Incentivizing High-quality User-Generated Content

Arpita Ghosh
Yahoo! Research
Santa Clara, CA, USA
arpita@yahoo-inc.com

Preston McAfee
Yahoo! Research
Burbank, CA, USA
mcafee@yahoo-inc.com

ABSTRACT

We model the economics of incentivizing high-quality user generated content (UGC), motivated by settings such as online review forums, question-answer sites, and comments on news articles and blogs. We provide a game-theoretic model within which to study the problem of incentivizing high quality UGC, in which contributors are strategic and motivated by exposure. Our model has the feature that both the quality of contributions *as well as* the extent of participation is determined endogenously in a free-entry Nash equilibrium.

The model predicts, as observed in practice, that if exposure is independent of quality, there will be a flood of low quality contributions in equilibrium. An ideal mechanism in this context would elicit both high quality and high participation in equilibrium, with near-optimal quality as the available attention diverges, and should be easily implementable in practice. We consider a very simple elimination mechanism, which subjects each contribution to rating by some number A of viewers, and eliminates any contributions that are not uniformly rated positively. We construct and analyze free-entry Nash equilibria for this mechanism, and show that A can be chosen to achieve quality that tends to optimal, along with diverging participation, as the number of viewers diverges.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social & Behavioral Sciences

General Terms

Economics

Keywords

User generated content (UGC), Attention economics, Quality of online content, Game theory

1. INTRODUCTION

User generated content (UGC) refers to a very wide spectrum of online content generated by end-users, ranging all the way from collaborative information sites like Wikipedia, to question-answer forums and discussion boards like Y!Answers and TripAdvisor, to reviews– product reviews on Amazon or movie reviews on IMDB and Y!Movies, or Yelp reviews– all the way to content on blogs and social media like Facebook, as well as comments on blogs and news articles as on Slashdot or Digg. On the one hand, user generated content constitutes a large fraction of the high-quality, easily accessible content that makes the Web useful. On the other hand, because there is no barrier to entry unlike traditional publishing, it also attracts junk and spam.

The extent to which quality of contributions is an issue varies widely across the broad spectrum of user generated content: Wikipedia and some other crowdsourcing applications seem to attract largely high-quality content, while on social media applications like Facebook and Twitter users can control and personalize the source of the content they see. The quality issue is more pertinent in the context of reviews, such as product or service reviews, comments on news articles or blog posts, and responses in discussion boards or question-answer forums, which are more prone to poor quality content, and where users have little control over what content they see. Why do some websites have great user generated content (UGC), while others have awful content? What incentives should a website manager employ to create great UGC?

In this paper, we address this question from a game-theoretic perspective. We employ a model with strategic contributors, where the primary motivator of contributors is *exposure*, or viewer attention– a contributor is motivated by the amount of exposure her content will receive[1]. Without some connection between quality of a contribution and amount of exposure, such exposure-motivated contributors will flood a site with low quality contributions, as is indeed observed in practice. Is there a way to allocate the available attention from viewers amongst the contributions– a mechanism– that encourages high-quality contributions, while also maintaining a high level of participation? An ideal mechanism is one which produces only high quality equilibria with high levels of participation, *and* is easily implementable in practice– for instance, eliciting high quality equilibria should not rely upon having a very precise estimate of quality. In fact, as the number of viewers increases, increasing the amount

[1]Note that this does not rely on any assumption about *why* contributors seek greater exposure– it could be because of a type of altruism (a contributor believes her contribution is valuable, and wants it to benefit the largest number), or a contributor seeking the largest possible audience for self-expression, or a host of other social-psychological factors.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0632-4/11/03.

of available attention, is there a simple mechanism that elicits a near optimal quality of contributions in equilibrium?

Our Contributions. Our primary contribution is providing a game-theoretic model within which the design and performance of mechanisms for incentivizing high-quality user generated content can be analyzed. Our model has the feature that both the *quality* of contributions, as well as the *number* of contributors are endogenously determined in a free-entry Nash equilibrium (§2). We emphasize the endogenous choice to contribute– the problem of producing UGC is not just the creation of high quality content, but also *encouraging* the production of content, and a model with exogenous participation misses a salient factor in most UGC settings. While we restrict ourselves to a simple cost-benefit function for contributors in this paper, the general framework we introduce could be useful in capturing a much broader class of contributor motivations and design goals for user generated content, see §7.

