, N, and intervals $(l/2^N, (l+1)/2^N]$ and $(j/2^N, (j+1)/2^N]$, such that for every $n \geq N$ we have $h'\left((G^n)^{-1}(\overline{z}')\right) k(\overline{z}') - h'\left((G^n)^{-1}(\underline{z}')\right) k(\underline{z}') > \delta$ for any $\overline{z}' \in (j/2^N, (l+1)/2^N]$ and $\underline{z}' \in (l/2^N, (j+1)/2^N]$.

Denote by I the infimum of the values $h'\left((G^n)^{-1}(z)\right) k(z)$ for $n \geq N$ and z in the former interval, and by S the supremum of the values $h'\left((G^n)^{-1}(z)\right) k(z)$ for $n \geq N$ and z in the latter interval. Thus, we have that $I - S \geq \delta$. Define functions $(H^n)^{-1}$ by increasing the value of $(G^n)^{-1}$ on $(j/2^N, (j+1)/2^N]$ by ε, and decreasing the value of $(G^n)^{-1}$ on $(l/2^N, (l+1)/2^N]$ by ε, so the budget constraint is maintained. For sufficiently small $\varepsilon > 0$, the former change increases (8) at least by $\left(\varepsilon/2^N\right)(I - \delta/3)$, and the latter change decreases (8) at most by $\left(\varepsilon/2^N\right)(S + \delta/3)$. This increases the value of (8) by at least $\delta\varepsilon/\left(3 \cdot 2^N\right)$ (for all $n \geq N$), since $I - S \geq \delta$.

If functions $(H^n)^{-1}$ are monotone, they are inverse distribution functions, and the value of (8) with $(H^n)^{-1}$ instead of G^{-1} exceeds, for large enough n, the value of (8) for G^{-1}. If functions $(H^n)^{-1}$ are not monotone, define $\left(\widetilde{H}^n\right)^{-1}$ by setting its value on interval $(0, 1/2^n]$ to the lowest value of $(H^n)^{-1}$ over intervals $(0, 1/2^n], (1/2^n, 2/2^n], ..., ((2^n - 1)/2^n, 1]$, setting its value on interval $(1/2^n, 2/2^n]$ to the second lowest value of $(H^n)^{-1}$ on these intervals, etc. The value of (8) with $(\widetilde{H}^n)^{-1}$ instead of G^{-1} is higher than with $(H^n)^{-1}$ instead of G^{-1}, because k is an increasing function.

The second condition in Case 1 and the condition in Case 2 are obtained by analogous arguments, noticing that $z_{\min} > 0$ (since $k(0) < 0$) and, since k is increasing and continuous, the inequality $h'(m) k(z) \geq \lambda$ for $z > z_{\max}$ is equivalent to $h'(m) k(z_{\max}) \geq \lambda$.

Proof of Proposition 5. Let $z_{\min}^m$, $z_{\max}^m$, and λ^m denote $z_{\min}$, $z_{\max}$, and λ for a given m. The proof of Proposition 4 shows that $z_{\min}^m < z_{\max}^m$ for all m. We claim that λ^m weakly increases with m. Suppose to the contrary that $\lambda^{m'} > \lambda^{m''}$ for some $m' < m''$.

Since $h'\left((G_{\max}^m)^{-1}(z)\right) k(z) = \lambda^m$ for all $z \in (z_{\min}^m, z_{\max}^m]$ and h' is decreasing, $h'(0) k(z) \geq \lambda^m$ for all $z \in (z_{\min}^m, z_{\max}^m]$, and since k is continuous, we have $h'(0) k(z_{\min}^m) \geq \lambda^m$. Since we also have $h'(0) k(z_{\min}^m) \leq \lambda^m$ (because we are in Case 1 of Section 5.1), we obtain $h'(0) k(z_{\min}^m) = \lambda^m$. Since k is increasing, this implies that $z_{\min}^{m'} > z_{\min}^{m''}$. In particular, we have (a): $(G_{\max}^{m'})^{-1}(z) = 0 \leq (G_{\max}^{m''})^{-1}(z)$ for all $z \leq z_{\min}^{m'}$, and the inequality is strict for

$z \in (z_{\min}^{m''}, z_{\min}^{m'})$. Since $h'\left((G_{\max}^{m})^{-1}(z)\right)k(z) = \lambda^m$ for all $z \in (z_{\min}^{m}, z_{\max}^{m}]$ and h' is decreasing, we have (b): $(G_{\max}^{m'})^{-1}(z) \leq (G_{\max}^{m''})^{-1}(z)$ for all $z \in (z_{\min}^{m'}, \min\{z_{\max}^{m'}, z_{\max}^{m''}\}]$. If $z_{\max}^{m'} \geq z_{\max}^{m''}$, then we have (c): $(G_{\max}^{m'})^{-1}(z) \leq m' < (G_{\max}^{m''})^{-1}(z) = m''$ for $z > \min\{z_{\max}^{m'}, z_{\max}^{m''}\}$. If $z_{\max}^{m'} < z_{\max}^{m''} \leq 1$, then $h'(m')k\left(z_{\max}^{m'}\right) \geq \lambda^{m'}$ (because we are in Case 1 of Section 5.1). But $h'\left(\left(G_{\max}^{m''}\right)^{-1}\left(z_{\max}^{m'}\right)\right)k\left(z_{\max}^{m'}\right) = \lambda^{m''}$, so $\lambda^{m'} > \lambda^{m''}$ implies that $\left(G_{\max}^{m''}\right)^{-1}\left(z_{\max}^{m'}\right) \geq m'$. Thus, as the inverse of any *cdf* is increasing, we again obtain (c), except that this time $(G_{\max}^{m''})^{-1}(z) \leq m''$. Now, (a), (b), and (c) imply that the budget constraint cannot be satisfied with equality by both $G_{\max}^{m'}$ and $G_{\max}^{m''}$, which completes the proof that λ^m weakly increases with m.

