Optimal Timing of Selection Contests

Dmitry Lubensky  
Indiana University

Michael Stevens  
US Department of the Treasury*

January 28, 2015

Abstract

A principal uses a contest to select one agent from a group to perform a future task. In choosing the timing of the contest the principal faces a tradeoff: agents’ types evolve and thus a later contest is more accurate, however, an agent’s effort in the contest diminishes her task performance and the less time until the task, the costlier the effort. We explore the optimal timing decision and show that later contests are better whenever types are more disperse and whenever ability at the task gives agents an edge in the contest. We find that the expected performance of the selected agent is invariant to the average ability in the group and can either increase or decrease with the number of participants. We also recast the timing decision as a more general screening problem in which the principal suffers the signaling costs of the chosen agent. We argue that the optimal screening mechanism is equivalent to that in Chakravarty and Kaplan (2013), is stochastic, and favors agents with low hazard ratios.

*The views expressed in this paper are those of the authors and do not necessarily reflect the policy of the U.S. Department of Treasury.
1 Introduction

Going into the 2008 Beijing Olympics hopes were particularly high for American swimmers Michael Phelps and Katie Hoff, both world record holders and slated to compete in numerous events. Yet at the Olympics while Phelps met and exceeded expectations Hoff under-performed. Why the disparity in the two athletes’ performances? Putting aside explanations relying on swimmer-specific idiosyncracies, their performances can be understood by considering the athletes’ strategic environment. Swimmers can peak only for a short period and they time their peak for a particular date. Each could peak closer to the trials to improve their chances of making the team but, by doing so hurt their Olympic performance. Katie Hoff faced stiff domestic competition and was forced to peak closer to the trials; Michael Phelps maintained a comfortable lead over his American rivals and was able to peak closer to the Olympics.\(^1\) For USA Swimming, Katie Hoff’s mediocre performance was avoidable: had they chosen a significantly earlier date for the trials, she would have had the opportunity to recover and peak again at the Olympics.\(^2\) But holding the trials early would have also come at a cost, since the best swimmer at the trials might no longer still be the best swimmer at the Olympics. In choosing the optimal time to hold the trials, USA Swimming had to balance the accuracy of their selection with the cost of not allowing swimmers sufficient time to recover.

This problem facing USA Swimming is one faced by many organizations: how to time the selection of an agent when waiting longer makes the selection more accurate but also more costly. Take as another example a political party choosing a candidate for a general election. Candidates attempt to position themselves on the political spectrum to match the views of their party’s median voter, but in order to win the general election they would be best off at the median of the entire population. Given that shifting one’s political platform takes time, holding a later primary hurts the chances of the eventual primary winner. At the same time, a lengthy primary election can reveal characteristics about a candidate which are independent of his or her platform but are important in determining electability in the general election.\(^3\) In this sense a later primary ensures that the winner is the most likely of the group to win the general election. Party organizers must weigh the accuracy benefits of a later primary with the costs of a running a more extreme candidate in the general election.\(^4\)

A similar tradeoff is also found in the workplace. A manager must select which of several analysts to promote to associate. Delaying the promotion decision allows the manager

\(^{1}\)In each event USA Swimming selects the top two performers. Going into the Olympic trials, Michael Phelps’ average lead over the eventual third place finishers was roughly twice that of Katie Hoff, in percentile terms.

\(^{2}\)In fact Australia, another traditional swimming power, held their trials a full year in advance.

\(^{3}\)For example, if candidate is a particularly inspiring public speaker or is has damaging secrets from his or her past.

\(^{4}\)The timing of the primary can be broadly understood as the time at which the party selects a winner, which often happens prior to the party’s convention.
to ascertain which of his analysts is most highly skilled. However, the longer the promotion decision is delayed, the more time is spent by the eventual winner performing the tasks of analyst instead of the more productive tasks of an associate.

In all of these examples, a principal faces a group of agents whose types he does not know and must choose one to perform a task. To make this decision, the principal chooses a time at which all agents compete and uses the results of this competition to make the selection. Agents divert resources from their final task to improve their performance in the competition and the principal must keep this in mind when choosing the timing optimally. The aim of this paper is to explore this timing decision and to ascertain which factors induce the principal to choose a later more accurate contest and which favor an earlier and noisier selection.

We take two approaches to address this question. First we note that from the point of view of the principal, the choice of timing is indirectly a choice of an allocation rule and a corresponding incentive compatible transfer function. That is, for any given selection time, agents choose how many resources to divert to the contest from their final performance, i.e. their transfer, and in an equilibrium conditional on these actions agents will have some probability of winning, i.e. the allocation rule. The choice of timing can be thought of as a blunt tool in the more general problem of choosing an optimal incentive-compatible screening mechanism, and our first approach is to solve that more general problem. Specifically, we consider a setting in which agents are privately informed about their types and simultaneously send costly signals to the principal, who in turn commits to an allocation rule as a function of the signals. The principal’s payoff is the performance of his chosen agent, hence he wants to choose the best agent but suffers that agent’s signaling cost. We show that ex-ante the principal prefers to allocate not necessarily to agents with the highest type but to those with the lowest hazard ratio, roughly speaking those agents facing the least competition. The optimal mechanism is stochastic, that is for some collections of signals the principal will allocate using a lottery between some or all of the agents. While this may seem suboptimal ex-post, a stochastic mechanism lessens the returns to costly signaling and reduces competition ex-ante. Thus, it is best to include some noise in the selection process.

The stochastic optimal allocation rule relies on commitment by the principal and may be difficult to implement in practice. Taken at face value, the mechanism requires the principal to sometimes not select the best agent despite having perfectly inferred all participants’ types through their signals. The principal cannot always commit to this behavior, just as, for instance, a party chairman may not find it feasible to flip a coin between two candidates when one has received significantly more votes. On the other hand, the party chairman is able to commit to an early primary and this can indirectly infuse a stochastic component into the decision process. Given that a candidate’s types evolves over time, the person chosen early is not necessarily the best one at the time of the general election.\footnote{Or more precisely, if an agent’s type evolves so that it is best at the time of the final task, she may not have necessarily been chosen during the selection contest.}
Restricting the principal to making only a timing decision then reduces the set of allocation rules that he can implement, and in our second approach we look for the optimum in this reduced set. We consider a continuous time environment in which agents privately observe the evolution of their types over time and choose an effort level for the selection contest that comes at the expense of their final performance. We assume that the cost of effort decreases as the time between the selection and the final task grows so that agents are better able to recover given more time. In choosing a later selection time, the principal benefits from the option value of picking the agent with the best shocks but potentially pays the price in higher effort costs. We show that if agents’ ability to divert resources from the final task is unrestricted, agents divert the same amount of resources from the final task regardless of the timing of the contest. That is, as there is less time to recover between the selection and the final task and effort becomes costlier, all agents adjust their effort to exactly offset this. Hence, it is always optimal to select agents as late as possible.