Our model predicts that unless the extent of exposure is determined by the quality of a contribution there will be a large number of low-quality contributions in a free-entry equilibrium. A natural candidate mechanism based on quality is a proportional mechanism, where the extent of exposure (or the fraction of viewers who see the contribution) is proportional to the rated quality of the contribution. We show that these mechanisms improve over showing all contributions equally, but are unable to produce very high quality even as the number of potential viewers diverges, in addition to requiring precise estimates of quality (§3). This motivates us to analyze simple *elimination* mechanisms, which subject contributions to rating by some small number of viewers and reject contributions that are not uniformly highly rated. We construct and analyze free entry Nash equilibria for these mechanisms (§4.1) and show that there is a tradeoff between the quality of contributions and the number of contributions, which can be controlled by varying the stringency of the test of inclusion, that is, the number of viewers who must rate the contribution as high quality (§4.2). In §4.3, we show that as the number of viewers diverges, we can achieve both optimal quality and a diverging number of contributions in equilibrium in the limit, using these elimination mechanisms. Finally, we show that we can do even better in settings where limiting contributions might be acceptable: we prove that the elimination mechanism combined with restricting the number of displayed contributions permits the production of a large number of near optimal contributions while also sending the fraction of rejected content to zero in the limit.

Related Work. An entirely different approach to the design of human computation systems and user generated content is based on identifying the social psychological motives, also called incentives, of contributors in these systems and designing user interfaces and systems to better reward those motives to encourage participation. There is a vast literature in social psychology that employs this approach, see for example [6, 7]. A game-theoretic approach complements this approach, and can also benefit from this literature in appropriately modeling cost and benefit functions for contributors.

There is a much smaller literature that addresses user generated content and crowdsourcing from a game theoretic perspective, see, for instance, [3, 2, 4]. To the best of our knowledge, this literature has thus far largely addressed specific systems or websites; also, none of this work addresses the issue of *quality* of user generated content, particularly in a general framework.

There is a large and growing literature on incentives in online reputation systems; for a survey, see [1], also [5]. Our work differs from this literature on reputation systems in one important respect: we are concerned about rating the quality of each individual piece of *content*, not the *user* that produced it. On the one hand, our approach therefore does not use all the possible information available to rate a contribution (specifically, the identity of the creator). On the other hand, independently judging each piece of content irrespective of the identity of the user who produced it helps us avoid some of the most important problems in the context of reputation systems, such as sybil-proofing, white-washing or cheaply creating new identities to shed bad reputation, building up and then cashing in on high reputations, hijacked reputations, and the thorny problem of translating expertise from one topic to another. Thus our work complements the reputation system literature.

2. A MODEL

The main components of our model are as follows. There are *contributors*, who produce content (such as a review for a product or a comment on a news article), and *viewers*, who consume, and possibly rate, the content. The generated content is of varying *quality*, the 'goodness' of which can be described by a one-dimensional value. Contributors derive value from receiving viewer attention for their content– which increases with the amount of attention received, and incur a cost to producing content, which is increasing in quality (see §7 for a discussion). Contributors are strategic: they make the decision of whether or not to participate, *i.e.*, whether to generate content, as well as what quality content to produce, selfishly to maximize their utility. Viewers are non-strategic and simply provide utility to contributors and possibly rate their content.

A mechanism in our setting is a way to distribute the available viewer attention amongst the participating contributors, possibly based on the number of participating contributors as well as the quality of each contribution. We will be interested in simple, easily implementable, mechanisms that incentivize high-quality participation by contributors. We now formally describe the model.

Content Model. We model the goodness of each contribution, or unit of content, with a quality q. The value q is measured in units such that the probability that a viewer would answer the question "Is this good content?" is q. This means the quality $q \in [0, 1]$.