By $h'(0)k(z_{\min}^{m}) = \lambda^m$, we obtain that $z_{\min}^{m}$ also weakly increases with m. If $z_{\max}^{m'} > z_{\max}^{m''}$ for $m' < m''$, then $h'\left(\left(G_{\max}^{m'}\right)^{-1}\left(z_{\max}^{m''}\right)\right)k(z_{\max}^{m''}) = \lambda^{m'}$ and $h'(m'')k\left(z_{\max}^{m''}\right) \geq \lambda^{m''} \geq \lambda^{m'}$, which would imply that $\left(G_{\max}^{m'}\right)^{-1}\left(z_{\max}^{m''}\right) \geq m'' > m'$. Thus, $z_{\max}^{m}$ also weakly increases with m. Moreover, $z_{\max}^{m}$ converges to 1 as m diverges, because otherwise the budget constraint would be violated for large enough values of m. Because $h'((G_{\max}^{m})^{-1}(z))k(z) = \lambda^m$, we have that $(G_{\max}^{m''})^{-1}(z) \leq (G_{\max}^{m'})^{-1}(z)$ for all $z \leq z_{\max}^{m'}$.[51]

Notice that $z_{\max}^{m} = 1$ for sufficiently large m. Otherwise, the condition $h'(m)k(z_{\max}) \geq \lambda$ cannot be satisfied for large enough m, by the assumption that $h'(y) \to 0$ as $y \to \infty$. And if $z_{\max}^{m} = 1$ for some m, then $(G_{\max}^{m'})^{-1} \equiv (G_{\max}^{m})^{-1}$ for all $m' \geq m$, because $(G_{\max}^{m'})^{-1}(z) \leq (G_{\max}^{m})^{-1}(z)$ for all $z \leq z_{\max}^{m}$ and both $G_{\max}^{m}$ and $G_{\max}^{m'}$ satisfy the budget constraint with equality. This completes the proof.

Proof of Lemma 2. The idea of the proof is analogous to that of the proof of Lemma 1. As in that proof, suppose that $z_{\min} < z_{\max}$, that is, we are in Case 1; the condition in Case 2 is obtained by analogous arguments. For an optimal distribution G, approximate G^{-1} by a sequence of inverse distribution functions $((G^n)^{-1})_{n=1}^{\infty}$ that are constant on intervals $(j/2^n, (j+1)/2^n]$ and with values in the set $\{0, 1/2^n, 2/2^n, ..., (2^n-1)/2^n, 1\}$. If (19) is violated, we construct functions $(H^n)^{-1}$ (also constant on intervals $(j/2^n, (j+1)/2^n]$ and

[51] Therefore, $(G_{\max}^{m})^{-1}$ converges pointwise to some G^{-1} on $[0,1)$, even when $h'(y) \nrightarrow 0$ as $y \to \infty$. We cannot conclude, however, that this G^{-1} is an inverse *cdf*. For example, G^{-1} can be a constant function equal to 0.

with values in the set $\{0, 1/2^n, 2/2^n, ..., (2^n-1)/2^n, 1\}$) such that the value of the target function (14) with $(H^n)^{-1}$ instead of G^{-1} exceeds, for large enough n, that of (14) for G^{-1}. This part of the proof replicates the argument from the corresponding part of the proof of Lemma 1, and will be omitted. If for a large enough n function $(H^n)^{-1}$ is nondecreasing, it is an inverse distribution function. We then obtain a contradiction to the optimality of G^{-1}, which completes the proof. If function $(H^n)^{-1}$ is not monotone, we define another function $\left(\widetilde{H}^n\right)^{-1}$ whose value on interval $(0, 1/2^n]$ is equal to the lowest value of $(H^n)^{-1}$ over intervals $(0, 1/2^n], (1/2^n, 2/2^n], ..., ((2^n-1)/2^n, 1]$, whose value on interval $(1/2^n, 2/2^n]$ is equal to the second lowest value of $(H^n)^{-1}$ on these intervals, and so on. We will complete the proof by showing that the value of (14) is no lower for $(\widetilde{H}^n)^{-1}$ than for $(H^n)^{-1}$ for sufficiently large n's.

To show this, we will consider only two adjacent intervals $(j/2^n, (j+1)/2^n]$ and $(l/2^n, (l+1)/2^n]$ (that is, $j+1=l$) such that $(H^n)^{-1}(z) = U$ on $(j/2^n, (j+1)/2^n]$ and $(H^n)^{-1}(z) = D$ on $(l/2^n, (l+1)/2^n]$, where $D < U$, and estimate the effect on (14) of changing the value of $(H^n)^{-1}$ on $(j/2^n, (j+1)/2^n]$ to D and changing the value of $(H^n)^{-1}$ on $(l/2^n, (l+1)/2^n]$ to U. We will use the same symbol $(\widetilde{H}^n)^{-1}$ to denote the function obtained from $(H^n)^{-1}$ as a result of this change, and we will sometimes use symbol Δ to denote $1/2^n$.