This result sheds some light on the underlying mechanism and we treat it as a benchmark case. However, the assumption that agents are unrestricted in how many resources they can divert is strong and violated in many applications. For instance in swimming, even if athletes peak for the trials, they can once again peak for the Olympics given enough time to recover. By choosing a trials date sufficiently early the principal can thus restrict the amount of resources participants can divert. To capture this, we consider a simplified timing model in which the principal chooses between an early and a late contest. The early contest is sufficiently early so that agents will fully recover; hence, they are unable to divert resources from their final performance. The late contest is one in which an agent’s ability to divert resources is unconstrained. We assume that all agents start with the same type at the early contest and since types evolve, each agent’s type is expected to be distributed according to some function $F$ at the time of the late contest. We show that the principal’s payoff to holding a late contest is independent of the mean of $F$ but grows in the dispersion of $F$. We also show that the effect of adding more agents to the late contest is ambiguous as there is a tradeoff between choosing a type from a higher order statistic but also inducing more competition. In addition, we consider selection contests of skill in which higher types have a natural advantage and prove this improves the returns to a later contest.

This paper is concerned with the question of how to time a selection contest, and answering this question overlaps with several literatures. A selection contest is in essence a screening mechanism, and as such our setting is strategically similar to those in the seminal works of Spence (1973) and Myerson (1981), as well as in the myriad papers spawned thereafter. Aside from establishing a general framework, Myerson (1981) also provides several techniques that we adopt to characterize equilibria. A point of departure from these classic works is that the designer can suffer the costs (or transfers) along with the agents. This idea is present in more recent work such as Hartline and Roughgarden (2008) and Chakravarty and Kaplan (2013). Both papers consider a setting in which agents’ efforts are wasteful to the principal, with
the latter explicitly characterizing the optimal mechanism for a benevolent social planner that wants to assign a good but suffers the sum of all the signaling costs of the agents. Our paper is motivated by a selfish designer who cares only about the costs expended by the agent that he chooses, yet we show that the two problems are strategically equivalent and the same mechanism maximizes both objectives. We re-interpret Chakravarty and Kaplan’s result in the context of competition, showing that the optimal allocation rule assigns the good to those who have to compete the least. We also differ in spirit from Chakravarty and Kaplan (2013) in addressing the issues of commitment and implementation by focusing on the timing question in particular.

In the part of the paper that deals with the timing model, we ask questions that are often addressed in the field of tournament design, such as how to assign a winner (Lazear and Rosen (1981), Moldovanu and Sela (2001)) and how many participants to include (Fullerton and McAfee (1999)). The answers to these questions are quite different in our setting in which a principal wishes to minimize effort as compared to a tournament design setting in which the principal wants to maximize effort.

The rest of the paper proceeds with Section 2, in which we present a model of costly screening. There we establish the equivalence to the general solution from Chakravarty and Kaplan (2013) and offer an interpretation for our particular application. In Section 3 we drop the assumption that a principal can commit to an allocation rule and explicitly model a principal that must choose a time for a selection contest. Then Section 4 concludes.

2 A Model of Costly Screening

We consider a setting in which a principal must choose one among \( N \) agents to perform a task. Each agent \( i \) has type \( \theta_i \geq 0 \) which describes her ability to perform the task. Types are drawn privately and independently from distribution \( F(\cdot) \) and are not observed by the principal. An agent \( i \) can signal her type by taking a publicly observable costly action \( e_i \geq 0 \). The cost of this action is that it diverts resources away from the agents’ performance of the task conditional on being selected. For instance, \( e_i \) can represent how far a candidate in a primary election moves her platform away from that of the population’s median voter; the cost of this move is a reduced chance of winning the general election and is only borne if the candidate wins the primary. An agent’s performance conditional on being chosen is

\[
y_i = \theta_i - e_i.
\]

Agents choose \( e_i \) simultaneously and the principal can commit to a mechanism

\[
M : e \rightarrow x
\]

in which the principal maps each action profile \( e = \{e_1, ..., e_N\} \) into a profile of allocations \( x(e) = \{x_1(e), ..., x_N(e)\} \), where \( x_i(e) \) is the probability that agent \( i \) is selected. We refer to the vector \( x^M(e) \) as the allocation rule under mechanism \( M \). We explicitly allow
for the allocation mechanism to be stochastic since a central question of our analysis is whether a stochastic mechanism is optimal. A mechanism $M$ is feasible if for every profile $e$, $\sum_i x_i^M(e) \leq 1$.

Agent $i$’s payoff is her performance on the task, and her von Neumann-Morganstern expected utility function is

$$u_i(x_i(e), e_i, \theta_i) = x_i(e)(\theta_i - e_i).$$

Recall that agent $i$ only suffers the cost of $e_i$ if she is chosen to perform the task.

The principal’s payoff is his chosen agent’s performance, hence his von Neumann-Morganstern utility is

$$u_0(x(e), e, \theta) = \sum_i x_i(e)(\theta_i - e_i) = \sum_i u_i(x_i(e), e_i, \theta_i).$$

An equilibrium is characterized by two conditions: expected utility maximizing behavior of each agent conditional on mechanism $M^*$

$$e^*_i(\theta_i, M^*) = \arg \max_{e_i} E_{e_{-i}}[x_i^{M^*}(e_i, e_{-i})(\theta_i - e_i)],$$

and the principal’s optimal choice of mechanism $M^*$

$$M^* = \arg \max_M E_{\theta} \left[ \sum_i x_i^M(e^*(\theta, M))(\theta_i - e^*_i(\theta_i, M)) \right]$$

$$= \arg \max_M E_{\theta} \left[ \sum_i u_i(\theta, M) \right].$$

One way to think about the optimal allocation mechanism $M^*$ is as a tradeoff between choosing accurately and inducing costly signaling. As extreme examples, an allocation rule $x_i(e) = \frac{1}{N}$ would choose randomly but induce $e_i = 0$ by all participants while $x_i(e) = \mathbb{I}\{e_i = \max\{e_1, ..., e_N\}\}$ would always select an agent with the highest action but would induce high signaling costs. As we will show, the optimal mechanism uses either the latter of these two allocation rules or is a lottery amongst those with the highest effort levels, depending on vector $e$.