Since the quality of a contribution q is defined to be the probability that a random viewer likes it, we can use viewer votes (such as the standard 'Thumbs Up/Down' button) to estimate q, so mechanisms incorporating the quality q can be implemented (to different degrees of accuracy) using viewer votes.

Contributor Model. There is an infinite[2] pool of potential contributors, although the number of contributors who

[2]The equilibrium number of contributors will never be larger than the number of viewers, as we will see, so infinite really only means as many potential contributors as viewers, which

actually decide to generate content, *i.e.*, the number of participating contributors K, will be some finite number that is *endogenously* determined by whether or not there are any remaining profits from participation. We emphasize the endogenous choice to contribute– our model does *not* assume a fixed number of contributors; rather, the number of contributors is also determined by strategic choices based on the available benefit from participation.

Contributors decide whether or not to participate, and the quality q of the content they will produce if participating.

The *cost* incurred by a contributor depends on the quality of content she chooses to produce: a contributor can produce content of quality q at a cost $c(q)$, which is increasing in q (*i.e.*, producing higher quality content is more costly). We make the following assumptions on the cost function c:

1. $c(0) > 0$: Participating, even with 0 quality content, is strictly costlier than not participating.
2. $\lim_{q\to 1} c(q) = \infty$: Perfect quality is infinitely costly, *i.e.*, making every viewer happy is nearly impossible.
3. $c(q)$ and $c'(q)$ are finite for all q strictly less than 1. (We note that this is a weaker condition than the standard assumption of convex cost functions.)

An example function satisfying these conditions is $c(q) = c(1-q)^{-r}$, $r > 0$.

Contributors derive *benefit* from receiving attention from viewers; we assume that this benefit is directly proportional to, and in fact without loss of generality (by appropriately scaling the cost functions), exactly equal to the amount of attention received. Let K be the number of participating contributors, producing content of quality $q_1, q_2, \ldots, q_K$. Suppose there are M viewers. The distribution of the total available attention from these M viewers amongst the K participating contributors is determined by the mechanism $\mathcal{M}$ being used to display the content. The benefit derived by a contributor i is exactly its allocation or share of the available attention awarded to it by the mechanism, which we denote by $V(q_i, q_{-i})$. It is reasonable to assume that V is non-increasing in K, *i.e.*, the share of attention does not increase (holding quality constant) if the total number of competing contributors increases.

A contributor i's utility from generating content of quality q_i is the difference between her benefit, which also depends on the number and possibly the quality of other contributions q_{-i}, and her cost:

$$\pi_i(q_i, q_{-i}) = V_i(q_i, q_{-i}) - c(q_i).$$

Solution Concept. We will be interested in the levels of participation and qualities that occur in an equilibrium. However, we need to be careful about the notion of equilibrium we use – since we did not assume a fixed number of participating contributors, the equilibrium concept will need to account for the fact that *participation is endogenously determined*. We use the solution concept of a *free entry Nash equilibrium*: as in a standard Nash equilibrium, the quality chosen by each participating contributor must be a best response to the qualities chosen by the remaining participants. In addition, because the choice of *whether* to participate or not (which is distinct from participating with $q = 0$ because $c(0) > 0$) is a strategic choice as well, the number of participants K must be such that while each current participant derives a nonnegative payoff, no further contributors can benefit from joining. Consider a symmetric equilibrium, where the only relevant quantities are the number of participants K and their common chosen quality q^*[3]. Then, K must be such that the profits with K participants $\pi(q^*(K), K) \geq 0$, but $\pi(q^*(K+1), K+1) \leq 0$.

For simplicity, we use the continuous version of this definition of the number of participants K: instead of finding an integer K, we will simply solve for K satisfying $\pi(q^*, K) = 0$ (which need not be an integer), and choose q^* to be the best response to this 'number' of participants[4].

