# SOCIAL STATUS AND THE DESIGN OF OPTIMAL BADGES

NICOLE IMMORLICA, GREG STODDARD, AND VASILIS SYRGKANIS

Abstract. Many websites rely on user-generated content to provide value to consumers. Often these websites incentivize user-generated content by awarding users *badges* based on their contributions. These badges confer value upon users as a symbol of social status. In this paper, we study the optimal design of a system of badges for a designer whose goal is to maximize contributions. We assume users have heterogeneous abilities drawn from a common prior and choose how much effort to exert towards a given task. A user's ability and choice of effort determines the level of contribution he makes to the site. A user earns a badge if his contribution surpasses a pre-specified threshold. The problem facing the designer then is how to set badge thresholds to incentivize contributions from users. Our main result is that the optimal total contribution can be well-approximated with a small number of badges. Specifically, if status is a concave function of the number of players with lower rank, then a single badge mechanism that divides players in two status classes suffices to yield a constant approximation, whilst for more general functions we show that typically logarithmic, in the number of players, badges suffice. We also show that badge mechanisms with a small number of badges have nice structural stability properties for sufficiently large number of players.

## 1. Introduction

A number of popular sites on the web today are driven by user-generated content. Review sites such as Yelp and TripAdvisor need users to rate and review restaurants and hotels, social news aggregators like Reddit rely on users to submit and vote upon articles from around the web, and question and answer sites like Stack Overflow and Quora depend on their users to ask good questions and provide good answers. Stack Overflow, a user-driven Q&A site for programming questions, is one of the most successful of these sites; over five million questions have been asked, with over 60% of those questions receiving a satisfactory answer. An often-cited reason for the high quality of the Stack Overflow community is the clever system of *reputation points* and *badges* it employs. While the main actions on SO are posting and answering questions, a variety of other activities such as flagging posts for moderation, rating contributions, etc are all necessary to maintain a healthy and vibrant discourse on the site. Stack Overflow motivates question and answers through the use of reputation points but all of these secondary activities are motivated by awarding *badges*. A badge is a small symbol displayed on a user's profile and posts and is typically awarded for accomplishing a fixed task, such as answering ten questions, reviewing 100 restaurants, or completing an online course. Most badge systems are not designed to promote competition among users but simply to reward individual accomplishments. As such, badges do not have the exogenously imposed scarcity that ranking systems (top 10 contributors) or contest systems (medals for first, second, and third place) have. In this paper, we examine the incentives created by these badge schemes and focus on designing badge schemes that maximize the user contributions to a site.

Badge systems have become particularly popular in web communities over the past few years. In addition to Stack Overflow, the Huffington Post recently implemented a badge system to reward actions from their commenters; these badges are now prominently displayed next to usernames in the comment sections. Foursquare, a mobile social networking app, awards badges for visiting certain locations or accomplishing certain tasks while there. The Mozilla Foundation is leading an initiative called the Open Badge Project which hopes to set an open standard for awarding, collecting, and displaying badges. Their ultimate goal is provide a persistent collection of badges that can be displayed anywhere on the web as a proof of acquired skills and achievements.

With all this excitement and energy surrounding badge systems, one fundamental question emerges: why do people care about badges? On the surface level, a badge is just a small sticker or group of pixels on a profile, so why should it be that badges incentivize people to exert effort to earn them? In this work, we posit the primary value of a badge is derived from its ability to confer social status to it's owner. The most valuable badges not only signal that a user has accomplished a certain task but they also signal that relatively few users have earned them. Indeed, many systems include a top level of badges such as the "gold" badges on Stack Overflow or the "Superuser" badge on Huffington Post, which are described to be difficult to earn and only awarded to the most committed users. As more users earn a particular badge, that badge loses its ability to distinguish a member within the community and thus becomes less valuable. This work studies optimal badge design when badges are a means for establishing relative social status.

The goal of the designer is to maximize the total contributions that users make towards a single task on the site. Then the main question is how to design a set of badges to maximize the total output that users contribute to that task? How many badges does the designer need to set and how should they be awarded? For example, if Yelp wants to maximize the number of reviews they receive, should they award just one badge for contributing at least 50 reviews or should they have one badge for 10 reviews and another badge for 100 reviews? In our model, users have heterogenous abilities for certain tasks on a site; users of low ability need to exert more effort to produce the same output as high ability users. This is effort is costly to users, so they will balance the value they get from earning a badge with the effort it takes to do so. Users all have the same base value for a badge (normalized to 1) but the value of a badge diminishes as more users acquire it. The way in which each agent's value for a badge diminishes with the number of others users that have it is given by a *status function*.

In keeping with the general design and ethos of badge systems, we study badge mechanisms which use a small number of *threshold badges*. A threshold badge is one which is awarded to any user who contributes more than a fixed threshold, such as a badge for answering 10 questions, another badge for answering 100 questions, etc. In this sense, badges are used to reward absolute accomplishments rather than relative ones. Thus the designer cannot use the ex-post scarcity imposed by schemes such as awarding only the top 10 contributors or ranking players into $1^{\text{st}}, 2^{\text{nd}}, \ldots$, to motivate agents. We insist that our mechanism use a small number of badges for two reasons. On a technical level, a badge mechanism with an arbitrarily large number of badges can essentially implement a ranking mechanism, so allowing for an arbitrarily large amount of badges would not elucidate the fundamental differences between threshold badges and rankings. Furthermore, badge systems tend to have a small number of badges in practice. StackOverflow, for example, divides their badge systems into just three tiers, bronze badges, silver badges, and gold badges. Bronze badges are awarded for small contributions, such as editing 10 posts, silver badges for medium contributions, editing 80 posts, and gold badges for high contributions, editing 500 posts. Similarly, The Huffington Post divides their badge system into two categories ("level 1" and "level 2"), and many other systems (such as traveler status with airlines) employ the "bronze,silver,gold" reward structure. A system with a small number of badges can be advantageous in many scenarios, such as when it is costly to implement additional badges or when an overly complex badge structure may overwhelm users. In this work, we describe the size of a badge mechanism by the number of badges it specifies. The results in this paper show that small badge mechanisms are competitive with both arbitrarily large badge mechanisms and general ranking mechanisms.

Our main result is that the optimal small badge mechanism receives at least a constant fraction of the contributions elicited by the optimal mechanism, but the precise definition of "small" is determined by the shape of the status function. When the status function is concave, i.e. the marginal value of increasing status decreases as status gets higher, a single badge is sufficient to achieve a 4-approximation to the optimal mechanism. Further, if the status function is linear, this result is improved to a 2-approximation. However, if the concavity condition fails, $\log(n)$ badges are necessary for a badge mechanism to approximate the optimal mechanism within a constant factor, where $n$ is the number of agents. For general status functions, if the status of the highest-contributing agent is no more than $B$ times the status of the lowest-contributing agent, then a badge mechanism with $\log(B)$ badges gives a 2-approximation to the optimal mechanism. For a number of a natural status functions, this bound $B$ is at most the number of agents.

The remainder of the paper is laid out as follows. In section 3 we formally describe our model and we characterize the equilibria of this model in section 4. In section 5 we derive the form of the optimal mechanism and use that characterization in section 6 to show that small badge mechanisms are approximately optimal.

## 2. Related Work

The paper that is most closely related to ours is that of Moldovanu, Sela, and Shi [12], and in fact, the specification of the game is identical to our setup (which is also used in [5]): agents exert effort towards some task, are assigned into categories based on their output, and gain (or lose) social status based on the rank of their category. When the distribution of agents' abilities satisfy a monotone hazard rate assumption, they find the optimal mechanism assigns agents to unique categories and further show that a coarse partitioning of agents into one of two categories (winners or losers) can elicit at least half of the total effort as the optimal mechanism. Our work differs in two important ways. First, their mechanism assigns each agent into a category based on his relative position in the ranking of all contributions while our mechanism assigns agents into categories only based on absolute thresholds for contributions. Second, [12] considers a specific social status function that is a linear function of the number of players you beat and the number of players you lose to. Specifically, each player gets a utility of 1 for each player he beats and a utility of $-1$ for each player that he loses to. In our setting we consider arbitrary status functions of the number of players that a person beats (is ranked strictly higher). [12] shows that only two status classes are sufficient to yield a 2-approximation to the optimal mechanism for monotone hazard rate distributions. We show that for any concave status function a single badge mechanism, which also discriminates in two status classes, is a 4-approximation to the optimal mechanism for the more general class of regular functions. Thus our setup is more general than the one in [12] in two ways: first the more general class of regular distributions is used, and an arbitrary status function is analyzed. The restriction is that this status function is dependent only on one parameter: the number of players that are ranked strictly lower, rather than on two parameters as in

[12]: the number of players ranked strictly lower and the number of players ranked strictly higher.