We use a mechanism design approach to characterize $M^*$, and we do so in two steps. First, we invoke the revelation principle to show that the allocation rule $x^{M^*}$ is the same as an allocation rule $x^{DM^*}$ which maximizes the sum of the expected utilities of the agents in a direct revelation mechanism. Then, we use a result in Chakravarty and Kaplan (2013) which characterizes $x^{DM^*}$.
From Myerson (1981), we know that if \((x^M(e^*(\theta, M)), e^*(\theta, M))\) is an equilibrium in our setting then in a direct revelation game \((x^{DM}(\theta), \tau(\theta, x^{DM}))\) is a truth-telling equilibrium if \(x^{DM}(\theta) = x^M(e^*(\theta, M))\) and \(\tau(\theta, x^{DM}) = e^*(\theta, M)\), that is

\[
x^{M^*}(e) = \arg\max_{x^M} E_\theta \left[ \sum_i x_i^M(e^*_i(\theta, x^M))(\theta_i - e^*_i(\theta_i, x^M)) \right]
\]

\[
= \arg\max_{x^{DM}} E_\theta \left[ \sum_i x_i^{DM}(\theta)(\theta_i - \tau_i(\theta, x^{DM})) \right]
\]

\[
= \arg\max_{x^{DM}} E_\theta \left[ \sum_i u_i(\theta, x^{DM}) \right]
\]

\(= x^{DM^*}(\theta)\).

The revelation principle is invoked in going from the first to second line, equating choosing an allocation \(x^M\) conditional on agents playing equilibrium \(e^*(\theta, x^M)\) to choosing allocation \(x^{DM}\) given that “transfers” \(\tau(\theta, x^{DM})\) are constructed so that truth-telling is optimal.

**Proposition 1** (From Chakravarty and Kaplan (2013), Proposition 1)

Let \(x^{DM^*}(\theta)\) be the allocation rule in a direct revelation mechanism that maximizes the sum of participants’ expected utilities, as in (3). Let \(\theta_M = \max\{\theta_1, \ldots, \theta_N\}\). There exists a family of disjoint intervals \(I_k = [\theta_k, \theta_{k+}]\), \(k = 1, \ldots, K\) such that

\[
\text{if } \theta_M \in I_k \text{ for some } k \text{ then } \quad x_i(\theta) = \begin{cases} 
1/\#(\theta_j \in I_k) & \text{if } \theta_i \in I_k \\
0 & \text{otherwise}
\end{cases}
\]

\[
\text{else if } \theta_M \notin I_k \text{ for any } k \text{ then } \quad x_i(\theta) = \begin{cases} 
1/\#(\theta_j = \theta_M) & \text{if } \theta_i = \theta_M \\
0 & \text{otherwise}
\end{cases}
\]

We refer the interested reader to Chakravarty and Kaplan (2013) for a characterization of the intervals \(I_k\) along with a proof of the proposition. The authors call intervals \(I_k\) the “lottery” regions and the rest of the type space as the “contest” regions. Whenever the highest type is in a lottery region, the allocation is a random assignment among all others in the region.

In what follows we provide some intuition for why such an allocation rule is optimal with an example.

Consider a setting with \(N = 2\) agents and assume that the lowest type is \(\underline{\theta} = 0\). Using the law of iterated expectations, we can rewrite the principal’s objective in the direct mechanism from equation (3)

\[
E_\theta \left[ \sum_i u_i(\theta, x) \right] = 2 \int_{\theta_i} E_{\theta_j} \left[ u_i(\theta_i, x_i) \right] f(\theta_i) d\theta_i \tag{4}
\]
In other words, the principal wants to maximize the ex-ante expected utility of a participant.

Let \( X(\theta_i) \) denote the ex ante probability that an agent of type \( \theta_i \) is allocated the object. Following Myerson (1981), an allocation rule \( x(\theta) \) can be obtained in equilibrium if it satisfies the following four constraints:

\[
E_{\theta_j}[u_i(\theta_i, x_i)] = E_{\theta_j}[u_i(\theta, x_i)] + \int_{\theta}^{\theta_i} X(\hat{\theta}) d\hat{\theta} \quad (5)
\]

\( X(\theta_i) \) non-decreasing \quad (6)

\[
E_{\theta_i}[u_i(\theta_i, x_i)] \geq 0 \quad (7)
\]

\[
\sum_i x_i(\theta) \in [0, 1] \quad \text{and} \quad 0 \leq x_i \leq 1 \quad (8)
\]

Conditions (5) and (6) are derived from first and second order conditions for truthful reporting, respectively. Condition (7) is a participation constraint and (8) is a feasibility constraint.

To solve the constrained optimization, we solve a relaxed program by ignoring (6), and then use a technique from Myerson (1981) called ironing to apply that constraint. Plugging (5) into the objective function (4), we obtain

\[
\max_x 2E_{\theta_j}[u_i(\theta, x_i)] + 2 \int_{\theta}^{\theta_i} X(\hat{\theta}) f(\theta_i) d\theta_i
\]

\[
\Leftrightarrow \max_x 2 \int_{\theta_i}^{\theta_i} X(\theta_i)(1 - F(\theta_i)) \ d\theta_i
\]

\[
\Leftrightarrow \max_x 2 \int_{\theta_i}^{\theta_i} \left( \int_{\theta_j}^{\theta_i} x_i(\theta_i, \theta_j) f(\theta_j) \ d\theta_j \right) (1 - F(\theta_i)) \ d\theta_i
\]

The second line follows because, with non-negative transfers, type \( \theta = 0 \) gets a utility of at most zero, and, by the participation constraint, must get exactly zero. Next we claim without proof that (8) binds so that \( x_1(\theta_1, \theta_2) + x_2(\theta_1, \theta_2) = 1 \) and that we can restrict attention to symmetric allocations, so \( x_1(a, b) = x_2(b, a) \). This allows us to rewrite the objective as

\[
\max_x 2 \int_{\theta_i}^{\theta_i} x_i(\theta_i, \theta_j)(1 - F(\theta_i) + (1 - x_i(\theta_i, \theta_j)) f(\theta_i) (1 - F(\theta_i)) \ d\theta_j \ d\theta_i
\]