The exchange of D and U does not affect the integrand of (14) on the intervals lower than $(j/2^n, (j+1)/2^n]$. It affects the value of $\widetilde{L}$ on interval $(j/2^n, (j+1)/2^n]$, increasing it by some $\widetilde{\Delta}_j$, as well as the value of $\widetilde{L}$ on interval $(l/2^n, (l+1)/2^n]$, increasing it by some $\widetilde{\Delta}_l$. As a result of the change in $\widetilde{L}$ on the two intervals, (14) increases by

$$\begin{aligned} & \Delta\left[c^{-1}(\widetilde{L}((j+1)/2^n) + \widetilde{\Delta}_j) - c^{-1}(\widetilde{L}((j+1)/2^n))\right] \\ & +\Delta\left[c^{-1}(\widetilde{L}((l+1)/2^n) + \widetilde{\Delta}_l) - c^{-1}(\widetilde{L}((l+1)/2^n))\right] \\ = \; & \Delta\left[c^{-1}(\widetilde{L}((l+1)/2^n) + \widetilde{\Delta}_l) - c^{-1}(\widetilde{L}((j+1)/2^n))\right] \\ & +\Delta\left[c^{-1}(\widetilde{L}((j+1)/2^n) + \widetilde{\Delta}_j) - c^{-1}(\widetilde{L}((l+1)/2^n))\right]. \end{aligned}$$

Observe that

$$\left[\widetilde{L}((l+1)/2^n) + \widetilde{\Delta}_l\right] - \left[\widetilde{L}((j+1)/2^n)\right] = \left[F^{-1}((j+1)/2^n) - F^{-1}(j/2^n)\right][h(U) - h(D)].$$

This is easiest to see by looking at Figure 5, in which the graph of L for $(\widetilde{H}^n)^{-1}$ is obtained from the graph of L for $(H^n)^{-1}$ by moving it to the right by the darkened rectangle, and moving it to the left by the shaded rectangle. By definition, $\widetilde{L}$ for $(H^n)^{-1}$ on $(j/2^n,(j+1)/2^n]$ is equal to the area of the rectangle $[0,h(U)]\times[0,F^{-1}(j/2^n)]$ minus the area to the left of the graph of L for $(H^n)^{-1}$ on the interval $[0,F^{-1}(j/2^n)]$. Similarly, $\widetilde{L}$ for $(\widetilde{H}^n)^{-1}$ on $(l/2^n,(l+1)/2^n]$ is equal to the area of the rectangle $[0,h(U)]\times[0,F^{-1}(l/2^n)]$ minus the area to the left of the graph of L for $(\widetilde{H}^n)^{-1}$ on the interval $[0,F^{-1}(l/2^n)]$. So, the difference between the latter and the former areas consists only of the shaded rectangle.

Similarly,

$$\left[\widetilde{L}((j+1)/2^n)+\widetilde{\Delta}_j)\right]-\left[\widetilde{L}((l+1)/2^n)\right]=\left[F^{-1}((j+1)/2^n)-F^{-1}(j/2^n)\right][h(U)-h(D)].$$

Using the mean value theorem, the increase in (14) caused by changing the value of $\widetilde{L}$ on intervals $(j/2^n,(j+1)/2^n]$ and $(l/2^n,(l+1)/2^n]$ can be expressed as

$$\begin{aligned}&\Delta(c^{-1})'\left(\widetilde{L}_l\right)\left[F^{-1}((j+1)/2^n)-F^{-1}(j/2^n)\right][h(U)-h(D)]\\&+\Delta(c^{-1})'\left(\widetilde{L}_j\right)\left[F^{-1}((j+1)/2^n)-F^{-1}(j/2^n)\right][h(U)-h(D)]\end{aligned}$$

for some $\widetilde{L}_l$ between $\widetilde{L}((l+1)/2^n)+\widetilde{\Delta}_l$ and $\widetilde{L}((j+1)/2^n)$, and some $\widetilde{L}_j$ between $\widetilde{L}((j+1)/2^n)+\widetilde{\Delta}_j$ and $\widetilde{L}((l+1)/2^n)$.

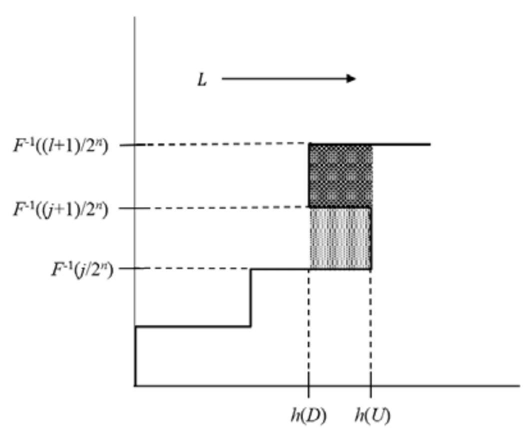

Figure 5: Making H^{-1} monotonic

The exchange of D and U also affects the integrand of (14) on the intervals higher than $(l/2^n, (l+1)/2^n]$. We can estimate this change in the integrand, as we did for the intervals $(j/2^n, (j+1)/2^n]$ and $(l/2^n, (l+1)/2^n]$, by using the mean value theorem. On each interval $(m/2^n, (m+1)/2^n]$, where $m > l$, the integrand increases by

$$(c^{-1})'(\widetilde{L}_m)\{\left[F^{-1}((j+1)/2^n) - F^{-1}(j/2^n)\right][h(U) - h(D)]$$

$$-\left[F^{-1}((l+1)/2^n) - F^{-1}(l/2^n)\right][h(U) - h(D)]\}$$

for some $\widetilde{L}_m$.