Dubey and Geanakoplos [5] consider the same contest for status but assume each player has complete information about the types of other players. In contrast to the incomplete-information game, they find that the optimal mechanism will often group multiple players into the same status category which illustrates the differences in the incomplete versus complete informations settings.

Hopkins and Kornienko [9] consider a model where consumers care about their social status, as determined by their consumption of a "positional" good, and study how shifts in equality in the societal distribution of income affect conspicuous consumption. They use a connection between first-price auctions and their status model in order to reason about the strategic behavior induced by status concerns. Our work extends this connection by drawing a link between a wider class of social status functions and allocations resulting from a general mechanism.

There's a growing literature on the role and design of incentives systems in online settings such as question and answer websites (e.g. StackOverflow, Quora, etc). This line of the literature considers how to award virtual points on a website in order to maximize some objective for the designer. In the context of Q&A sites, Ghosh and Hummel [7] show that a simple scheme of giving a large amount of points to the winning answer and a smaller amount of points to each remaining answer can maximize a large class of objective functions. Further, they show always choosing the best contribution to be the winning contribution actually *decreases* the designer's value. Jain, Chen, and Parkes [10] study the design of Q&A reward schemes in a more complex environment where users contribute answers sequentially and the value of a contribution may vary based on the set of other contributions the designer receives, such as the case where contributions may compliment each other. They show there is no mechanism which can achieve the efficient outcome for every type of valuation for the designer. Hartline et al [4] study crowdsourcing contests where the principal only receives value from the highest quality submission, as is the case with crowdsourcing tasks such as the design of a poster or a logo. Using tools from optimal mechanism design, they show the optimal crowdsourcing contest gives all the prize money to the best contribution and nothing to all other contributions. Cavallo and Jain [3] shows that winner-take-all contests are not welfare-maximizing when there's uncertainty in the conversion from effort to quality and instead propose a mechanism where every contributor earns a minimum amount of money.

Easley and Ghosh [6] consider a model of earning badges in the presence of noisy observations. They show that when a user's value for a badge is exogenously determined, the optimal badge structure is to award the badge only to a fixed number of top contributors. They also extend their model to the case when a user's value for a badge changes based on the number of other users who have earned that badge and interestingly show that a designer can take advantage of user uncertainty over the number of badge earners in order to incentivize more contributions. Depending on the shape of the badge value function (as a function of mass of users who have earned that badge), hiding the number of winners may increase or decrease contributions in equilibrium. Kleinberg et al [1] find empirical evidence that people are motivated by badges; using the public logs from Stack Overflow, they show that as users get closer to the threshold for winning a particular badge, they spend comparatively more effort performing the action associated with that badge. They also develop a theoretical model describing a how a single user may change his behavior to earn badges, and how to design badge thresholds under this model.

## 3. Badge Utilities and Social Status

We consider the following game theoretic model of contributions to a user-driven site. There are $n$ players each having ability $v_i$ with respect to some task which the designer wishes to incentivize. Ability is a context-specific trait such as programming knowledge in the case of Stack Overflow or inherent ability to write good restaurant reviews. If player $i$ invests effort $e_i$ on this task, then she produces an output of $b_i = v_i \cdot e_i$. A user's output is their observed contribution to the site, such as the number of questions answered or the quality of a restaurant review. The goal of the designer is to maximize the sum of the output of all players $\sum_i b_i$.

**Badge Mechanism.** Before players invest effort, the designer defines a *badge mechanism*, which is a set of $m$ badges associated with this task. Specifically, each badge $j \in [m]$ is awarded to a user if her output $b_i$ is at least some threshold $\theta_j$. As an example badge mechanism, if the designer sets the badge thresholds $\theta = (10, 100, 200)$, it indicates that he awards badge 1 for answering 10 questions, badge 2 for 100 questions, and badge 3 for 200 questions. Without loss of generality, we assume $\theta_1 \leq \ldots \leq \theta_m$. Given a player's output $b_i$ we denote a player's *rank* $r(b_i)$ as the highest badge that player $i$ wins, i.e. the highest threshold $\theta_j$ such that $b_i \geq \theta_j$. If a player doesn't win any badge then we assume his rank is zero.

**Player Utilities.** In this paper, a player's value for a badge is determined endogenously by its ability to signal social status within the community. If a player receives multiple badges, they only value the highest badge they won. Thus their value is determined by their rank. Given a profile of the ranks of all players $r = (r_1, \ldots, r_n)$, a player's value for her rank is some continuous decreasing function $S(\cdot) : [0,1] \to \mathbb{R}^+$ of the proportion of people that have an equal or higher rank (and zero in the case when she receives no badge). For example, $S(t)$ could be the proportion of the population that a player beats, $S(t) = 1 - t$, or it could be the inverse of the proportion of the population that is at least as good as the player $S(t) = \frac{1}{t}$. We return to these two examples later in the paper.

Let $t_i(r) = \frac{|\{j \in [n] : r_j \geq r_i\}|}{n}$ denotes this proportion. A player's utility is quasi-linear in their value for their rank and the effort they exerted:

$$\hat{u}_i(r, e_i) = S(t_i(r)) \cdot 1_{r_i > 0} - e_i \tag{1}$$

We refer to the first term as the *status value* and we denote it as $x_i(r) = S(t_i(r)) \cdot 1_{r_i > 0}$.

Given a particular badge mechanism, the rank of a player is determined by the profile of outputs $b = (b_1, \ldots, b_n)$. Also observe that a player's effort is a function of their output and ability, so we can rewrite a player's utility as:

$$\hat{u}_i(b; v_i) = x_i(r(b)) - \frac{b_i}{v_i} \tag{2}$$

**Additional Assumptions.** We assume that each player's ability $v_i$ is private information and is drawn independently and identically from a commonly known, atomless, and regular[1] distribution $F$ with support $[0, \bar{v}]$ and density $f(\cdot)$. Hence, any badge mechanism defines a game of incomplete information among the $n$ players. We analyze the Bayes-Nash equilibria of this game. A BNE is a profile of mappings from abilities $v_i$ to outputs $b_i(v_i)$ such that each player maximizes their utility in expectation over the abilities of the rest of the players. For all $v_i$, $b_i'$:

$$\mathrm{E}_{v_{-i}}\left[u_i(b(v); v_i)\right] \geq \mathrm{E}_{v_{-i}}\left[u_i(b_i', b_{-i}(v_{-i}); v_i)\right] \tag{3}$$

[1] A distribution $F$ is regular if $v - \frac{1-F(v)}{f(v)}$ is (weakly) monotone in $v$.

Since each player is maximizing his utility conditional on his value, we can equivalently assume that a player's utility is:

$$u_i(b; v_i) = v_i \cdot \hat{u}_i(b; v_i) = v_i \cdot x_i(r(b)) - b_i \tag{4}$$

because any strategy which maximizes their original utility $\hat{u}$ also maximizes $u$. This new formulation allows us to see that this game is equivalent to an all-pay auction where a player's allocation is a function of the ranks assigned to all players.

To summarize the timing of the game, the designer first announces a set of badges $\theta$. Each agent then learns his ability $v_i$ and simultaneously decides how much output $b_i$ to contribute to the site. The designer observes the output of each agent and awards badges based on the thresholds for each badge.

**Main Question.** The designer of the badge mechanism is interested in maximizing the expected total output produced by the players: $\mathrm{E}[\sum_i b_i(v_i)]$, which is exactly the revenue of the corresponding all-pay auction. We consider two main questions:

(1) *Given a distribution of abilities, how should the designer set badge thresholds to maximize expected revenue?*
(2) *How many badges are sufficient to closely approximate the revenue achievable by any mechanism in this context?*

## 4. Uniqueness and Characterization of Bayes-Nash Equilibrium

We begin our analysis by characterizing the Bayes-Nash equilibria induced by any badge mechanism. This model falls within the class of symmetric ranking mechanisms of Hartline and Chawla [4] and therefore has a unique Bayes-Nash equilibrium. This equilibrium is symmetric and monotone, i.e. each player uses the same strategy mapping $b_i(v_i) = b(v_i)$ and this mapping is monotone in the player's ability. Thus we just need to characterize this unique symmetric Bayes-Nash equilibrium in order to analyze this game.

We start by observing that if a player gets rank $r_i$, her output should be exactly the threshold to win that badge, $\theta_{r_i}$, because producing more output would cost more effort and would not increase her value. Since the equilibrium mapping is monotone in ability, the equilibrium is defined by a set of thresholds in the ability space of the players $a_1, \ldots, a_m$, such that if player $i$ has ability $v_i \in [a_t, a_{t+1})$ then he produces output $b(v_i) = \theta_t$. If $v_i < a_1$ then $b(v_i) = 0$ and if $v_i \geq a_m$ then $b(v_i) = \theta_m$.