Expression (10) can be maximized pointwise, and by inspection the optimal allocation rule must satisfy

\[
x_1(\theta_1, \theta_2) = \begin{cases} 
1 & \text{if } \frac{1-F(\theta_1)}{f(\theta_1)} \geq \frac{1-F(\theta_2)}{f(\theta_2)} \\
0 & \text{otherwise}
\end{cases}
\]

(11)
In the mechanism design literature, \( \theta_i \) is said to have a “virtual type” given by the inverse hazard ratio \( \frac{1-F(\theta_i)}{f(\theta_i)} \). The hazard ratio is a measure of how competitive the neighborhood is around a particular type, and the principal wants to allocate probability to types that will not be forced to compete very hard.\(^6\)

This allocation rule will generally not satisfy condition (6), as higher types will tend to have lower virtual types. For instance, if \( \theta_i \) is distributed uniformly on \([0, 1]\) then (11) results in interim probabilities given by \( X(\theta_i) = 1 - \theta_i \), which is everywhere decreasing. To account for this, one must use the ironing algorithm outlined in Myerson (1981). The basic idea is that over regions in which (11) causes \( X(\theta_i) \) to be decreasing, one must increase the allocation to higher types at the expense of lower types, and the optimal way to do so is to identify intervals such that whenever two types are from the same interval their allocation is decided by a random lottery. Thus, the optimal screening mechanism is noisy and sometimes does not select the agent with the highest realized type, despite types being fully revealed in equilibrium.

### 3 Timing Models

A key assumption in Section 2 is that the principal commits to the allocation rule and in many applications this assumption is difficult to defend. For example, imagine a party chairman flipping a coin between two candidates despite one receiving more votes than the other, or holding the US Olympic trials and then rolling a die to determine which of the top six performers goes to the Olympics. Aside from having incentive to choose the revealed-best agent, there is also institutional pressure to do so.

Yet while the principal is pressured to choose the winner of the selection contest, he can make ex-ante decisions about the design of the contest so that the winner is not necessarily the highest type. We consider a setting in which the decision is about \textit{when} to hold the selection contest. Agents have types that evolve stochastically over time, hence an earlier selection makes it less likely that the chosen agent will be best at the time of the task. However agents can also divert resources toward the selection contest, and the less time between selection and the task the costlier that diversion. The tradeoff in the timing decision is analogous to that in the previous section: accurate selection versus inducing harmful effort. In essence, the timing structure now restricts the set of available stochastic allocation rules and we look for the “constrained-best” optimum within this set.

\(^6\)This is in contrast to a principal who maximizes revenue and defines virtual types by \( \theta_i - \frac{1-F(\theta_i)}{f(\theta_i)} \). This principal prefers agents with high hazard ratios so that they compete away their information rents.
3.1 A Model in Continuous Time

A principal commits to a time $t \in [0, 1]$ in which to hold a selection contest. There are $N$ agents and all have type $\theta_0$ at time $t = 0$. Each agent’s type evolves independently and stochastically over time such that for any $t > t'$, $\theta_i(t')$ is a mean preserving spread of $\theta_i(t')$. Let $F(\theta_i|t)$ denote the distribution of $\theta_i(t)$ and assume that for any $t$ the support has finite and strictly positive upper and lower bounds.\(^7\) As an example, one may think of $\ln(\theta_i(t))$ following Brownian motion, though many other stochastic processes would also be admissible.

Each agent privately observes the evolution of her type and at time $t$ chooses action $e_i \in [0, \infty]$, to which we refer as effort. The agent that wins the selection contest is the one with the highest score $s(e_i, \theta_i)$. For now, assume the score strictly increases in effort and does not directly depend on type\(^8\) so that

$$s(e_i, \theta_i) = s(e_i).$$

Each agent’s action $e_i$ comes at a cost $c(e_i, t)$ to the agent’s final task performance

$$y_i = \theta_i(1) - c(e_i, t).$$

Note that an agent’s final task performance depends on her type in period $t = 1$. However by assumption $E[\theta_i(1)|\theta_i(t)] = \theta_i(t)$, hence from the risk-neutral perspectives of both the agent and the principal, choosing type $\theta_i$ at time $t$ has the same payoff as choosing type $\theta_i$ at time $t = 1$.

The cost function $c(e_i, t)$ is weakly increasing and continuous in effort and in time. Agents that choose $e_i = 0$ divert no resources from the final task, thus $c(0, t) = 0$ for any $t$. The set of feasible costs at any time is $[0, \bar{c}(t)]$, where $\bar{c}(t) \equiv \lim_{e \to \infty} c(e, t)$. Note that since $c_e$ and $c_t$ are both weakly positive, it must be that $\bar{c}(t)$ is weakly increasing as well. As before, agent $i$’s expected utility is $u_i = x_iy_i$ and the principal’s expected utility is $u_0 = \sum_i x_iy_i$.

For a given selection time $t$, agents solve

$$\max_{e_i} \Pr \left( s_i(e_i) > \max_{j \neq i} s_j \right) \left( \theta_i - c(e_i, t) \right) \quad \Leftrightarrow \quad \max_{e_i \in [0, \bar{c}(t)]} \Pr \left( s_i(e_i(c_i, t)) > \max_{j \neq i} s_j \right) \left( \theta_i - c_i \right)$$

\(^7\)This is an assumption made for mathematical convenience. In our setup, an agent with a negative type would choose not to participate and to compute equilibrium strategies would require accounting both for the distribution of types and the number of participants.

\(^8\)An agent’s score in the selection contest can depend both on her effort and her type. For example, in an Olympic trial a swimmer’s chance of winning depends not only on how close she is to her peak but also on her ability. The assumption here is made to match the setup in the previous section; later we consider how relaxing this assumption affects the timing decision.
The selection contest can then be interpreted as a first price auction in which each agent chooses bid function \( c(\theta_i, t) \) bounded by the budget constraint \( \bar{c}(t) \). The principal anticipates expected equilibrium behavior for each time \( t \) and solves

\[
\max_t \ E_{F(\theta_i, t)} \sum_i \left( \theta_i(t) - c(\theta_i, t) \right)
\]

The timing choice is then a choice of a distribution of bidders \( F(\theta_i, t) \) along with their corresponding bidding strategies \( c(\theta_i, t) \). A later selection contest has the benefit of sampling from a distribution with higher variance but can also be associated with higher effort costs.