Definition 2.1 (Free entry Nash equilibrium). *Fix N. Define a "free entry Nash equilibrium in pure strategies" to be values (K, q^*) so that*

- *q^* maximizes $\pi(q, q^*, K)$: q^* is a Nash equilibrium quality choice given the number of participants K and their chosen qualities q^*,i.e., q^* is a best response to $K-1$ other participants contributing content of quality q^*.*
- *$\pi(q^*, K) = 0$: The values (K, q^*) are such that each participant's payoff from participating is zero, i.e., it is individually rational for each of the K contributors to participate, but no more potential contributors can derive a positive payoff by participating (recall that we assumed that V_i, the share of attention to the ith contributor is decreasing in K).*

That is, a free entry Nash equilibrium consists of an endogenously determined number of contributors K and a quality q^* such that no more contributors want to participate, and no participating contributor wants to deviate from her chosen quality (or from her decision to participate).

3. MECHANISM CHOICE AND EQUILIBRIUM QUALITY

We now investigate the impact of the choice of mechanism on qualities generated in equilibrium.

First consider the situation in many UGC forums, which display all comments or reviews irrespective of quality. A simple model for this is to say that contributors are assigned at random to receive views from the M viewers.[5]. If K contributors provide content, each contributor receives an equal share of attention, which is M/K. An individual

is precisely the situation in UGC because of open participation.

[3]While equilibria for arbitrary mechanisms that depend in an arbitrary way upon the qualities can indeed be asymmetric, we will focus here on symmetric equilibria where all participating contributors choose the same quality. In fact, for all but one mechanism that we consider, the only possible equilibria are indeed symmetric; we discuss asymmetric equilibria separately for that mechanism.

[4]The actual number of participants corresponding to this equilibrium will be the nearest integer smaller than K, since V_i decreases with K; the equilibrium quality q^* is the best response for $\lfloor K \rfloor$, rather than K, participants. We use the continuous version since integrality of K will be unimportant for us.

[5]Actual sites vary in how content is displayed– some sites display oldest first, while others show newest first. We model display order as random purely for simplicity, and because it captures independence from quality.

contributor receives payoff $M/K - c(q)$, which is uniquely maximized by setting $q = 0$. The total number of contributions at which contributors are indifferent to joining satisfies $M/K - c(0) = 0$, or $K = M/c(0)$ (*i.e.*, the $K+1$th contributor does not want to join because his payoff $\frac{M}{K+1} - c(0)$ will be negative).

Thus this baseline mechanism has a unique equilibrium with $q^* = 0$, and $K = \frac{M}{c(0)}$: it produces the maximum possible number of contributions under individual rationality, all with the worst possible quality $q^* = 0$, in equilibrium.

A Proportional Mechanism. The baseline mechanism demonstrates that ignoring the quality of contributions will lead to poor equilibrium qualities. To induce high quality, therefore, the quality of a contribution must be used to decide what share of the attention it receives.

Some sites display contributions in decreasing order of estimated quality based on viewer ratings. It is reasonable to assume that viewers scan the page from top to bottom, so that contributions higher up in the order receive more attention. This suggests a natural candidate mechanism for analysis, namely a *proportional* mechanism: suppose we have some way to divide the total attention from the M viewers proportional to the quality of each contribution q_i. The proportional mechanism gives the ith contribution a share $\frac{q_i}{\sum_{j=1}^{K} q_j}$ of the M units of attention.

As we show below, the proportional mechanism is indeed an improvement over the baseline and solves the problem of eliminating content with $q = 0$, but the equilibrium quality can never approach 1, even in the limit as the number of viewers goes to infinity:

PROPOSITION 3.1. *Every participant chooses equilibrium quality $q^* > 0$ in the proportional mechanism for all M; however, $\lim_{M\to\infty} q^* < 1$.*

PROOF. Consider a vector of qualities $q_1, \ldots, q_K$; the payoff of the ith contributor is

$$\pi_i(q_i) = M\frac{q_i}{q_i + \sum_{j\neq i} q_j} - c(q_i).$$

First note that with this payoff function, $q_i = 0$ is never a best response to any set of remaining qualities, and therefore will never occur in a (symmetric or asymmetric) equilibrium: if $q_j > 0$ for any $j \neq i$, i's share of M is 0, and her payoff is negative since $c(0) > 0$, so not participating is a profitable deviation. If all contributors choose $q_j = 0$, irrespective of the allocation computed with $q_j = 0\ \forall\ j$, there is some contributor who can strictly improve her payoff by increasing q to ϵ since this increases her share to 1, *i.e.*, get all the available attention, which is clearly a profitable deviation for $M > c(0)$ (if $M < c(0)$, there are no participants, so the statement still holds). Therefore no contributor ever chooses 0 quality in an equilibrium.