So given a badge mechanism with contribution thresholds $\theta = (\theta_1, \ldots, \theta_m)$, the resulting BNE is defined by a vector of thresholds on player ability, but it remains to compute those ability thresholds. We compute these thresholds using the Bayes-Nash equilibrium characterization of Myerson [13] for quasi-linear utility, single-dimensional type environments. For a fixed mapping $b(\cdot)$, we denote with $\tilde{x}_i(v_i) = \mathrm{E}_{v_i-i}[x_i(r(b(v))]$, the expected status value player $i$ gets, assuming each player uses strategy $b(\cdot)$. $\tilde{x}_i(v_i)$ is often referred to as the interim allocation of player $i$. Since our setting is completely symmetric, the interim allocation function is the same for all players, and we will denote it with $\tilde{x}(\cdot)$. Now observe that the expected utility of a player with ability $v_i$, assuming that everyone employs strategy $b(\cdot)$, is :

$$\mathrm{E}_{v_{-i}}[u_i(b(v); v_i)] = v_i \tilde{x}(v_i) - b(v_i) \tag{5}$$

Myerson's equilibrium characterization, when applied to our setting, states that $b(\cdot)$ is an equilibrium only if $\tilde{x}(v_i)$ is monotone in $v_i$ and:

$$b(v_i) = v_i \tilde{x}(v_i) - \int_0^{v_i} \tilde{x}(z)\, dz \tag{6}$$

Additionally, if the function $b(\cdot)$ spans the whole region of feasible bids, then the latter is an if and only if statement. The latter won't be true in our setting. However, the proposed strategy $b(\cdot)$ will include all the feasible undominated bids, which is all the badge threshold bids. Hence, if we find a strategy that satisfies the above equation then that strategy will be an equilibrium.

Now consider the mapping $b(\cdot)$ that corresponds to some vector of value thresholds $a$ as explained previously. For a player with value $v_i \in [a_t, a_{t+1}]$ we can compute explicitly his expected status value as follows: we know that every player with value $v_j \geq a_t$ will be assigned a rank at least as high as player $i$. Thus the status value of player $i$ is the same as the social status of a player with value $a_t$ and is simply the expected value of $S\left(\frac{T+1}{n}\right)$ where $T$ is the random variable denoting the number of players other than player $i$ that have value above $a_t$. This can be computed as follows:

$$\tilde{x}\left(a_t\right) = \sum_{\nu=0}^{n-1} S\left(\frac{\nu+1}{n}\right) \cdot \beta_{\nu,n-1}\left(1 - F\left(a_t\right)\right) \tag{7}$$

where $\beta_{\nu,n}(x) = \binom{n-1}{\nu} \cdot x^{\nu} \cdot (1-x)^{n-1-\nu}$, denotes the Bernstein polynomial. It is also easy to see that if $S\left(x\right)$ is a strictly increasing function then $\tilde{x}\left(a_t\right)$ is also strictly increasing. Additionally, since $F$ is an atomless distribution, $\tilde{x}\left(a_t\right)$ is continuous and differentiable and $\tilde{x}\left(\bar{v}\right) = S\left(1/n\right)$ and $\tilde{x}\left(0\right) = S\left(1\right)$.

Using the step nature of the interim allocation function and the fact that by the initial analysis, a player with value $v_i \in [a_j, a_{j+1}]$ bids $\theta_j$, the equilibrium characterization (6) simply becomes:

$$\theta_j = a_1 \cdot \tilde{x}\left(a_1\right) + \sum_{t=2}^{j} a_t \cdot \left(\tilde{x}\left(a_t\right) - \tilde{x}\left(a_{t-1}\right)\right) \tag{8}$$

This relation is depicted in Figure 1. Observe that the above defines a system of $m$ equalities

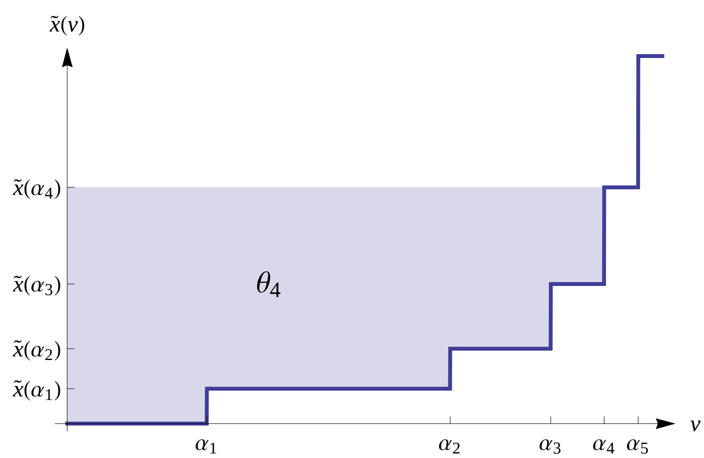


FIGURE 1. Relation between interim allocation probability and badge thresholds at equilibrium.

that have a unique solution. An equivalent way of phrasing the above equations is that: $a_1 \cdot \tilde{x}(a_1) = \theta_1$ and for $j \in [2, m]$:

$$a_j \cdot \left(\tilde{x}\left(a_j\right) - \tilde{x}\left(a_{j-1}\right)\right) = \theta_j - \theta_{j-1} \tag{9}$$

Thus, given a profile of badge thresholds $\theta$, we can iteratively solve for the profile of ability thresholds $a$ as follows: find a solution to the equation $\theta_1 = a_1\tilde{x}(a_1)$. Observe that $\bar{v}\cdot\tilde{x}(\bar{v}) = \bar{v}\cdot S(1/n)$, $0\cdot\tilde{x}(0)=0$, $v\cdot\tilde{x}(v)$ is continuous increasing. If $\theta_1 < \bar{v}\cdot S(1/n)$ then a unique solution exists. Otherwise, no player is willing to bid as high as $\theta_1$ and the recursion stops. Subsequently, find the solution $a_2$ to the equation: $\theta_2 - \theta_1 = a_2(\tilde{x}(a_2) - \tilde{x}(a_1))$. For similar reason, either a unique such solution exists or no player (not even a player with ability $\bar{v}$) is willing to bid $\theta_2$ rather than bid $\theta_1$ and we can stop the recursion. Then solve for $a_3,\ldots,a_m$ in the same way.

The following corollary of the above discussion summarizes a property that will prove very useful in subsequent sections:

**Corollary 1.** *There is a one-to-one correspondence between a vector of ability thresholds $a = (a_1,\ldots,a_m)$ and a vector of badge thresholds $\theta = (\theta_1,\ldots,\theta_m)$ such that: 1) the badge mechanism with badge thresholds $\theta$ has a unique Bayes-Nash equilibrium were a player with ability $v_i \in [a_j, a_{j+1})$ will win badge $j$, a player with ability $v_i \geq a_m$ will win badge $m$ and a player with ability $v_i < a_1$ will win no badge, 2) this is the unique badge mechanism that implements these ability thresholds. When this is the case, we say $\theta$ implements $a$.*

Instead of setting badge thresholds, a designer could optimize over thresholds in the ability space and then compute the badges needed to implement the ability thresholds. Intuitively, this design can be thought of as the direct-revelation equivalent of the badge mechanism design. This formulation of the problem is easier because we no longer need to worry about the induced equilibrium behavior. For the remainder of the paper, we focus on this version of the problem.

*Quantile Space.* Another convenient way of characterizing a badge mechanism is through the notion of the upper quantile of a distribution. Given a distribution $F$ of abilities, the quantile of ability $v_i$ is the probability that a random sample from $F$ is at least as high as $v_i$:

$$q(v_i) = 1 - F(v_i) \tag{10}$$

For an atomless continuous distribution $F$, there is a one-to-one correspondence between abilities and quantiles. Define $v(q_i) = F^{-1}(1-q_i)$ to be the ability corresponding to quantile $q_i$. We can also define the interim allocation of a player as a function of his quantile rather than his ability.

$$\hat{x}(q_i) = \tilde{x}(v(q_i))) \tag{11}$$

It is easy to see that there is a one-to-one correspondence between a vector of thresholds $a = (a_1,\ldots,a_m)$ in ability space and a vector of thresholds $\kappa = (\kappa_1,\ldots,\kappa_m)$ in quantile space. A higher ability corresponds to a lower quantile, with $q(\bar{v}) = 0$ and $q(0) = 1$, and therefore if $a_1 \leq \ldots \leq a_m$, we have $\kappa_1 \geq \ldots \geq \kappa_m$. Hence, the designer can equivalently think of designing his thresholds in quantile space rather than ability space. By the above observation and Corollary 1, for any vector of thresholds $\kappa$ in the quantile space, there exists a vector of badge thresholds $\theta$ that will implement $\kappa$ in the unique Bayes-Nash equilibrium.