We now examine this tradeoff in more detail.

### 3.1.1 Contests in Which \( \bar{c}(t) \) Does Not Bind

First as a benchmark we examine the optimal timing decision when at every \( t \), \( \bar{c}(t) \) is large enough so that it does not bind.\(^9\) In this situation we will show that in the timing decision there is no tradeoff for the principal. In holding a later contest the effort cost of the winner remains constant while the benefit of the option value of waiting continues to accrue. Thus, waiting until the end is optimal.

**Proposition 2** Suppose \( \bar{c}(t) \) is sufficiently large at every \( t \) so as not to bind. Then the optimal selection time is \( t = 1 \).

**Proof Sketch**

Given that \( \bar{c}(t) \) does not bind it must be that for any realization of types \((\theta_1(t), ..., \theta_N(t))\) the highest type wins for sure. This implies that type \( \theta_i(t) \) has an interim probability of winning \( X(\theta_i(t)) = F^{N-1}(\theta_i, t) \). Using the analog of expression (9) for \( N \) agents, we can write the principal’s payoff as

\[
u_0(t) = N \int_{\theta_i} F^{N-1}(\theta_i, t)(1 - F(\theta_i, t)) \ d\theta_i
\]

We need to show that \( \frac{\partial \nu_0}{\partial t} > 0 \) by making use of the fact that as \( t \) increases the distribution \( F \) undergoes a mean-preserving spread. Roughly speaking a mean preserving spread makes bidding less competitive by making types more disperse and increases the expected type of the winner. Thus, the expected surplus of the winner also increases. See Appendix A for details.

\[\blacksquare\]

### 3.1.2 Contests in Which \( \bar{c}(t) \) May Bind

From Proposition 2 we learn that when the cost of effort in the selection contest is unconstrained agents will exactly offset the cost savings of an earlier selection contest with higher effort, leaving the principal unable to affect the effort costs by his choice of timing. However,

\(^9\)For instance \( \bar{c}(t) \geq \max_{\theta} \text{supp}(F(\theta, t)) \)
in many applications when the principal chooses the selection contest sufficiently early, even in maximizing their selection contest performance agents are only able to incur a limited cost to their final task performance. For instance, imagine an Olympic trials that takes place a full year prior to the Olympics. An athlete can peak at the trials and, with a full year to recover, peak again at the Olympics. Similarly, a political candidate can move her platform all the way to that of the party’s median voter and still have sufficient time until the general election to move it back to the population median. In both situations, by running the selection contest early enough the principal restricts the agents’ ability to incur costs.

We accommodate for this by now considering a setting in which in some periods the cost constraint binds. In such a period $t$ the equilibrium in the selection contest is characterized by a cutoff type $\hat{\theta}_i$ such that $c(\theta_i, t)$ is the solution to (12) for $\theta_i \leq \hat{\theta}_i$ and $c(\theta_i, t) = \bar{c}(t)$ for $\theta_i > \hat{\theta}_i$. The cutoff type is indifferent between the two bids. If the cutoff type bids according to (12), she wins if and only if all other agents have types below $\hat{\theta}_i$. Hence this cutoff type’s interim probability of winning is $X_i(\hat{\theta}_i, t) = F_{N-1}(\hat{\theta}_i, t)$. If she were to bid $\bar{c}(t)$ she would win against all types below $\hat{\theta}_i$ and tie with any types above $\hat{\theta}_i$. With $N$ agents there are several combinations in which ties can occur and the interim probability of winning is given by

$$\bar{X} \equiv \sum_{k=0}^{N-1} \frac{1}{k+1} \binom{N}{k} F(\hat{\theta}_i)^{N-1-k} (1 - F(\hat{\theta}_i))^k$$

(13)

Note that $\bar{X}(\hat{\theta}_i) > F_{N-1}(\hat{\theta}_i, t)$ since the possibility of tying with higher types makes the interim probability of winning jump discretely. When $N = 2$ the expression simplifies to

$$\bar{X}(\hat{\theta}_i) = F(\hat{\theta}_i) + \frac{1}{2} (1 - F(\hat{\theta}_i))$$

The cutoff type $\hat{\theta}_i$ is then defined implicitly by the indifference condition

$$F_{N-1}(\hat{\theta}_i, t)(\hat{\theta}_i - c(\hat{\theta}_i, t)) = \bar{X}(\hat{\theta}_i)(\hat{\theta}_i - \bar{c}(t))$$

(14)

In order for the equality to hold, it must be that when bidding the ceiling, the discrete jump in the interim probability of winning is accompanied by a corresponding discrete jump in the bid; thus $\bar{c}(t) > c(\hat{\theta}_i, t)$. Figure 1 depicts an example of a selection contest equilibrium with a binding ceiling constraint.

Having characterized what happens when the bid ceiling binds in the selection contest, we will argue shortly that without additional structure the timing of the selection contest is a complex problem that does not easily lend itself to comparative statics. To this end, we first ask a simpler question: Does a lower bid ceiling make the principal better off? That is if it were the case that an earlier selection contest reduces the maximal cost agents can incur and if we ignore the lost option value of waiting, does the principal benefit from an earlier contest?
The answer is not necessarily, and the reasoning behind this answer comes from the discussion of the optimal mechanism in Section 2. Recall that equation (10) allows us to express the payoff to the principal in any incentive compatible mechanism only in terms of the allocation rule. Recall also that ceteris paribus, in a collection of agents \((\theta_1, \ldots, \theta_N)\) the principal wants to assign all the probability to the agent with the highest virtual type \(1 - F(\theta_i) / f(\theta_i)\). In this sense, a lower bid ceiling is just a different allocation rule, and depending on the distribution of types \(F\), not necessarily a better one.

**Lemma 3** Depending on the distribution of types \(F\), the principal can be either better or worse off with a lower bid ceiling. That is, there exist both positive and negative realizations of \(\frac{\partial u}{\partial \bar{c}}\) for some distributions.

See Appendix B for a proof of Lemma 3.

Lemma 3 highlights that the optimal timing problem is difficult to solve in general. Waiting to hold the contest always involves the benefit of cherry-picking those agents who have received a positive shock. However, the expected cost incurred by the winner in the selection contest can either increase or fall with time and, moreover, this effect can be non-monotonic. We now impose some additional structure on the problem that will provide tractability while still preserving the central tradeoffs.