Now we prove the second part of the claim. Consider a symmetric equilibrium, where there are K participants each generating content of quality q^*. In a free entry Nash equilibrium, the values K, q^* need to simultaneously satisfy

$$q^* = \arg\max \pi(q) = \arg\max M\frac{q}{q + (K-1)q^*} - c(q);$$

$$M\frac{q^*}{Kq^*} - c(q^*) = 0, \text{ or } c(q^*) = \frac{M}{K}.$$

(The first condition says that no participant can benefit by choosing a quality different from q^* (for instance, by increasing q^* slightly to get a greater share of attention), given that the remaining $K-1$ participants have chosen q^*. The second condition says that a potential $K+1$th participant would derive negative payoff and so would not want to participate; in addition, it is individually rational for each current contributor to participate.)

The first-order condition for q^* to maximize $\pi(q)$ is

$$\left.\frac{\partial \pi_(q)}{\partial q}\right|_{q^*} = \frac{M(K-1)q^*}{K^2 q^{*2}} - c'(q^*) = 0,$$

or

$$q^* c'(q^*) = M(K-1)/K^2.$$

Dividing this by by $c(q^*) = M/K$, we get

$$\frac{q^* c'(q^*)}{c(q^*)} = \frac{(K-1)}{K}.$$

Since $\lim_{q\to 1} c(q) = \infty$, $\lim_{q\to 1} \log c(q) = \infty$ as well. So $c'(q)/c(q)$, which is the derivative of $\log c$, must go to infinity as well, while the right-hand side $(K-1)/K$ does not. So we cannot have $q \to 1$ in an equilibrium of the proportional mechanism. □

The proportional mechanism therefore improves upon the baseline mechanism by disincentivizing $q = 0$, *i.e.*, it eliminates the worst reviews. Ideally, we would like to be able to drive the equilibrium qualities to 1 in the limit as the number of viewers, M, diverges to infinity; however, as we saw above, this cannot be achieved with the proportional mechanism.

In fact, the proportional mechanism has another difficulty: it requires a precise estimation of the qualities q_i, which is challenging because obtaining a precise estimate requires rating by a large number of viewers which itself contributes to the utility, complicating the strategic choices. This difficulty is shared by another natural mechanism which displays only the contributions with the P highest qualities: with perfect observation of quality, this mechanism induces equilibrium qualities $q^* = c^{-1}(M/P)$ which tends to 1 as M diverges. However, obtaining this high-quality equilibrium requires accurate quality estimates and sequential evaluation since otherwise one can cut quality and make higher profit but still be part of the top P. Also practically speaking, enforcing a predetermined, quality-independent limit on the number of displayed entries may not be desirable in all UGC contexts, for instance discussion forums for issues or comments on news, which is another reason to not use such a mechanism.

3.1 An Elimination Mechanism

We now discuss a very simple mechanism $\mathcal{M}_e$ that also uses the quality of content, albeit differently from the proportional mechanism, to determine a contribution's share of the total attention: each contribution is shown to (up to) A viewers who vote yes or no on whether the content is good. A unit of content is displayed only if some number B of the A viewers vote yes. Once displayed, each such qualifying contribution receives an equal share of the available attention; this can be achieved in expectation by showing a different random permutation to each viewer. Note that if none of the contributions qualify, *i.e.*, receive the required B yeses, no content is displayed to any viewers outside of those who voted on the contributions.

Observe that this elimination-based mechanism uses qualities somewhat differently from the proportional mechanism: the proportional mechanism used the relative values of the qualities to decide on the allocation, whereas this mechanism uses the *absolute* quality of each unit of content. So in $\mathcal{M}_p$, every contributor is guaranteed some nonzero share of the attention, but the magnitude of the share depends on the qualities generated by other contributors; in $\mathcal{M}_e$, a contributor gets either no share at all, or a share that only depends on the *number* of other qualifying contributors, and which of these two happens depends only on her own choice of quality. Note that $\mathcal{M}_e$ does not require or rely on a precise estimate of q_i (in fact, the description of $\mathcal{M}_e$ does not directly invoke the q_i values at all).