Given a badge mechanism defined by a vector of quantile thresholds $\kappa$, the interim allocation of a player with quantile $q_i \in [\kappa_{t+1}, \kappa_t]$ is simply:

$$\hat{x}(q_i) = \hat{x}(\kappa_t) = \sum_{\nu=0}^{n-1} S\left(\frac{\nu+1}{n}\right)\cdot\beta_{\nu,n-1}(\kappa_t) \tag{12}$$

Using the equilibrium characterization we can compute the expected revenue produced by a set of badges with thresholds $\kappa = (\kappa_1,\ldots,\kappa_m)$ in the quantile space as follows: If a player has quantile $q_i \in [\kappa_j, \kappa_{j+1}]$ then by Equation (8) we get that he submits a bid of $\theta_j = \sum_{t=1}^{j} v(\kappa_t)\left(\hat{x}(\kappa_t) - \hat{x}(\kappa_{t-1})\right))$.[2] The latter happens with probability $\kappa_j - \kappa_{j+1}$. Thus a player's expected payment, after some rearrangements, is:

$$\begin{aligned}
\mathrm{E}_{v_i}[b(v_i)] &= \sum_{j=1}^{m}\sum_{t=1}^{j}(\kappa_j - \kappa_{j+1})v(\kappa_t)\left(\hat{x}(\kappa_t) - \hat{x}(\kappa_{t-1})\right) \\
&= \sum_{t=1}^{m}\sum_{j=t}^{m}(\kappa_j - \kappa_{j+1})v(\kappa_t)\left(\hat{x}(\kappa_t) - \hat{x}(\kappa_{t-1})\right) \\
&= \sum_{t=1}^{m}\kappa_t \cdot v(\kappa_t)\cdot\left(\hat{x}(\kappa_t) - \hat{x}(\kappa_{t-1})\right)
\end{aligned}$$

Let $R(x) = v(x)\cdot x$, which is usually denoted as the revenue function of the distribution [2, 8], since it corresponds to the revenue of a monopolist, facing a single buyer and posting a price such that he sells with probability $x$.Thus the total revenue of the mechanism is simply:

$$\mathrm{E}\left[\sum_i b_i(v_i)\right] = n\cdot\sum_{t=1}^{m} R(\kappa_t)\cdot\left(\hat{x}(\kappa_t) - \hat{x}(\kappa_{t-1})\right) \tag{13}$$

$$= n\cdot\sum_{t=1}^{m}\left(R(\kappa_t) - R(\kappa_{t+1})\right)\cdot\hat{x}(\kappa_t) \tag{14}$$

## 5. Optimal Ranking Mechanism

In the previous section, we observed that a badge mechanism induces a ranking among players; the players who win the highest badge are assigned the highest rank, players who win the second highest badge are assigned the second highest rank, and so on. These badge-induced rankings are just one example of a general mechanism which collects contributions from agents and assigns ranks based on that output. For example, we could collect contributions and then assign the highest contributor to the highest rank, the second highest contributor to the next rank, etc. Another mechanism could assign the 10 highest contributors to the highest rank, then the next 100 highest contributors to the second highest rank, and the lowest rank to every other player.

Along the same lines, one could ask the question of how does a badge mechanism with a fixed or small number of badges compare with respect to a mechanism with a continuum of badges (observe that adding more freedom to the designer to use more badges can only increase optimal revenue). A continuum of badges essentially, can simulate any symmetric ranking mechanism where rank is monotone in ability. For instance, it can simulate the mechanism where every player is assigned a distinct rank.

Each of these mechanisms would incentivize players to invest effort and contribute to the site because it gives them a way to achieve social status. In this section, we characterize the optimal ranking mechanism. We show that the optimal mechanism is a fully discriminatory one, that assigns a distinct rank to each player in decreasing order of submitted quality. In the next section we show that a simple badge mechanism provides a good approximation to the optimal mechanism.

We first introduce the optimal mechanism design problem. Suppose the designer is allowed to run an arbitrary mechanism that asks each player to report their ability, and then outputs

[2] For notational convenience assume that $\kappa_0 = 1$ with $\hat{x}(\kappa_0) = 0$ and $\kappa_{m+1} = 0$ with $\hat{x}(\kappa_{m+1}) = 1$.

a rank profile $r = (r_1, \ldots, r_n)$ and an output profile $b = (b_1, \ldots, b_n)$, such that each player $i$ contributes $b_i$ and is assigned rank $r_i$. Each player's utility for any such output is of the form $u_i(r, b_i) = v_i \cdot x(r) - b_i$, where $x(r)$ is as defined in the Section 3. The only restriction we make on the mechanism is that each player's expected utility is non-negative, i.e. it is rational for participate in the mechanism. What is the mechanism that produces the highest expected revenue in Bayes-Nash equilibrium?

This scenario is the classic single-parameter optimal mechanism design setting with quasi-linear utilities. By the Revelation Principle the designer can restrict himself to direct Bayesian incentive compatible mechanisms of the form: players report their abilities, the designer decides an allocation of ranks based on the reported abilities, and then asks the players to output some $b_i$ such that truthfully reporting ability is an equilibrium for all players in the resulting game of incomplete information.

Before computing the revenue-optimal allocation, we compute the welfare-optimal allocation. The social welfare in this setting is defined as the sum of all agents utilities plus the utility for the designer. The utility for the designer is equal to the sum of the output for the agents. For a given vector of abilities $v$, an allocation of rankings $r$ and a vector of agent outputs $b$, the social welfare is as follows.

$$SW(v, r, b) = SW_{players} + SW_{designer} = \sum_i v_i \cdot x(r) \tag{15}$$

In words, each agent's contribution to the social welfare is equal to his ability weighted by status value. The following lemma shows welfare is maximized[3] by assigning assigning each player a different rank in decreasing order of ability.

**Lemma 2.** *If $v_1 \geq v_2 \geq \ldots \geq v_n$ then the optimal social welfare is achieved by assigning $r_1 = n, r_2 = n-1, \ldots, r_n = 1$, producing welfare of:*

$$\sum_i v_i \cdot x(r) = \sum_i v_i \cdot S\left(\frac{i}{n}\right)$$

*Proof.* The statement follows by the following arguments: first it is easy to see that the rank should be monotone non-decreasing in the value, since if $v_i > v_j$ and $r_i < r_j$ then we can increase welfare by swapping the ranks of player $i$ and player $j$. Additionally, if for some set of values $v_i \geq v_{i+1} \geq \ldots v_{i+k}$ we have $r_i = \ldots = r_{i+k}$ then by discriminating $v_i$ to be strictly higher than the rest of the values increases the welfare. More concretely, for any $j > i$ we can set $r'_j = r_j + 1$. The satisfaction of all players $j > i$ doesn't change since the number of people that have rank at least as high as them remains the same. Additionally, the satisfaction of player $i$ strictly increases, since the number of people ranked at least as high as him, strictly decreased. ■

Myerson's characterization states that the expected revenue of any mechanism is equal to its expected *virtual welfare*, the sum of each agent's *virtual value*. To continue the analogy between the all-pay auction setting and our model, we analogously define the *virtual ability* of an agent with ability $v_i$ to be

$$\phi(v_i) = v_i - \frac{1 - F(v_i)}{f(v_i)} \tag{16}$$

[3]This lemma is not characterizing the welfare-optimal because we transformed utilities in equation 4. However, this lemma is useful for characterizing the revenue-optimal mechanism.

Then the virtual welfare of any ranking mechanism is $\sum_i \phi(v_i) \cdot \tilde{x}(v_i)$. The expected total output of any mechanism in any equilibrium profile $(b_1(\cdot), \ldots, b_n(\cdot))$ is equal to the virtual welfare.

$$\mathrm{E}_v \left[ \sum_i b_i(v_i) \right] = \mathrm{E}_v \left[ \sum_i \phi(v_i) \cdot \tilde{x}(v_i) \right] \tag{17}$$

This characterization and lemma 2 imply that the optimal mechanism orders players by decreasing virtual ability and assigns them a unique rank, so long as their virtual ability is positive. Players, with negative virtual value are assigned rank 0. In section 3 we assumed the ability distribution $F$ is regular, so virtual ability is monotone in ability and ordering by virtual ability is equivalent to ordering by ability. The optimal mechanism then asks each player to contribute an output of $b_i(v_i)$ according to the payment identity (6).