### 3.2 A Two-Period Model

The principal now has only two choices of timing, an early contest at \(t = 0\) or a late contest at \(t = 1\). The early contest is defined as early enough so that agents can fully recover and peak again for the final task, hence \(c(0) = 0\). As before at \(t = 0\) all agents have type \(\theta_0\). Since all agents have the same type and will expend no costs, it follows that the payoff to
the principal from holding the selection contest in period $t = 0$ is

$$u_0(t = 0) = \theta_0$$

In the late contest agents are distributed according to $F(\theta_i, 1) = F(\theta_i)$. Recall that since types evolved over time through a mean-preserving stochastic process, $E[\theta_i] = \theta_0$. We assume that $c(1)$ is large enough so as not to bind. This principal’s payoff from holding the selection contest in period $t = 1$ is

$$u_0(t = 1) = N \int_{\theta_i} F^{N-1}(\theta_i) (1 - F(\theta_i)) \, d\theta_i$$

(15)

Note that $u_0(t = 0)$ is a fixed value and thus in examining the timing decision we can solely focus on what affects $u_0(t = 1)$. For example, what about the shape of $F$ makes a later selection more appealing? What happens to the benefit of a later selection when there are more participants? We provide first an example with a parameterized uniform distribution and then a series of lemmas to answer these questions more generally.

As a benchmark, first consider a distribution $F(\theta_i)$ that is uniform on $[\theta_0 - \alpha, \theta_0 + \alpha] \subset \mathbb{R}^+$. The expected surplus to the principal is

$$u_0(t = 1) = N \int_{\theta_i} F^{N-1}(\theta_i) (1 - F(\theta_i)) \, d\theta_i$$

This surprisingly simple expression yields immediate comparative statics. The benefit to a later selection contest is (i) independent of $\theta_0$, (ii) increases in the amount of noise, and (iii) decreases in the number of participants. Intuitively, increases in $\theta_0$ get competed away by all the agents; a greater $\alpha$ adds to the option value of waiting; and a larger number of participants eats away at the winner’s surplus through increased competition. In what follows, we will show that (i) and (ii) hold in general, while there is a countervailing effect that makes (iii) depend on the distribution.

First, we show that the payoff to the later contest depends only on the shape of $F$ and is not a function of where $F$ is centered.

**Lemma 4** For some $a > 0$, let $G(\theta_i) = F(\theta_i - a)$ so that distribution $G$ is obtained by shifting $F$ to the right by $a$. Then $u_0(t = 1|F) = u_0(t = 1|G)$.

**Proof** Follows from applying a change of variables to (15). ■

One can think of an agent’s type as composed of a common component $\theta_0$ and an idiosyncratic component $\theta_i - \theta_0$. The lemma states that a later selection contest erodes away the
common component through competition, and while this relies on the fact \( c \) does not bind, the qualitative message still holds when the assumption is relaxed. One implication is that if a principal can make a costly investment which can improve all agents’ types, he should weigh this investment carefully since the eventual winner will have lost much of its benefit through competition. On the other hand, when holding the selection contest early the principal gets the full return on such an investment.

While the mean of \( F \) has no effect on the payoff of a later contest, given that there is an option value associated with waiting, we may expect that when \( F \) is more disperse a later contest is more beneficial. However, the shape of the distribution also affects the level of competition, thus the result is not immediate and requires a more careful analysis.

**Lemma 5** If \( G(\theta_i) \) is a mean preserving spread of \( F(\theta_i) \) then \( u_0(t = 1|G) \geq u_0(t = 1|F) \).

Lemma 10 is proven in Appendix A.

This result is a generalization of the finding in the uniform case in which \( u_0(t = 1) \) is increasing in \( \alpha \). Note that MPS is a partial ordering over distributions and that the converse of the lemma is not true. That is, one can construct examples in which \( u_0(t = 1|G) > u_0(t = 1, F) \) but \( G \) is not a mean preserving spread of \( F \). The two lemmas combine into one of our main propositions.

**Definition 1** The distribution of \( Y \) is a spread of the distribution of \( X \) iff there exists a random variable \( Z \) and constant \( c \) such that (i) \( E[Z|X] = c \) for all values of \( X \) and (ii) \( Y \overset{d}{=} X + Z \).

**Proposition 6** If \( G(\theta_i) \) is a spread of \( F(\theta_i) \) then \( u_0(t = 1|G) \geq u_0(t = 1|F) \).

Next we consider how the number of agents \( N \) figures into the timing decision. The payoff to the early \( t = 0 \) contest is independent of \( N \). However, at \( t = 1 \) the number of agents matters in two ways. On one hand, the identity of the highest of \( N \) draws is on average higher when \( N \) is higher. On the other hand, an increase in the number of agents makes the contest more competitive, inducing the eventual winner to bid a higher cost. As we will show, either effect can dominate.

We consider a parameterized family of distributions \( F(\theta_i) = \theta_i^a \) with support on \([0, 1]\) and \( a > 0 \). Using equation (15) we can compute the principal’s payoff in the later contest as a function of \( N \).

\[
\begin{align*}
u_0(N) &= \int_0^1 \theta_i^a(N-1)(1 - \theta_i^a) \text{d}\theta_i = \frac{aN}{(1 + a(N - 1))(1 +aN)} \\
u_0(N + 1) - u_0(N) &= \frac{a}{(1 + a(N - 1))(1 + a(N + 1))(1 +aN)(1 - a(N + 1))} \\
&\geq 0 \text{ if and only if } a \leq \frac{1}{N + 1}
\end{align*}
\]

14
When \( a \) is sufficiently small the principal is better off with an extra agent, otherwise he is better off with one fewer agent. With more agents the field becomes more competitive but also the odds of drawing a very good agent increase, and the margins of both of these effects are sensitive to the shape of the distribution. In the particular family that we have chosen, for small values of \( a \) most agents are already crowded in a small interval at the low end of the distribution and the marginal increase in competition from an additional agent is negligible relative to the marginal increase in the likelihood that an extra agent will draw a much higher type.

**Lemma 7** The principal’s payoff to the later contest can increase or decrease with the number of participants.

### 3.3 Contests of Skill

Thus far we have assumed that the selection contest is one of will in the sense that higher types win because they are willing to pay a higher cost in order to win. But often, selection contests are also contests of skill so that of two agents putting in the same effort, the agent with the higher type will score higher in the contest. Thus, in this section we assume that \( s(e_i, \theta_i) \) is increasing in the second argument and analyze how this change affects the optimal timing decision.