There are two natural models, corresponding to different relative rates of content generation and viewer voting, within which to analyze the strategic behavior of contributors under the elimination mechanism $\mathcal{M}_e$.

- In the *sequential* model, a contributor can see how many other contributors have participated *and* qualified so far.
- In the *simultaneous* model, a contributor can see how many other contributors have participated so far, but *not* whether they qualified to be displayed or not.

The sequential and simultaneous models capture different settings: sequential entry models the situation where ratings come in quickly relative to the content (for instance, such as the thumbs up/down ratings on comments on Y!News), whereas simultaneous entry models a situation where the viewer ratings come in slowly relative to contributions (for instance, in a question-and-answer forum like Y!Answers, or Yelp reviews). As we will see, this difference in whether or not a contributor knows how many previous contributions will eventually be displayed will affect the strategic choice facing each contributor, and correspondingly the equilibrium qualities and number of participating contributors. However, much of the analysis, as well as the qualitative results, are very similar for these models.

4. ANALYSIS OF $\mathcal{M}_e$

We now analyze equilibrium behavior in the elimination mechanism $\mathcal{M}_e$. For simplicity (and want of space), we do this under the following assumptions:

- $B = A$: A contribution must be approved of by *each* viewer who rates it. This special case of $\mathcal{M}_e$ has a dramatically simpler analysis and as we will see, allows adequate freedom to achieve the properties we need, including sending the equilibrium qualities to 1 in the limit. In practice, setting $B < A$ is a better choice since it is more *robust* to manipulation or error, *i.e.*, can withstand a voter who always votes a contribution negatively either out of malicious intent or random error. While we do not include the analysis here, it is possible to choose $B < A$ to achieve $q^* \to 1$.
- Let there be M viewers in addition to the viewers who vote on content (of which there are no more than A per contribution). We will assume that contributors only derive utility from displaying to these M viewers. This is a particularly reasonable assumption since we will want A, the maximum number of votes required per contribution, to be a vanishingly small fraction of the total number of available viewers; we note though that our analysis and results do extend to including the utility from displaying to these (up to) A viewers in the payoffs as well[6].
- We will assume, to begin with, that the total amount of attention brought to a page is independent of the number of displayed contributions K (specifically, we assume that the M viewers bring a total attention of M). We will generalize our results to a natural distribution of attention within the population in §5, where we allow the amount of attention from these M viewers to grow with K.

In both the sequential and simultaneous models, the probability that a contributor who decides to participate with quality q is chosen to be displayed is q^A. In the sequential model, contributors will choose q optimally under the common belief that that they will be sharing attention with $K-1$ other *contributions* if displayed (recall that only the number, and not the qualities, of other participants affects the final share of attention for a qualifying contribution.) In the simultaneous model, contributors will again choose q optimally, but assuming that there are $K-1$ other *participants* (as opposed to $K-1$ other contributions), of which some random number will pass the A tests and be displayed. The additional information in the sequential model changes the payoffs and strategic behavior since it mitigates the randomness in how many other contributors will be sharing attention relative to the simultaneous model. We begin with deriving the payoffs in both models.

Sequential Entry. In the sequential entry model, a contributor can see, *before* contributing, how many contributors have qualified so far. A free entry Nash equilibrium in the sequential model is a pair of values (K, q^*) where K is the final number of qualifying participants– that is, in this free entry equilibrium, when a contributor arrives, she checks whether there are already K qualifying contributors. If yes, she does not generate any content; if there are fewer than K qualifying contributors, she generates content with quality q^*.