Setting $z = (j+1)/2^n$ and dividing the aggregate increase in (14) by $[h(U) - h(D)]$ (which appears in all expressions), we obtain

$$\Delta[(c^{-1})'\left(\widetilde{L}_l\right) + (c^{-1})'\left(\widetilde{L}_j\right)]\left[F^{-1}(z) - F^{-1}(z-\Delta)\right] \tag{24}$$

$$-\{\left[F^{-1}(z+\Delta) - F^{-1}(z)\right] - \left[F^{-1}(z) - F^{-1}(z-\Delta)\right]\}\sum_{m=l+1}^{2^n-1}\Delta(c^{-1})'(\widetilde{L}_m).$$

By using the mean value theorem twice on (24), once in the first line and once in the second line (for function $g(z) = F^{-1}(z+\Delta) - F^{-1}(z)$ on interval $[z-\Delta, z]$), we obtain

$$\Delta^2[(c^{-1})'\left(\widetilde{L}_l\right) + (c^{-1})'\left(\widetilde{L}_j\right)]\left[\frac{1}{f(F^{-1}(z-\zeta))}\right] - \tag{25}$$

$$\Delta\left[\frac{1}{f(F^{-1}(z+\eta))} - \frac{1}{f(F^{-1}(z-\Delta+\eta))}\right]\sum_{m=l+1}^{2^n-1}\Delta(c^{-1})'(\widetilde{L}_m),$$

where $0 \le \zeta, \eta \le \Delta$.

Applying the mean value theorem again, (25) is equal to

$$\Delta^2[(c^{-1})'\left(\widetilde{L}_l\right) + (c^{-1})'\left(\widetilde{L}_j\right)]\left[\frac{1}{f(F^{-1}(z-\zeta))}\right] + \Delta^2\left[\frac{f'(F^{-1}(z'))}{f^3(F^{-1}(z'))}\right]\sum_{m=l+1}^{2^n-1}\Delta(c^{-1})'(\widetilde{L}_m) \tag{26}$$

for some $z' \in [z-\Delta+\eta, z+\eta]$. By continuity of f and f', (26) is equal to

$$\Delta^2[(c^{-1})'\left(\widetilde{L}_l\right) + (c^{-1})'\left(\widetilde{L}_j\right)]\left[\frac{1}{f(F^{-1}(z))}\right] + \Delta^2\left[\frac{f'(F^{-1}(z))}{f^3(F^{-1}(z))}\right]\sum_{m=l+1}^{2^n-1}\Delta(c^{-1})'(\widetilde{L}_m) + o(\Delta^2), \tag{27}$$

where $o(\Delta^2)$ is an expression that tends to zero faster than Δ^2.

To determine the sign of (27), consider two cases: (1) If $f'(F^{-1}(z)) \geq 0$, then (27) is positive for sufficiently small Δ's, since its first component is strictly positive, and the second component is nonnegative; (2) If $f'(F^{-1}(z)) < 0$, then the first component is no smaller than

$$\frac{2\Delta^2}{\overline{c}f(F^{-1}(z))},$$

and the second component is no smaller than

$$\Delta^2 \left[\frac{f'(F^{-1}(z))}{f^3(F^{-1}(z))}\right] \sum_{m=l+1}^{2^n-1} \frac{\Delta}{\underline{c}} = \Delta^2 \left[\frac{f'(F^{-1}(z))}{f^3(F^{-1}(z))}\right] \frac{(1-z)}{\underline{c}} + o(\Delta^2)$$

So, (27) is positive for sufficiently small Δ's by Assumption 2. This completes the proof of (19).

Proof of Proposition 7. We first observe that if G^{-1} crosses $\widetilde{G}^{-1}$ once, from below, that is, if for some $\hat{z} \in (0,1)$ we have $G^{-1}(z) \leq \widetilde{G}^{-1}(z)$ for $z \in [0, \hat{z}]$ and $G^{-1}(z) \geq \widetilde{G}^{-1}(z)$ for $z \in [\hat{z}, 1]$, then $\widetilde{G}$ SOSD G. Indeed, since $Y < m(1 - F(x^*))$ and the condition on $Q(z)$ implies that $\widetilde{F}$ FOSD F, we have that $Y < m\left(1 - \widetilde{F}(\widetilde{x}^*)\right)$. Thus, G and $\tilde{G}$ have the same expectation Y, so SOSD holds if and only if for every $t \in [0,1]$ we have that

$$\int_0^t G(y)\,dy \geq \int_0^t \widetilde{G}(y)\,dy. \tag{28}$$

Since

$$\int_0^1 G(y)\,dy = \int_0^1 \widetilde{G}(y)\,dy = 1 - Y, \tag{29}$$

a sufficient condition for (28) is that $\widetilde{G}$ crosses G once, from below. This happens if and only if G^{-1} crosses $\widetilde{G}^{-1}$ once, from below.

We will now show that G^{-1} crosses $\widetilde{G}^{-1}$ once, from below. For expositional convenience only, suppose that $m = 1$. By Case 1 in Lemma 1, we have for any $z \in (z_{\min}, z_{\max}]$ that

$$h'\left(G^{-1}(z)\right) = \lambda / k(z), \tag{30}$$

and similarly for $\widetilde{z} \in (\widetilde{z}_{\min}, \widetilde{z}_{\max}]$.