First, note that although it may seem intuitive that reducing agents’ cost of obtaining a high score in the selection contest leads to lower costs and better final task performance, we have already implicitly shown that generally this is not the case. Specifically, note that in Section 3.1 no assumptions were placed on the function \( s(e_i) \) other than it being increasing. Hence, if we were to change the contest so that we have a new scoring function \( \tilde{s}(e_i) = s(2e_i) \), thus making a given score half as expensive, the equilibrium utility to the principal would remain unchanged. We will show that it is not the absolute value of \( s(e_i, \theta_i) \) that matters but rather how it varies with \( \theta_i \).

For clarity, we continue to evaluate the two period case, specifically the contest at \( t = 1 \) with a \( \tilde{c} \) that does not bind. Assume that \( s_2(e_i, \theta_i) > 0 \). Let \( c(s, \theta_i) \) be the cost borne by type \( \theta_i \) in obtaining a score of \( s \) in the selection contest. An agent’s expected utility is

\[
u_i(s_i, \theta_i) = \Pr(s_i > s_{j\neq i}) (\theta_i - c(s_i, \theta_i))
\]

**Claim 8** There is a separating equilibrium in which \( s(\theta_i) \) is strictly increasing, thus the interim probability of winning is \( X_i(\theta_i) = F_{N-1}(\theta_i) \).

**Proof Sketch**

In the original setting with \( s(e_i, \theta_i) = s(e_i) \), this result was guaranteed by the single crossing property since higher types have a higher marginal willingness to pay and the same marginal cost. In our setting, higher types still have a higher marginal willingness to pay and a smaller marginal cost.

\[\blacksquare\]
Since the equilibrium is separating, we can use the standard direct mechanism tools to compute agents’ expected equilibrium utility. Given some equilibrium function \( s(\theta_i) \), let \( c(\theta, \theta_i) \equiv c(s(\theta), \theta_i) \) be the cost to type \( \theta_i \) of choosing the equilibrium score of type \( \theta \). An agent’s expected utility can be rewritten in terms of her type and her report is

\[
\begin{align*}
    u_i(\theta_i) &= F^{N-1}(\theta_i) - c(\theta, \theta_i) \\
    \frac{du_i}{d\theta} &= \frac{\partial u_i}{\partial \theta}(\theta_i, \theta_i) + \frac{\partial u_i}{\partial \theta_i} \\
    &= 0 + F^{N-1}(\theta_i) - c_2(\theta_i, \theta_i)
\end{align*}
\]

That \( \frac{\partial u_i}{\partial \theta_i}(\theta_i, \theta_i) = 0 \) follows because reporting one’s own type is a best response. Expression (16) is similar to the incentive constraint from Myerson (1981), with the last term accounting for the heterogeneous costs. An agent’s expected utility in the contest is then

\[
\begin{align*}
    u_i(\theta_i) &= u_i(\hat{\theta}) + \int_{\hat{\theta}} F^{N-1}(\hat{\theta}) - c_2(\hat{\theta}, \hat{\theta}) \, d\hat{\theta} \\
    &= \int_{\hat{\theta}} F^{N-1}(\hat{\theta}) - c_2(\hat{\theta}, \hat{\theta}) \, d\hat{\theta}
\end{align*}
\]

The last line follows because type \( \hat{\theta} \) wins with probability zero. Since \( s_2(e_i, \theta_i) > 0 \), it must be that \( c_2(s_i, \theta_i) < 0 \). Comparing (5) with (17), it is evident that when higher types are more skilled in the selection contest, the interim expected utility of all agents and thus the expected utility of the principal are higher.

**Proposition 9** The more is the selection contest one of skill, the more beneficial it is to hold it at \( t = 1 \). That is, given two selection contests \( s^A(e_i, \theta_i) \) and \( s^B(e_i, \theta_i) \), if \( s^A_2 > s^B_2 \) at every point, \( u_0(t = 1|s^A) > u_0(t = 1|s^B) \).

**4 Conclusion**

We consider the decision of a principal that times a selection contest for a group of agents whose types he does not know. The principal uses the outcome of the contest to choose the agent best suited to perform a task, and any effort that agents put forth in order to win the contest hurts their task performance. Agents’ types evolve over time making a later selection more accurate, however as time passes effort in the contest becomes costlier to the agents’ task performance conditional on being selected. We examine the determinants of the principal’s decision in light of this inherent tradeoff.

Considering the timing problem directly, we find that while effort is less costly in earlier contests much of these savings are wiped out by agents simply putting in more effort. Hence it is not enough for a principal to pick a slightly earlier contest; he must pick one early enough so as to physically constrain agents from being able to divert resources from the final task. We show that the benefits of a later contest are only a function of the dispersion of
agents’ types and not a function of the average of their types, which means a later contest is only optimal if the dispersion of ability is sizeable relative to the average ability. We also show that including more agents in a later contest is not necessarily beneficial, as doing so trades off drawing a potentially higher type against the effects of increased competition. Lastly we show that in selection contests in which skill gives agents an edge, a later contest is more favored because the eventual winner is likely to incur a smaller cost. Improving all agents’ ability in the selection contest however provides no benefit.

Recasting this problem as one of designing an optimal screening mechanism provides several insights. Namely, due to the fact that competition is costly for the principal, he wants to design a noisy contest so that the incentive for effort is muted. Along the same lines, when taking the effects of competition into consideration, the principal wants to design a mechanism that favors those agents that are “isolated” in the distribution and thus compete away relative little of their surplus.

The optimal screening mechanism provides a first best benchmark but is not always implementable, as it requires a principal to commit to actions ex-ante that he may not want to take ex-post. The timing games on the other hand require no commitment by the principal, as every action he takes is best conditional on his beliefs. Yet timing games are just a subset of the bigger set of implementable mechanisms, and an important extension would be to characterize this set and the optimum within it.

References


Appendix A

Proposition 2 and Lemma 10 follow directly from this lemma:

Lemma 10 If $G(\theta_i)$ is a mean preserving spread of $F(\theta_i)$ then

$$ \int_{\theta_i} G(\theta_i)^{N-1}(1 - G(\theta_i))d\theta_i \geq \int_{\theta_i} F(\theta_i)^{N-1}(1 - F(\theta_i))d\theta_i $$

Proof Rothschild and Stiglitz (1970) have shown that if $G$ is a mean preserving spread of $F$ then there exists a sequence $F_i$ so that $F_0 = F$ and $F_i \to G$ such that $F_n$ and $F_{n+1}$ are separated by what the authors call a “single Mean Preserving Spread” (single MPS). Our goal is then simplified to prove the proposition only for $F$ and $G$ separated by a single MPS.