**Corollary 3** (Optimal Ranking Mechanism)**.** *If players are distributed i.i.d. according to a regular distribution $F$, then the optimal mechanism asks from players to report abilities, assigns rank 0 to any player with ability $v_i \leq \eta$, where $\eta$ is the solution to the equation $\phi(\eta) = 0$ (monopoly reserve), and assigns distinct ranks to the rest of the players in decreasing order of their abilities. Last it asks from each player to submit an output:*

$$b_i(v_i) = v_i \tilde{x}(v_i) - \int_0^{v_i} \tilde{x}(z)\, dz \tag{18}$$

*where*

$$\tilde{x}(v_i) = \sum_{\nu=0}^{n-1} S\left(\frac{\nu+1}{n}\right) \beta_{\nu,n-1}(1 - F(v_i)) \tag{19}$$

Observe that the interim allocation of a player under a badge mechanism is equal to the interim allocation of the lowest ability player in his badge under the optimal mechanism. Thus the interim allocation of a badge mechanism can be thought of as a downwards rounding of the interim allocation of the optimal mechanism.

Equivalently, the optimal mechanism can be described in the quantile space, as ordering players in increasing quantile after discarding player with quantile higher than the quantile $\kappa^*$ corresponding to the monopoly reserve $\eta$ and asking a player with quantile $q_i$ to submit a bid of:

$$b_i(q_i) = v(q_i) \cdot \hat{x}(q_i) - \int_1^{q_i} \hat{x}(z)\, dz \tag{20}$$

where

$$\hat{x}(q_i) = \sum_{\nu=0}^{n-1} S\left(\frac{\nu+1}{n}\right) \beta_{\nu,n-1}(q_i) \tag{21}$$

for any $q_i \leq \kappa^*$ and 0 otherwise. The revenue of the mechanism can be computed through the use of the revenue function $R(x) = x \cdot v(x)$:

$$\mathrm{E}_v \left[ \sum_i b_i(v_i) \right] = n \int_0^{\kappa^*} R'(q) \cdot \hat{x}(q)\, dq = n \left( R(\kappa^*) \cdot \hat{x}(\kappa^*) - \int_0^{\kappa^*} R(q) \cdot \hat{x}'(q)\, dq \right) \tag{22}$$

## 6. Approximating Optimal Revenue with Small Number of Badges

In this section, we show that simple badge mechanisms can achieve a constant approximation of the output generated by the optimal mechanism, under very generic assumptions on the social status function $S(\cdot) : [0,1] \to \mathbb{R}^+$. We break our analysis based on the convexity of the status function. We show that when social status is a concave function of the proportion of players ranked at least as high (i.e. the marginal increase in status decreases as you beat more and more players), then a single badge can achieve a 4-approximation to the optimal mechanism. On the contrary, when status is a convex function we show examples where a logarithmic in the number of players loss is necessary for any fixed number of badges. We also provide logarithmic upper bounds for several natural classes of non-concave status functions and give a generic upper bound that applies to any status function, stating that a number of badges that is logarithmic in the ratio between the highest and lowest status is sufficient.

6.1. **Decreasing Marginal Status Functions.** Concavity of the status function implies that the marginal increase from beating one extra person decreases as a player is closer to the top. For such status functions we show that a very coarse status discrimination is sufficient to incentivize players to invest close to the optimal effort.

**Theorem 4.** *If the social status function $S(\cdot)$ is concave, then a single badge, with quantile threshold equal to half the monopoly quantile, can achieve a 4-approximation to the optimal revenue.*

*Proof.* Let $OPT$ denote the revenue of the optimal mechanism and $APX$ the revenue of the single badge mechanism. A single badge mechanism at quantile $\kappa$ achieves revenue of

$$APX = n \cdot R(\kappa) \cdot \hat{x}(\kappa) \tag{23}$$

We will show that if we set $\kappa = \frac{\kappa^*}{2}$ then $4 \cdot APX \geq OPT$.

First we point out that if the social status function is concave, then $\hat{x}(q)$ is also a concave function. This follows from known facts about binomial distributions and derivatives of Bernstein polynomials. Thus $\hat{x}(q)$ is a decreasing concave function of $q$. From this fact it follows that:

$$\hat{x}\left(\frac{\kappa^*}{2}\right) \geq \frac{\hat{x}(0) + \hat{x}(\kappa^*)}{2} \geq \frac{\hat{x}(0)}{2} \tag{24}$$

Additionally, by regularity of the distribution we know that $R(q)$ is a concave function and additionally, for any $q \leq \kappa^*$, $R(q)$ is increasing (since, $\kappa^*$ is defined as the point where $R'(\kappa^*) = 0$). Additionally, $R(0) = 0$. From this it follows that:

$$R\left(\frac{\kappa^*}{2}\right) \geq \frac{R(\kappa^*)}{2} \tag{25}$$

Combining the two we get that the revenue of the single badge mechanism is at least:

$$APX \geq n \cdot \frac{R(\kappa^*) \cdot \hat{x}(0)}{4} \tag{26}$$

By the fact that $R(q)$ is increasing in the region $[0, \kappa^*]$ we have that $R(q) \leq R(\kappa^*)$. Since $\hat{x}(q)$ is decreasing in $q$ (i.e. $\hat{x}'(q) \leq 0$) we have that: $R(q)\hat{x}'(q) \geq R(\kappa^*)\hat{x}'(q)$. Thus, we can also upper bound the revenue of the optimal mechanism:

$$OPT \leq n\left(R(\kappa^*) \cdot \hat{x}(\kappa^*) - R(\kappa^*) \int_0^{\kappa^*} \hat{x}'(q)\, dq\right) = n \cdot R(\kappa^*) \cdot \hat{x}(0) \tag{27}$$

The theorem then follows. ■

**Example.** To gain some intuition on the latter theorem, we examine the special case where the social status function is simply the proportion of players that a player beats: i.e.

$$S(t) = 1 - t. \tag{28}$$

In this case the interim allocation takes the simple form: $\hat{x}(q) = \frac{n-1}{n}(1-q)$. Thus the revenue of the optimal mechanism becomes:

$$OPT = (n-1)\left(R(\kappa^*)(1-\kappa^*) + \int_0^{\kappa^*} R(q)dq\right) \tag{29}$$

Whilst, the revenue of a single badge mechanism is simply $APX = (n-1) \cdot R(\kappa) \cdot (1-\kappa)$. The proof of the existence of a threshold $\kappa$ that achieves a 4-approximation, has a nice graphical interpretation in this case as depicted in Figure 2. Theorem 4 in this example of a linear status function, boils down to showing that for a concave revenue curve there exist a $\kappa$ such that the shaded area in the right figure is at least 1/4 of the shaded area in the left figure. ■

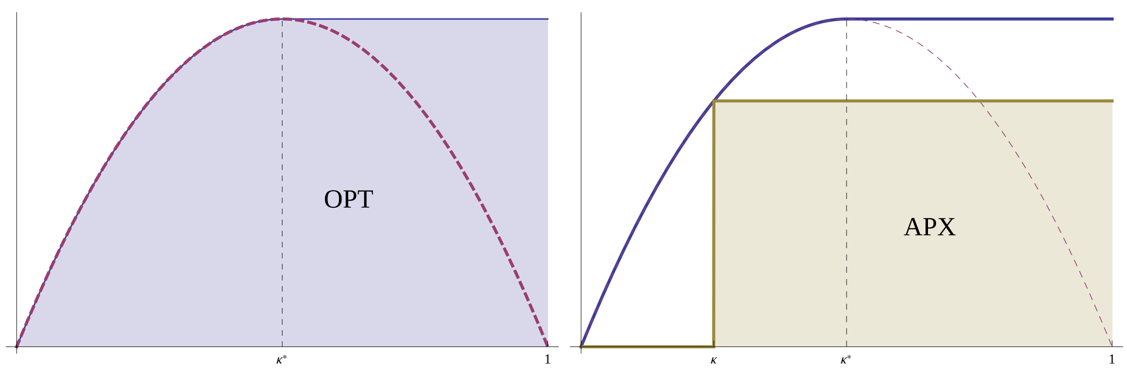


FIGURE 2. Left figure depicts the revenue of the optimal mechanism. The dashed line corresponds to the revenue curve (in this example of the uniform distribution) and the shaded region corresponds to the optimal revenue. The shaded region in the right figure depicts the revenue of a single badge mechanism with quantile threshold $\kappa$.

In fact for this specific status function a tighter 2-approximation can be achieved, via an application of Jensen's inequality and based on the latter graphical interpretation.