We will now show that the single crossing occurs if $\widetilde{k}/k$ strictly decreases on $(\underline{z}, \overline{z}]$, where $\underline{z} = \max\{z_{\min}, \widetilde{z}_{\min}\} \geq z^*$ and $\overline{z} = \min\{z_{\max}, \widetilde{z}_{\max}\}$. By (30) and the concavity of h, for

$z \in (\underline{z}, \overline{z}]$ we have that

$$G^{-1}(z) \leq \widetilde{G}^{-1}(z) \iff h'\left(G^{-1}(z)\right) \geq h'\left(\widetilde{G}^{-1}(z)\right) \iff \lambda/k(z) \geq \widetilde{\lambda}/\widetilde{k}(z) \iff \frac{\widetilde{k}(z)}{k(z)} \geq \frac{\widetilde{\lambda}}{\lambda}.$$

Thus, $G^{-1}(z) \leq \widetilde{G}^{-1}(z)$ if and only if $\widetilde{k}(z)/k(z) \geq \widetilde{\lambda}/\lambda$, and similarly with strict instead of weak inequalities (by strict concavity of h). This last observation implies that if $\widetilde{k}/k$ strictly decreases on $(\underline{z}, \overline{z}]$, then it crosses $\lambda/\widetilde{\lambda}$ at some point $\hat{z} \in (\underline{z}, \overline{z})$; if it did not cross $\lambda/\widetilde{\lambda}$, then (29) would be violated. We then have that $G^{-1}(z) < \widetilde{G}^{-1}(z)$ for $z \in (\underline{z}, \hat{z})$ and $G^{-1}(z) > \widetilde{G}^{-1}(z)$ for $z \in (\hat{z}, \overline{z}]$. Since G^{-1} and $\widetilde{G}^{-1}$ are continuous at $z_{\min}$ and $z_{\max}$, and $\widetilde{z}_{\min}$ and $\widetilde{z}_{\max}$, respectively, (part 2 of Proposition 4, and left-continuity of inverse cdfs), we have that $\underline{z} = z_{\min}$ and $\overline{z} = z_{\max}$, so $G^{-1}(z) \leq \widetilde{G}^{-1}(z)$ for $z \in [0, \underline{z}]$ and $G^{-1}(z) \geq \widetilde{G}^{-1}(z)$ for $z \in (\overline{z}, 1]$. Therefore, $G^{-1}(z) \leq \widetilde{G}^{-1}(z)$ for $z \in [0, \hat{z}]$ and $G^{-1}(z) \geq \widetilde{G}^{-1}(z)$ for $z \in [\hat{z}, 1]$, so G^{-1} crosses $\widetilde{G}^{-1}$ once, from below.

We will complete the proof of the proposition by showing that $\widetilde{k}/k$ strictly decreases on $(z^*, 1]$ (recall that $z^* \leq \underline{z}$) if $Q(z)$ is weakly convex and decreasing on $(z^*, 1]$. Observe that

$$k(z) = \left(-F^{-1}(z)(1-z)\right)' \text{ and } \widetilde{k}(z) = \left(-\widetilde{F}^{-1}(z)(1-z)\right)' = \left(-F^{-1}(z)Q(z)(1-z)\right)',$$

so

$$\frac{\widetilde{k}(z)}{k(z)} = \frac{\left(-F^{-1}(z)(1-z)Q(z)\right)'}{\left(-F^{-1}(z)(1-z)\right)'} = \frac{\left(-F^{-1}(z)(1-z)\right)'Q(z) + \left(-F^{-1}(z)(1-z)\right)Q'(z)}{\left(-F^{-1}(z)(1-z)\right)'} =$$

$$Q(z) + \frac{\left(-F^{-1}(z)(1-z)\right)Q'(z)}{k(z)}.$$

We need to show that $\left(\widetilde{k}(z)/k(z)\right)' < 0$ for $z > z^*$. Since $Q'(z) < 0$, it suffices to show that the derivative of the fraction is weakly negative. This holds if and only if

$$\left(\left(-F^{-1}(z)(1-z)\right)Q'(z)\right)'k(z)) - \left(-F^{-1}(z)(1-z)\right)Q'(z))k'(z) \leq 0.$$

And since

$$k(z) > 0 \text{ and } -\underbrace{\left(-F^{-1}(z)(1-z)\right)}_{-}\underbrace{Q'(z)}_{-}\underbrace{k'(z)}_{+} \leq 0,$$

where $k(z) > 0$ follows from $z > z^*$ and $k'(z) \geq 0$ by Assumption 1, it suffices to observe that

$$\left(-F^{-1}(z)(1-z)Q'(z)\right)' = \underbrace{k(z)}_{+}\underbrace{Q'(z)}_{-} + \underbrace{\left(-F^{-1}(z)(1-z)\right)}_{-}\underbrace{Q''(z)}_{+} \leq 0.$$

Proof of Proposition 8. This proof continues the proof of Proposition 7. By substituting $z = F(x)$ into (30) we obtain $h'\left(y^A(x)\right) = \lambda / P(x)$ for any $x \in \left(F^{-1}(z_{\min}), F^{-1}(z_{\max})\right]$. Recall that $\widetilde{k}/k$ strictly decreases on $(z^*, 1]$ (see the proof of Proposition 7). Since $\widetilde{k}(1) = k(1) = 1$, we have that $\widetilde{k}(z)/k(z) > 1$ for all $z \in (z^*, 1)$. This implies that $\widetilde{\lambda} > \lambda$. Indeed, for $\widetilde{\lambda} \leq \lambda$ we would have $z_{\min} \geq \widetilde{z}_{\min}$ and $z_{\max} \geq \widetilde{z}_{\max}$ by (10) and (9), and by the conditions of Lemma 1 for all $z \in (z_{\min}, \widetilde{z}_{\max}]$, we would have

$$h'\left(G^{-1}(z)\right) = \lambda / k(z) > \widetilde{\lambda}/\widetilde{k}(z) = h'\left(\widetilde{G}^{-1}(z)\right) \Rightarrow \widetilde{G}^{-1}(z) > G^{-1}(z).$$

This, together with $z_{\min} \geq \widetilde{z}_{\min}$ and $z_{\max} \geq \widetilde{z}_{\max}$, would imply that $\widetilde{G}^{-1}(z) \geq G^{-1}(z)$ for all z, and $\widetilde{G}^{-1}(z) > G^{-1}(z)$ for a positive measure of z's, which would violate (29).