We restate the Rothschild and Stiglitz (1970) definitions here. It is useful to define the function

$$ s(x) = \begin{cases} 
\alpha & \text{for } a < x < a + t \\
-\alpha & \text{for } a + d < x < a + d + t \\
-\beta & \text{for } b < x < b + t \\
\beta & \text{for } b + e < x < b + e + t \\
0 & \text{otherwise} 
\end{cases} $$

with

$$ \alpha, \beta \geq 0 $$
$$ a \leq a + t \leq a + d \leq a + d + t \leq b \leq b + t \leq b + e \leq b + e + t $$
$$ ad = \beta e $$

An example of such a function is depicted in Figure 2. When the function $s(x)$ is added to a probability density function, $f(x)$, it shifts $\alpha$ probability mass down a distance $d$ and $\beta$ probability mass up a distance $e$, and since $ad = \beta e$, this preserves the mean. A couple of things to note here. Given that probability density functions are defined so $g(x) = $
Figure 2: A single MPS $s(x)$

$f(x) + s(x)$, it must be that $G(x) > F(x)$ for $a < x < a + d + t$ and $G(x) < F(x)$ for $b < x < b + e + t$; for all other $x$, it must be that $G(x) = F(x)$. Furthermore,

$$
\int_{a}^{a+d+t} G(x) - F(x) \, dx = \int_{b}^{b+e+t} F(x) - G(x) \, dx = \Delta
$$

Thus to go from $F$ to $G$ we first add a mass of $\Delta$ over the interval $[a, a+d+t]$ and then take that mass away over the interval $[b, b+e+t]$.

**Claim 11**

\[
\int_{a}^{a+d+t} G(x)^{N-1}(1-G(x)) - F(x)^{N-1}(1-F(x)) \, dx \\
> (F(a + d + t) + \Delta)^{N-1}(1 - (F(a + d + t) + \Delta)) - F(a + d + t)^{N-1}(1 - F(a + d + t)).
\]

and

\[
\int_{b}^{b+e+t} G(x)^{N-1}(1-G(x)) - F(x)^{N-1}(1-F(x)) \, dx \\
< (F(b) + \Delta)^{N-1}(1 - (F(b) + \Delta)) - F(b)^{N-1}(1 - F(b)).
\]

**Proof** Follows from the fact that the function $F^{N-1}(1 - F)$ is concave in $F$ and maximized at $F = \frac{1}{2}$.  

Intuitively consider maximizing the function $\int F(\theta)^{N-1}(1 - F(\theta))d\theta$ by adding a mass to a particular realization of $F(\theta)$. Because $F^{N-1}(1 - F)$ is concave in $F$ and maximized at $F = \frac{1}{2}$ the best place to add this mass is as far left as possible. Conversely, if we were to be removing a mass, the best place to do so would be as far right as possible.

Putting the inequalities from the claim together, an immediate corollary is that

\[
\int_{a}^{a+d+t} G(x)^{N-1}(1-G(x)) - F(x)^{N-1}(1-F(x)) \, dx \geq \int_{b}^{b+e+t} G(x)^{N-1}(1-G(x)) - F(x)^{N-1}(1-F(x)) \, dx
\]

which implies that $\int G(x)^{N-1}(1 - G(x)) \, dx > \int F(x)^{N-1}(1 - F(x)) \, dx$. This concludes the proof of the lemma.
Appendix B

Lemma 3 is proven below.

Proof The principal’s expected payoff in this selection contest is given by equation (10)

\[ u_0 = N \int_{\theta_{\bar{i}}}^{\theta_i} X(\theta_i, t)(1 - F(\theta_i, t)) d\theta_i \]

\[ = N \int_{\theta_{\bar{i}}}^{\theta_i} F^{N-1}(\theta_i, t)(1 - F(\theta_i, t)) d\theta_i + N \int_{\theta_{\bar{i}}}^{\theta_i} (\bar{X}(\hat{\theta_i}) - F^{N-1}(\theta_i))(1 - F^{N-1}(\theta_i)) d\theta_i \]

It follows from equation (14) that \( \frac{\partial \hat{\theta}_i}{\partial \bar{c}} > 0 \), hence it suffices to examine \( \frac{\partial u_0}{\partial \hat{\theta}_i} \). Note that only the second term is a function of \( \hat{\theta}_i \).

\[
\frac{\partial u_0}{\partial \hat{\theta}_i} = N \left[ \frac{\partial \bar{X}}{\partial \hat{\theta}_i} \int_{\bar{\theta}_i}^{\theta_i} (1 - F^{N-1}(\theta_i)) d\theta_i - \left( \bar{X}(\hat{\theta}_i) - F^{N-1}(\theta_i) \right) \left( 1 - F^{N-1}(\theta_i) \right) \right]
\]

From equation (13),

\[
\frac{\partial \bar{X}}{\partial \hat{\theta}_i} = f(\hat{\theta}_i) \sum_{k=0}^{N-1} \frac{1}{k+1} \binom{N-1}{k} \left( (N - 1 - k)F^{N-2-k}(1 - F)^k - kF^{N-1-k}(1 - F)^{k-1} \right)
\]

\[
= f(\hat{\theta}_i) \sum_{k=0}^{N-1} \frac{(N - 1)!}{(N - 2 - k)!(k + 1)!} \left( \frac{1}{k + 2} \right) F^{N-2-k}(1 - F)^k > 0
\]

The second line follows from the first through a grouping of terms. We see that \( \frac{\partial \bar{X}}{\partial \hat{\theta}_i} \) is strictly positive and increases linearly in \( f(\hat{\theta}_i) \). Fixing the function \( F(\theta_i) \), the derivative \( \frac{\partial u_0}{\partial \hat{\theta}_i} \) is positive for large values of \( f(\hat{\theta}_i) \) and negative for small values. This is consistent with the idea that the principal wants to allocate to agents with high virtual types. An increase in \( \hat{\theta}_i \) means that this type will now lose to people above her instead of tying with them. The larger is \( f(\hat{\theta}_i) \), the lower is that agent’s virtual type and the better off is the principal in allocating away from her. ■