**Lemma 5.** *When the status function is linear $S(t) = \alpha \cdot (1-t)$ then a single badge with quantile threshold $\kappa = \min\{\kappa^*, 1/2\}$ achieves a 2-approximation to the optimal revenue.*

*Proof.* Wlog consider the case where $S(t) = 1 - t$. We distinguish two cases. If $\kappa^* \leq 1/2$, then it is easy to see that $OPT \leq (n-1) \cdot R(\kappa^*) \leq (n-1) \cdot R(\kappa^*) \cdot 2 \cdot (1-\kappa^*)$. Thus setting only one badge at the monopoly reserve is sufficient. When $\kappa^* > 1/2$, then we consider the concave curve defined as

$$\hat{R}(q) = \begin{cases} R(q) & q \leq \kappa^* \\ R(\kappa^*) & \text{o.w.}\} \end{cases}$$

Observe that $OPT = (n-1) \cdot \int_0^1 \hat{R}(q)dq$. By concavity of $\hat{R}(q)$ and applying Jensen's inequality we get that:

$$\int_0^1 \hat{R}(q)dq \leq \hat{R}(1/2) = R(1/2) \tag{30}$$

where in the last equality we used the fact that $\kappa^* > 1/2$ and thereby $R(1/2) < R(\kappa^*)$. Thus we get that: $OPT \leq (n-1)R(1/2)$. A single badge mechanism with quantile thresholds at $1/2$ gets revenue $(n-1)\cdot R(1/2)/2$. Thus a single badge mechanism with quantile $\kappa = \min\{\kappa^*, 1/2\}$ yields a 2-approximation to the optimal mechanism in any case. ∎

A graphical depiction of the latter 2-approximation theorem is given in Figure 3. The latter theorem can also be easily generalized to show that if $m$ badges are used, then a $\frac{m+1}{m-1}$ approximation is achieved.[4]

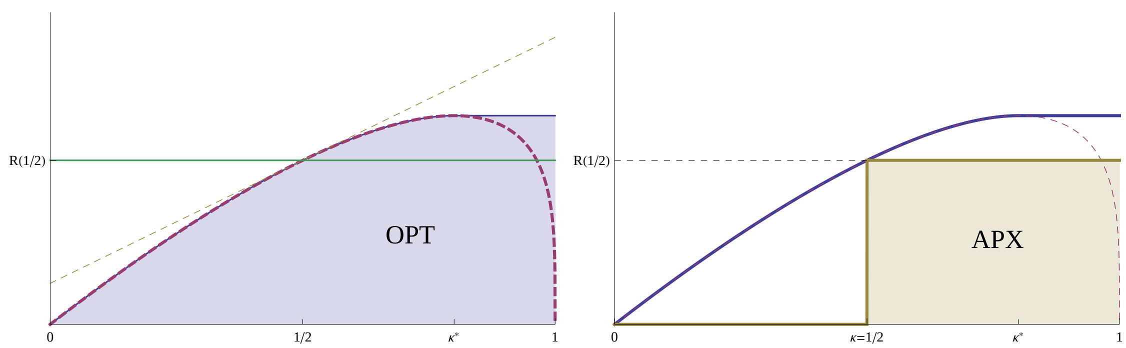


FIGURE 3. Left figure depicts the revenue of the optimal mechanism. Observe that the rectangle with height $R(1/2)$ is at least as large as the are below the concave curve, since the curve lies below the tangent at $1/2$. The revenue of a single badge at $\kappa = 1/2$ is depicted on the right and is half of the latter rectangle.

6.2. **Increasing Marginal Status.** In this section we analyze the revenue of badge mechanisms for settings where status has an increasing marginal behavior. The prototypical such status function that we will analyze is the case where the status of a player is inversely proportional to the proportion of players ranked at least as high:

$$S(t) = \frac{1}{t} \tag{31}$$

We start by portraying that for such a function a logarithmic in the number of players loss is necessary if the number of badges is constant.

[4]Consider iteratively placing badges at the center of the intervals defined by the previous set of badges and the points 0 and 1. After adding such a set of badges, by the same reasoning as in Lemma 5, the remnant of the optimal revenue that is not accounted for in the badge mechanism is reduced by half. Thus after $t$ such iterations we get that the resulting badge mechanism gets a $1 - \frac{1}{2^t}$ fraction of the optimal. Observe that if at step $t$ we had $m_t$ badges, then at step $t+1$ we have $2m_t + 1$ badges. Since $m_1 = 1$, we get that at step $t$ we have $2^t - 1$ badges. Thus if we use $m \in [2^t - 1, 2^{t+1} - 1)$ badges we can achieve $1 - \frac{1}{2^t} \geq 1 - \frac{2}{m+1}$ fraction of the optimal.

**Example.** (Logarithmic Loss) Consider the case where ability is distributed uniformly in $[0,1]$. The revenue function is then $R(q) = q(1-q)$, with derivative $R'(q) = 1 - 2 \cdot q$ and the monopoly quantile is $1/2$. Thereby the optimal revenue is:

$$\begin{aligned}\frac{OPT}{n} &= \int_0^{1/2} (1 - 2 \cdot q)\hat{x}(q)\, dq = \int_0^{1/2} (1 - 2 \cdot q)\frac{1-(1-q)^n}{q} dq \\ &\geq \int_0^{1/2} (1-2q)\frac{1-e^{-nq}}{q} dq \\ &= \log(n) + \frac{2\left(1 - e^{-\frac{n}{2}}\right)}{n} + \Gamma\left(0, \frac{n}{2}\right) + \gamma - 1 - \log(2) \\ &\geq \log(n) + \gamma - 1 - \log(2) = \Theta(\log(n))\end{aligned}$$

where $\gamma \approx .57$ is the Euler-Mascheroni constant and $\Gamma\left(0, \frac{n}{2}\right) = \int_{n/2}^{\infty} \frac{e^{-t}}{t} dt \to 0$.

On the other hand we note that the maximum social welfare achievable by any mechanism that uses $m$ badges is at most $n \cdot m$. For any bid profile $b$, the social welfare from any badge mechanism with $m$ badges is simply:

$$\sum_{t=1}^{m} \frac{|i : r_i(b) = t|}{|i : r_i(b) \geq t|/n} \leq \sum_{t=1}^{m} n = n \cdot m \tag{32}$$

Therefore, it trivially follows that the optimal revenue achievable with $m$ badges is at most $n \cdot m$. Thus as $n \to \infty$ the fraction of the optimal revenue achievable with $m$ badges converges to $\log(n)/m$. ■

To complement the latter example we show that if the mechanism designer sets $\log(n) + 1$ badges in a particular way, then the expected total output produced by agents is at least half of the total output in the optimal mechanism. Thereby, a logarithmic number of badges is not only necessary but is also sufficient for a constant approximation. Our analysis makes use of the *revenue equivalence principle*: if two mechanisms induce the same interim allocation function, then their revenue is the same. We achieve our main result by constructing a set of $\log(n) + 1$ badges such that each player's status value is at least half of their status value in the optimal mechanism. Combining this property with the revenue equivalence principles yields that our badge mechanism is a 2-approximation to the optimal mechanism.

**Theorem 6.** *If $S(t) = \frac{1}{t}$, the badge mechanism defined by setting quantile thresholds*

$$\kappa = (\kappa^*, \frac{\kappa^*}{2}, \ldots, \frac{\kappa^*}{2^{\log n}}) \tag{33}$$

*achieves at least half of the revenue of the optimal ranking mechanism.*

*Proof.* We first show that, for any player, the interim allocation in this badge mechanism is at least half of the interim allocation in the optimal mechanism. Thus by revenue equivalence, this badge mechanism achieves at least half of the revenue of the optimal mechanism.

Consider any quantile $q_i$. From corollary 3, we know the interim allocation function of the optimal mechanism is equation 19. Rewriting the interim allocation in quantile space yields that the interim allocation for a player of quantile $q_i$ is:

$$\hat{x}(q_i) = \begin{cases} \frac{1-(1-q_i)^n}{q_i} & \text{if } q_i \leq \kappa^* \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

Now we examine the interim allocation of the badge mechanism under quantile thresholds $\kappa = (\kappa^*, \frac{\kappa^*}{2}, \ldots, \frac{\kappa^*}{2^{\log n+1}})$. Denote this interim allocation with $\hat{x}^\kappa(q_i)$. If $q_i > \kappa^*$ then $\hat{x}^\kappa(q_i) = \hat{x}(q_i) = 0$. If $q_i \in [\kappa_j, \kappa_{j+1}) = \left[\frac{\kappa^*}{2^{j-1}}, \frac{\kappa^*}{2^j}\right)$ for some $j \in [m-1]$ then observe that the interim allocation under the badge mechanism is equal to the interim allocation of a player with quantile $\kappa_j$ in the optimal mechanism.

$$\hat{x}^\kappa(q_i) = \hat{x}(\kappa_j) = \frac{1-(1-\kappa_j)^n}{\kappa_j} \tag{35}$$

Now observe that $q_i \le \kappa_j \le 2q_i$, which yields:

$$\hat{x}^\kappa(q_i) = \frac{1-(1-\kappa_j)^n}{\kappa_j} \ge \frac{1-(1-q_i)^n}{2\cdot q_i} = \frac{\hat{x}(q_i)}{2} \tag{36}$$

Last consider a player $i$ with quantile $q_i \le \kappa_m = \frac{\kappa^*}{2^{\log(n)}} = \frac{\kappa^*}{n} \le \frac{1}{n}$. The interim allocation of such a player under the badge mechanism is:

$$\hat{x}^\kappa(q_i) = \hat{x}(\kappa_m) \ge \hat{x}\left(\frac{1}{n}\right) = n\cdot\left(1-\left(1-\frac{1}{n}\right)^n\right) \ge \frac{n}{2} \ge \frac{1}{2}\hat{x}(q_i) \tag{37}$$

Where we used the fact that the interim allocation is non-decreasing in ability and hence non-increasing in the quantile in the first inequality and the fact that $\hat{x}(q_i) \le n$ in the last inequality.