Thus, $\widetilde{\lambda} > \lambda$, so $\widetilde{z}_{\min} > z_{\min}$ and $\widetilde{z}_{\max} \geq z_{\max}$. And since $\widetilde{P}(x) \leq P(x)$ for all x,

$$h'\left(y^A(x)\right) = \lambda / P(x) < \widetilde{\lambda}/\widetilde{P}(x) = h'\left(\tilde{y}^A(x)\right) \Rightarrow \tilde{y}^A(x) < y^A(x)$$

for $x \in \left(F^{-1}(\widetilde{z}_{\min}), F^{-1}(z_{\max})\right]$. This, together with $F^{-1}(z_{\min}) < F^{-1}(\widetilde{z}_{\min})$ and $F^{-1}(z_{\max}) \leq F^{-1}(\widetilde{z}_{\max})$, implies that every type obtains a weakly lower prize under $\widetilde{F}$ than under F.

Proof of Proposition 9. If h is weakly convex, the result follows immediately from Proposition 3. Suppose that h is weakly concave, but not linear. By Proposition 4, we are in Case 1 of Lemma 1. Without loss of generality, we assume that the budget constraint holds with equality for the higher budget (and therefore also for the lower budget). If this were not the case, we would consider the intermediate budget for which the budget constraint holds with equality but the optimal distribution of prizes already consists of an atom at 0 and an atom at m. Then, we would first compare the lower budget with the intermediate one, and then the intermediate one with the higher one.

If we had that $\lambda_l \leq \lambda_h$, where λ_l and λ_h are the shadow prices for the lower and higher budget, respectively, then we would also have that $G_h^{-1}(z) \leq G_l^{-1}(z)$ for all z (see the proof of Proposition 8). But then the budget constraint would not hold with equality for the higher budget. Thus $\lambda_h < \lambda_l$, so $G_h^{-1}(z) \geq G_l^{-1}(z)$ for all z (again, see the proof of Proposition 8), which is a FOSD shift in the optimal prize distribution.

Proof of Proposition 10. Choose some prize distribution G. By looking at the areas below the graphs of L and L^{-1} in the square $[0,x]\times[0,L(x)]$, we obtain that the cost of the performance of type x in the mechanism that implements the assortative allocation satisfies

$$xL(x)-\int_0^x L(\widetilde{x})\,d\widetilde{x}=\int_0^{L(x)} L^{-1}(u)\,du.$$

Thus, the average performance (6) is equal to

$$\int_0^1\left(c^{-1}\left(\int_0^{L(x)} L^{-1}(u)\,du\right)\right)f(x)\,dx=\int_0^1\left(c^{-1}\left(\int_0^{h\left(G^{-1}(z)\right)} L^{-1}(r)\,dr\right)\right)dz, \qquad (31)$$

where the equality follows from the change of variables $z=F(x)$ and the identity $L\left(F^{-1}(z)\right)=h\left(G^{-1}(z)\right)$.[52] Since a FOSD shift in F decreases F and therefore L pointwise, it increases L^{-1} pointwise, and therefore increases (31).

Deriving the optimal prize distribution for the example in Section 8. Proposition 6 shows that $z_{\min}<z_{\max}$ and G may have atoms only at 0 and m. We now use the conditions in Case 1 of Lemma 2 to derive G. Define an auxiliary function $q(z)=(c^{-1})'(\widetilde{L}(z))$, plug $q(z)$ into (19), and differentiate with respect to z to obtain the differential equation $q\prime(z)z+2q(z)=0$ for $q(z)$.[53] Solving this equation, and substituting back into (19), we obtain $(c^{-1})'(\widetilde{L}(z))=\lambda/z^2$. By the definition (15) of $\widetilde{L}(z)$, we obtain $\left((c^{-1})'\right)^{-1}(\lambda/z^2)=zG^{-1}(z)-\int_0^z G^{-1}(t)\,dt$. If G^{-1} is differentiable, differentiating the last equality gives $(G^{-1})'(z)=(-2\lambda/z^4)(((c^{-1})')^{-1})'(\lambda/z^2)$.[54]

Since $c^{-1}(z)=\sqrt{z}$, we have $(c^{-1})'(z)=1/(2\sqrt{z})$, $\left((c^{-1})'\right)^{-1}(z)=1/(4z^2)$, and $\left(\left((c^{-1})'\right)^{-1}\right)'(z)$ $=-1/(2z^3)$. Thus, $G^{-1}(z)=z^3/(3\lambda^2)+y_{\min}$, where $y_{\min}$ is the "lowest prize" awarded. By parts 2 and 3 of Proposition 6, $z_{\min}=0$ and $y_{\min}=0$.

[52]Even though L^{-1} may be discontinuous, because G^{-1} may be discontinuous, it is monotonic, so the change of variables applies.

[53]The solution can be verified to be differentiable.

[54]We will show that an optimal prize distribution G with differentiable inverse G^{-1} exists. No other prize distribution will lead to higher average performance, since the average performance corresponding to any prize distribution can be approximated arbitrarily closely by the average performance corresponding to a prize distribution with a differentiable inverse.

Consider first $m \geq 4Y$. Suppose that $z_{\max} = 1$ and the entire budget is used. Substituting the expression for $G^{-1}(z)$ into the budget constraint with equality, we obtain $\lambda = 1/\sqrt{12Y}$, which gives $G^{-1}(z) = 4z^3 Y$. Thus, G^{-1} does not exceed $m \geq 4Y$. Substituting G^{-1} into the target function, the average performance is $\sqrt{Y/3}$, which increase in the budget Y, so it is indeed optimal to use the entire budget. Moreover, we cannot have $z_{\max} < 1$, because the budget constraint would be violated: on $[z_{\max}, 1]$ the prize would be m, higher than with $z_{\max} = 1$, and in order to have $G^{-1}(z_{\max}) = m$, the value of λ would have to be lower than that with $z_{\max} = 1$, which implies a pointwise higher value of G^{-1} on $[0, z_{\max}]$ than with $z_{\max} = 1$.

Now suppose that $m < 4Y$ and the entire budget is used. Then, we still have $G^{-1}(z) = z^3/\left(3\lambda^2\right)$ for $z \leq z_{\max}$, but this new λ is different from that for $m = 4Y$. (Otherwise, since $G^{-1}(z) = 4Y$ at $z = 1$ for the old λ, the entire budget would not be used.) This implies that $z_{\max} < 1$. Since the budget constraint is satisfied with equality, $\lambda = z_{\max}^2/(12(Y - m(1 - z_{\max})))^{1/2}$. Substituting this λ into $m = G^{-1}(z_{\max}) = z_{\max}^3/\left(3\lambda^2\right)$ gives that $z_{\max} = 4(m - Y)/(3m)$. Substituting the expression for $z_{\max}$ into the expression for λ, and substituting the resulting expression for λ into the expression for G^{-1} for $z \leq z_{\max}$, gives $G^{-1}(z) = 27z^3 m^4/\left(64(m - Y)^3\right)$. Substituting this G^{-1} into the target function, the average performance is $\sqrt{m - Y}\,(1 - 8(m - Y)/(9m))$.

This expression increases for Y in $[m/4, 5m/8]$, and decreases for Y in $[5m/8, m]$.[55] Therefore, this expression is the maximal average performance for Y in $[m/4, 5m/8]$. Any budget in excess of $5m/8$ will optimally not be used.

[55] No budget above m will ever be used, since m is the cost of awarding all types the highest possible prize.

12 Online Appendix

In this appendix, we discuss the solution to (4) for linear costs in the case when Assumption 1 is not satisfied. We will provide only heuristic, informal arguments, but the reader will see that making the arguments rigorous should not encounter any major difficulty.

Suppose that the range of z's is divided into a large number of small (infinitesimal) intervals, on which we will increase the value of G^{-1} progressively by small (infinitesimal) moves of a size Δ, until we exhaust the entire budget given by (5). By raising G^{-1} on such an interval by Δ, we increase, up to a first-order approximation, the objective function by $h'(\cdot)k(\cdot)\Delta$, where the value of k is taken at any point from the interval, and the value of h' is taken at the current value of G^{-1} on this interval. However, in order to maintain the monotonicity of G^{-1} when we increase its value on an interval I, we must increase its value also on all intervals I' higher than I on which the value of G^{-1} is the same as that on I. Thus, we always want to increase the value of G^{-1} on an interval I such that the average value of $h'(\cdot)k(\cdot)$ across all intervals I' is the highest.

When k is increasing, as assumed in the main text, the average value of $h'(\cdot)k(\cdot)$ is initially the highest for the highest interval. So, we begin with raising G^{-1} on the highest interval by Δ. When, in addition, h' is increasing, this makes the average value of $h'(\cdot)k(\cdot)$ even higher for the highest interval, without affecting the average values for the other intervals.[56] So, we raise G^{-1} on the highest interval until we reach its bound of m. Next, we raise G^{-1} on the second highest interval, and we continue in this manner until we exhaust the budget. This yields Proposition 3.

When, in turn, h' is decreasing (assume strictly for the sake of our argument), raising G^{-1} makes the average value of $h'(\cdot)k(\cdot)$ lower for the highest interval. This ultimately makes the average highest for the second highest interval. (Notice that the average on the second-highest interval will be equal just to the value on that interval, after we first raise the value on the highest interval.) And we will then begin raising G^{-1} on the second-highest interval. We will never raise G^{-1} on the second-highest interval to a value strictly higher than that on

[56] Actually the first jump even decreases the average on lower intervals, since it makes the highest interval no longer count in the average.

the highest interval, because k is increasing. Ultimately, we will make the average highest on the third-highest interval. We will not stop until we reach the lowest interval, because any nontrivial increase in G^{-1} on an interval reduces $h'(\cdot)$ by more than enough to offset the just slightly lower value of $k(\cdot)$ on the adjacent lower interval.[57] This yields Proposition 4.

These arguments imply a number of claims even when k is not increasing.

Claim 1. If h is weakly convex, so players are risk neutral or risk loving, then the optimal prize distribution consists of: (1) a mass of prize m and a mass of prize 0; or (2) a mass of prize m, a mass of some intermediate prize $y \in (0, m)$, and a mass of prize 0.