Thus we showed that for any quantile $q_i$, the interim allocation of player $i$ under the badge mechanism is at least half of his interim allocation under the optimal ranking mechanism.

$$REV_{badge} = \mathrm{E}\left[\sum_i \hat{x}^\kappa(q_i)\phi(v(q_i))\right] \ge \frac{1}{2}\mathrm{E}\left[\sum_i \hat{x}(q_i)\,\phi(v(q_i))\right] = \frac{1}{2}REV_{opt}$$

■

Figure 4 portrays the relation between the optimal and the badge interim allocation probabilities. Using similar analysis we can also show that the transition to approximate optimality

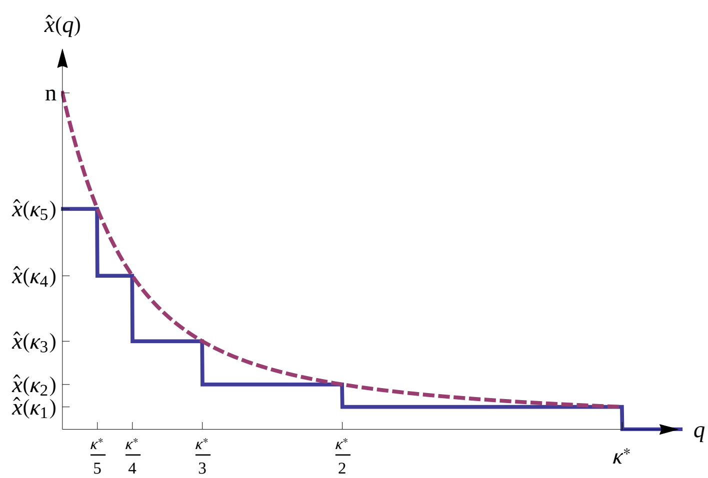


FIGURE 4. Optimal (dashed) versus badge mechanism (solid) interim allocation probabilities, with respect to quantiles.

is smooth. That is if the designer uses $m = \frac{\log n}{c} + 1$ badges then there exists a badge mechanism that achieves a $2^c$ approximation. This can be achieved by setting quantile thresholds of the form $(\kappa^*, \frac{\kappa^*}{t}, \frac{\kappa^*}{t^2}, \ldots, \frac{\kappa^*}{t^{m-1}})$, for $t = 2^{\frac{\log(n)}{m-1}} = 2^c$.

The above theorem can also be generalized for arbitrary social status functions, as follows.

**Theorem 7.** *The optimal badge mechanism with* $\log\left(\frac{S(1/n)}{S(1)}\right)$ *badges, achieves at least half of the optimal ranking mechanism.*

## 7. Structural Stability of Badge Mechanisms in Large Markets

We show that for a sufficiently large number of players the unique Bayes-Nash equilibrium of a badge mechanism with a fixed number of badges with quantile thresholds $\kappa_1, \ldots, \kappa_m$ becomes an ex-post equilibrium. The latter also has implications on the extensive robustness of the Bayes-Nash equilibrium to extensive form perturbations of the game, by applying the results of [11].

Our result implies that the single-badge approximate mechanism that we considered in the previous section, results in an approximate ex-post equilibrium for sufficiently large markets and hence becomes extensively robust. Formally, we use the following definition of an approximate ex-post equilibrium and of extensive robustness.

**Definition 8** ([11]). *A mixed Bayes-Nash equilibrium is* $(\epsilon, \rho)$ *ex-post Nash if with probability at least* $1-\rho$*, no player can improve his utility by more than additively* $\epsilon$*, after the instantiation of the types and strategies.*

Kalai [11] shows that if the Bayes-Nash equilibrium becomes an ex-post equilibrium at an exponential rate in the number of players then this implies that the Bayes-Nash equilibrium is extensively robust, in the following sense. Consider extensive form versions of the game where players take potentially many turns to play and revise their strategy and where information about players types can be arbitrarily revealed in between rounds. The sole assumption required is that for every possible strategy of the simultaneous game, a player has the option in the extensive form version to play that strategy at the first time he is allowed to play and stick to it for the rest of the game. A Bayes-Nash equilibrium of the simultaneous game is extensively robust, if the strategy of playing according to that equilibrium in the first round you are allowed to play and sticking to the same action for all subsequent plays, is a Bayes-Nash equilibrium of any extensive form version of the game as described above. It is $(\epsilon, \rho)$ extensively robust, if with probability $1-\rho$ no player has an $\epsilon$ improving deviation at any information set of the extensive form game. Kalai [11] shows that if a Bayes-Nash equilibrium becomes an $(\epsilon, \rho)$ ex-post equilibrium, with $\rho$ being exponential in the number of players, then it is also an $(\epsilon, \rho')$ extensively robust with $\rho'$ also decaying exponentially in the number of players.

To prove our result we need to adapt the techniques of Kalai [11] to our setting. Kalai [11] considers only settings with a finite set of actions and types and with continuous utilities. Our setting involves a continuum of types, a continuum of strategies and at a first glance a discontinuous utility. However, due to the monotonicity of the equilibrium, and the finiteness of the badges, the number of undomianted strategies is finite and the number of types is also essentially finite, since all that characterizes a players behavior is whether his type falls in one of finitely many ability intervals. Additionally, we can view the utility of a player as a function of the proportion of players that congest each badge, i.e. the proportion of players that fall within each ability interval. Under such a perspective the utility of a player is a continuous function of this proportions, for any fixed action of the player, due to the continuity of the status function. The latter allows us to apply techniques similar to those in [11] to show an exponential convergence to an ex-post Nash equilibrium.

**Lemma 9.** *Consider a badge mechanism with quantile thresholds $\kappa = (\kappa_1, \ldots, \kappa_m)$ and a $c$-Lipschitz status function $S(\cdot)$. For any positive $\epsilon$, there exists constants $\alpha$, and $\beta$ that depend on $\epsilon$ and on the instance of the game, such that any Bayes-Nash of the game with $n$ players is $(\epsilon, \alpha\beta^n)$ ex-post equilibrium.*

*Proof.* Let $X_t^i = 1_{q_i \geq \kappa_t}$ be the indicator variable, of whether player $i$ has quantile above $\kappa_t$. Let $X_t = \frac{\sum_{i\in[n]} X_t^n}{n}$ be the random variable that corresponds to the proportion of players with quantile above $\kappa_t$. Observe that $\Pr[X_t^i = 1] = \kappa_t$ and $\mathrm{E}[X_t] = \kappa_t$. By Chernoff-Hoeffding bounds and following a similar reasoning as in [**?**] we have:

$$\Pr\left[|X_t - \kappa_t| > \delta\right] \leq 2e^{-2\delta^2 n} \tag{38}$$

$$\Pr\left[|X_t - \kappa_t| > \delta \mid v_i\right] \leq 2e^{-2[(n\delta-1)/(n-1)]^2(n-1)} \tag{39}$$

By the union bound we have that the probability that there exists some $t \in [m]$ such that the above event happens is:

$$\Pr\left[\exists t \in [m] : |X_t - \kappa_t| > \delta\right] \leq 2\cdot(m+1)\cdot e^{-2\delta^2 n} \tag{40}$$

$$\Pr\left[\exists t \in [m] : |X_t - \kappa_t| > \delta \mid v_i\right] \leq 2\cdot(m+1)\cdot e^{-2[(n\delta-1)/(n-1)]^2(n-1)} \tag{41}$$

Let $\mathcal{E}(\delta) = \{\forall t \in [m] : |X_t - \kappa_t| \leq \delta\}$ denote the corresponding event and $\mathcal{E}^c(\delta)$ its complement. Also denote with $\rho_n = 2\cdot(m+1)\cdot e^{-2[(n\delta-1)/(n-1)]^2(n-1)}$. Thus we know that $\Pr[\mathcal{E}^c(\delta)] \leq \rho_n$.