Indeed, notice that even though k is no longer increasing, $h'(\cdot)k(\cdot)$ still takes the highest value on the highest interval I, since $k(1) = 1$ is the highest possible value of k. Raising G^{-1} increases the value of $h'(\cdot)k(\cdot)$ on the highest interval I, without affecting the value on the intervals lower than I. So, we keep raising G^{-1} on the highest I until reach the bound of m. And then we go to the second highest interval. However, as k is no longer increasing, we may at some moment happen to go to an interval lower than the next highest one; and then we may exhaust the remaining budget before reaching the bound of m on that interval I and all higher intervals on which the value of G^{-1} is the same as that on I. If this happens, the optimal prize distribution has the form as described in (2).

Corollary 2 *If h is weakly convex, and m is sufficiently large, then the optimal prize distribution consists of a mass of prize m and a mass of prize 0.*

Indeed, the previous argument implies that the budget is exhausted at the highest intervals.

Claim 2. If h is strictly concave, so players are risk averse, then any optimal prize distribution assigns a positive mass to the set of intermediate prizes $(0, m)$, and any optimal prize distribution awards all prizes up to the highest prize awarded. That is, G strictly increases on $[0, G^{-1}(1)]$.

[57] Of course, the continuity of k is essential here.

As before, we first raise G^{-1} on the highest interval I. Raising G^{-1} makes the value of $h'(\cdot)k(\cdot)$ on I lower. This ultimately makes the value of $h'(\cdot)k(\cdot)$ higher for the second highest interval. And we then begin raising G^{-1} on the second highest interval, and continue going down to lower and lower intervals. We will not stop until we exhaust the budget or run out of intervals on which $k(\cdot)$ is positive, because any nontrivial increase in G^{-1} on any interval reduces $h'(\cdot)$ by more than enough to offset a slightly lower value of $k(\cdot)$ on the adjacent highest interval.

Notice, however, that we may make discrete jumps on the way down to lower and lower intervals when $k(\cdot)$ takes higher values on lower intervals. This implies that even if h is strictly concave, an optimal prize distribution may have atoms at prizes other than 0 and m. Finally, it can be readily checked that Claim 2, except the statement that G strictly increases on $[0, G^{-1}(1)]$, holds true even when h is only weakly concave but not linear on $[0, m]$.

References

[1] **Barut, Yasar, and Dan Kovenock.** 1998. "The Symmetric Multiple Prize All-Pay Auction with Complete Information." *European Journal of Political Economy*: 14(4): 627-644.

[2] **Billingsley, Patrick.** 1995. "Probability and Measure." John Wiley and Sons, Inc.

[3] **Bulow, Jeremy I., and Jonathan Levin.** 2006. "Matching and Price Competition." *American Economic Review*, 96(3): 652-68.

[4] **Bulow, Jeremy I., and John Roberts.** 1989. "The Simple Economics of Optimal Auctions." *Journal of Political Economy*, 97(5): 1060-1090.

[5] **Clark, Derek J., and Christian Riis.** 1998. "Competition over More than One Prize." *American Economic Review*, 88(1): 276-289.

[6] **Glazer, Amihai and Refael Hassin.** 1988. "Optimal Contests." *Economic Inquiry*, 26(1): 133-143.

[7] **Hillman, Arye L., and John G. Riley.** 1989. "Politically Contestable Rents and Transfers." *Economics and Politics*, 1(1): 17–39.

[8] **Immorlica, Nicole, Greg Stoddard, and Vasilis Syrgkanis.** 2015. "Social status and badge design." in *Proceedings of the 24th International Conference on World Wide Web*, 473-483. ACM.

[9] **Kaplan Todd R. and Shmuel Zamir.** 2015. "Advances in Auctions," in Handbook of Game Theory, ed. by H. P. Young, and S. Zamir, vol. 4, pp. 949–984. North-Holland.

[10] **Moldovanu, Benny, and Aner Sela.** 2001. "The Optimal Allocation of Prizes in Contests." *American Economic Review*, 91(3): 542-58.

[11] **Moldovanu, Benny, Aner Sela, and Xianwen Shi.** 2007. "Contests for status." *Journal of Political Economy*, 115(2):338–363.

[12] **Myerson, Roger B.** 1981. "Optimal Auction Design." *Mathematics of Operations Research*, 6(1): 58-73.

[13] **Olszewski, Wojciech, and Ron Siegel.** 2016. "Large Contests." *Econometrica*, 84(2): 835-854.

[14] **Reny, Philip J.** 1999. "On the Existence of Pure and Mixed Strategy Nash Equilibria in Discontinuous Games." *Econometrica*, 67(5): 1029-1056.

[15] **Sethi, Suresh P., and Gerald L. Thompson**. 2000. "Optimal Control Theory, Applications to Management Science and Economics." Springer.

[16] **Siegel, Ron.** 2010. "Asymmetric Contests with Conditional Investments." *American Economic Review*, 100(5): 2230-2260.

[17] **Spence, Michael A.** (1973). "Job Market Signaling." *Quarterly Journal of Economics*, 87: 355-374.

[18] **Szymanski, Stefan.** (2003). "Job Market Signaling." *Journal of Economic Literature*, 41(4): 1137-1187.

[19] **Xiao, Jun.** 2016. "Asymmetric All-Pay Contests with Heterogeneous Prizes." *Journal of Economic Theory*, 163: 178-221.