Let $X = (X_1, \ldots, X_m)$ and $u_i(t, X; v_i) = S(X_t) - \frac{\kappa_t}{v_i}$, be the utility of player $i$ from bidding to win badge $t$, ex-post and $u_i(t, \kappa\ ;\ v_i) = S(\kappa_t) - \frac{\kappa_t}{v_i}$ be the hypothetical utility if the expectation occurs.

Consider a player $i$ who is winning badge $t$. Consider an ex-post instantiation of player's values where event $\mathcal{E}(\delta)$ holds and let $p = p_1, \ldots, p_t$ be the values of $X_1, \ldots, X_t$. Suppose that player $i$ has a deviation to some other badge $t'$ that yields him an improvement of more than $\epsilon$. Then for any ex-post instantiation $p'$ under which event $\mathcal{E}(\delta)$ happens we know, by $c$-Lipschitz continuity of $f(\cdot)$, that for all $t' \in [m]$:

$$|u_i(t', p'\ ;\ v_i) - u_i(t', p\ ;\ v_i)| = |f(p'_{t'}) - f(p_{t'})| \leq c\cdot\delta \tag{42}$$

Additionally, observe that:

$$\begin{aligned}\big|\left[u_i(t', p\ ;\ v_i) - u_i(t, p\ ;\ v_i)\right] - \left[u_i(t', p'\ ;\ v_i) - u_i(t, p'\ ;\ v_i)\right]\big| \leq \\ |u_i(t', p\ ;\ v_i) - u_i(t', p'\ ;\ v_i)| + |u_i(t, p\ ;\ v_i) - u_i(t, p'\ ;\ v_i)| \leq 2\cdot c\cdot\delta\end{aligned}$$

Thus if $t'$ was an $\epsilon$ profitable deviation under $p$ then it must be at least a $\epsilon - 2\cdot c\cdot\delta$ profitable deviation under $p'$, for any $p'$ in $\mathcal{E}(\delta)$.

When an instance $p' \notin \mathcal{E}(\delta)$ occurs, then deviating from $t$ to $t'$ can lead to a loss of at most: $-(\bar{v}+1)S(1/n)$. The reason is that player $i$ is getting utility at most $S(1/n)$ from any badge $t$ and at least $-S(1/n)\cdot\bar{v}$ by playing any other badge, since badges with $\kappa_t \geq S(1/n)\cdot\bar{v}$ can simply be ignored, since no player will ever consider them.

Thus the expected profit of player $i$ from deviating to $t'$ is:

$$\mathrm{E}[u_i(t', X\ ;\ v_i) - u_i(t, X\ ;\ v_i)] \geq (\epsilon - 2\cdot c\cdot\delta)(1-\rho_n) - \rho_n\cdot(\bar{v}+1)\cdot S(1/n) \tag{43}$$

Since $t$ is an equilibrium for player $i$ it must be that the latter quantity is negative. Thus for $\epsilon > 2 \cdot c \cdot \delta + \frac{\rho_n}{1-\rho_n} \cdot (\bar{v}+1) \cdot S(0)$, we get a contradiction.

Thus we get for any $\delta$, the Bayes-Nash equilibrium is an $(\epsilon, \rho)$ ex-post Nash equilibrium with:

$$\epsilon > 2 \cdot c \cdot \delta + \frac{\rho_n}{1-\rho_n} \cdot (\bar{v}+1) \cdot S(1/n) \tag{44}$$

$$\rho > 2 \cdot (m+1) \cdot e^{-2\delta^2 n} \tag{45}$$

For sufficiently small $\delta$ and sufficiently large $n$ the first inequality is always satisfied, yielding the Lemma. ■

## 8. Extensions and Further Directions

8.1. **Addition of Badges and Revenue Monotonicity.** A natural question is whether an extra badge at some prefixed quantile threshold can actually hurt revenue. We show that this cannot happen as long as the quantile threshold of this extra badge is above the monopoly quantile of the distribution.

**Lemma 10.** *Adding an extra badge with quantile threshold $\tilde{\kappa} \geq \kappa^*$ can only increase the revenue of an existing badge mechanism.*

*Proof.* Consider augmenting the badge mechanism with an extra badge with quantile threshold $\tilde{\kappa}$ and such that $\tilde{\kappa} \in [\kappa_{t+1}, \kappa_t]$ for some $t \in [1, m]$, with the convention that $\kappa_{m+1} = 0$ and $\kappa_0 = 1$. Then the revenue of this new mechanism is:

$$REV_{\kappa+\tilde{\kappa}} = \sum_{\sigma \neq t} (R(\kappa_t) - R(\kappa_{t+1}) \cdot \hat{x}(\kappa_t) + (R(\kappa_t) - R(\tilde{\kappa})) \cdot \hat{x}(\kappa_t) + (R(\tilde{\kappa}) - R(\kappa_{t+1})) \cdot \hat{x}(\tilde{\kappa})$$

By regularity of the distribution, we know that $R(\cdot)$ is monotone in the region $[0, \kappa^*]$. Since we assumed that $\tilde{\kappa} \leq \kappa^*$ we get that $R(\tilde{\kappa}) \geq R(\kappa_{t+1})$. By monotonicity of $\hat{x}(\cdot)$ we also have: $\hat{x}(\tilde{\kappa}) \geq \hat{x}(\kappa_t)$. Combining the above we get the Lemma:

$$\begin{aligned} REV_{\kappa+\tilde{\kappa}} \geq & \sum_{\sigma \neq t} (R(\kappa_t) - R(\kappa_{t+1}) \cdot \hat{x}(\kappa_t) + (R(\kappa_t) - R(\tilde{\kappa})) \cdot \hat{x}(\kappa_t) + (R(\tilde{\kappa}) - R(\kappa_{t+1})) \cdot \hat{x}(\kappa_t) \\ = & \sum_{\sigma \neq t} (R(\kappa_t) - R(\kappa_{t+1}) \cdot \hat{x}(\kappa_t) + (R(\kappa_t) - R(\kappa_{t+1})) \cdot \hat{x}(\kappa_t) = REV_\kappa \end{aligned}$$

■

8.2. **Convex Effort Costs.** We consider the generalization where each players cost for investing some effort $e_i$ is a convex function of $e_i$, rather than linear, that is fixed and known to the designer:

$$\hat{u}_i(r, e_i) = x_i(r) - e_i^\alpha \tag{46}$$

for some $\alpha \geq 1$, that is common knowledge. For this setting, we can still reformulate the problem as a single-parameter mechanism design problem with quasilinear utilities. The difference is that we will consider the translated utility:

$$u_i(b; v_i) = v_i^\alpha \hat{u}_i(r, e_i) = v_i^\alpha x_i(r(b)) - b_i^\alpha \tag{47}$$

The crucial part in achieving this translation is that the cost of a player depends only on his own bid and is not an expectation over other players values. Thus we can do a completely analogous analysis as in the quasi-linear utility model. The only difference will occur when translating the ability thresholds to badge thresholds, where we would replace $\theta_j$ with $\theta_j^\alpha$ in all the related Equations (8) and (9).

## References


[1] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Steering user behavior with badges. In *Proceedings of the 22nd international conference on World Wide Web*, pages 95–106. International World Wide Web Conferences Steering Committee, 2013.

[2] Jeremy Bulow and John Roberts. The simple economics of optimal auctions. *Journal of Political Economy*, 97(5):pp. 1060–1090, 1989.

[3] Ruggiero Cavallo and Shaili Jain. Efficient crowdsourcing contests. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 677–686. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[4] Shuchi Chawla, Jason D Hartline, and Balasubramanian Sivan. Optimal crowdsourcing contests. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 856–868. SIAM, 2012.

[5] Pradeep Dubey and John Geanakoplos. Grading exams: 100, 99, 98, or a, b, c? *Games and Economic Behavior*, 69(1):72–94, 2010.

[6] David Easley and Arpita Ghosh. Incentives, gamification, and game theory: An economic approach to badge design. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pages 359–376. ACM, 2013.

[7] Arpita Ghosh and Patrick Hummel. Implementing optimal outcomes in social computing: a game-theoretic approach. In *Proceedings of the 21st international conference on World Wide Web*, pages 539–548. ACM, 2012.

[8] Jason Hartline. *Approximation in Economic Design*. 2012.

[9] Ed Hopkins and Tatiana Kornienko. Running to keep in the same place: Consumer choice as a game of status. *American Economic Review*, pages 1085–1107, 2004.

[10] Shaili Jain, Yiling Chen, and David C Parkes. Designing incentives for online question-and-answer forums. *Games and Economic Behavior*, 2012.

[11] Ehud Kalai. Large robust games. *Econometrica*, 72(6):pp. 1631–1665, 2004.

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

[13] R. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.