By definition of q, the probability of a contribution qualifying, *i.e.*, receiving a yes vote from each of the A viewers, is q^A. Let M be the number of available viewers, and consider a free entry equilibrium (K, q^*). In this equilibrium, every qualifying contributor receives attention M/K. Thus, the expected payoff of each participating contributor, who qualifies with probability ${q^*}^A$ is

$$\pi_{\text{seq}}(q^*, K) = V_{\text{seq}}(q^*, K) - c(q) = \frac{M{q^*}^A}{K} - c(q).$$

Simultaneous Entry. By definition of q, the probability of a contribution qualifying, *i.e.*, receiving a yes vote from each of the A viewers, is q^A. However, because more

[6]Including this utility only introduces an additional term of $\frac{1-q^A}{1-q} - Aq^A$ in the payoff $\pi(q)$. The only difference is in the assumption on c we need: to prevent $q^* = 0$ from being an optimal solution (*i.e.*, to guarantee an interior optimum), we will require $c(0) > 1$ since otherwise the utility from displaying to the first evaluator itself pays for the cost of producing a zero-quality contribution.

than one contribution may qualify for display to the general population, even a qualifying contribution won't receive all units of attention, but will share it with the remaining qualifying contributions. Let M be the number of available viewers, and suppose $K-1$ other contributors decide to participate with quality q^*. Note that not all of these other $K-1$ contributors will qualify to be displayed: if k out of these qualify, which happens with probability $\binom{K-1}{k}{q^*}^{A^k}(1-{q^*}^A)^{K-1-k}$, our contributor receives $M/(k+1)$ views.

The expected number of qualified views of a contributor with quality q given that the other $K-1$ contributors use quality q^*, $V_{\text{sim}}(q,q^*,K)$, is

$$\begin{aligned} & q^A \sum_{k=0}^{K-1} \frac{M}{k+1}\binom{K-1}{k}{q^*}^{A^k}(1-{q^*}^A)^{K-1-k} \\ = & \frac{Mq^A}{K{q^*}^A}\sum_{k=0}^{K-1}\binom{K}{k+1}{q^*}^{A^{k+1}}(1-{q^*}^A)^{K-1-k} \\ = & \frac{Mq^A}{K{q^*}^A}(1-(1-{q^*}^A)^K). \end{aligned}$$

Thus, the total profits of a contributor with quality q in the simultaneous entry model are

$$\pi_{\text{sim}}(q,q^*,K) = \frac{Mq^A}{K{q^*}^A}(1-(1-{q^*}^A)^K) - c(q). \quad (1)$$

4.1 Equilibrium

We will now show how to construct a free entry Nash equilibrium in pure strategies (K,q^*), given M and A. First, since the profit at a free-entry Nash equilibrium is zero, we must have

$$\pi(q^*,K) = V(q^*,K) - c(q) = 0.$$

In addition, q^* must be a best response given that $K-1$ other (successful) contributors are choosing quality q^*, *i.e.*, q^* must maximize $\pi(q,q^*,K)$ for this K:

$$q^* = \arg\max_q V(q,q^*,K) - c(q).$$

Why should such a pair of values that simultaneously satisfy these two equations exist, and how do we proceed to find them? The standard approach to finding an equilibrium, which is to work with the first-order conditions for q^* to maximize the profit (*i.e.*, satisfy $\pi'(q,q^*,K)|_{q^*}=0$), does not work in our setting: because q^A is convex in q for $A>1$, the payoff functions (e.g. $\frac{M}{K}q^A - c(q)$) need not be concave[7]. Thus, the first-order conditions, while necessary, do not characterize a maximum of the payoff function.

Instead, we develop an alternative strategy, rewriting the payoff function by factoring out the 'convex' term q^A, and writing the remaining factor as the difference between two terms, where the first term depends only on the quality and number of *other* participants, and the second depends only on the quality q chosen by *this* participant. The zero profit and best response conditions together imply that q^* must be such that it minimizes this second term (independent of K) over q, and K must be chosen so that this minimum value is equal to the value of the first term. To the best of our knowledge, this is a new equilibrium construction; the details are in the proof below.

[7]In fact, the payoff function can be guaranteed to have convex regions for large enough A.

THEOREM 4.1. *A free entry Nash equilibrium in pure strategies always exists for each M, A. Further, the set of equilibria (K,q^*) are determined precisely by the minimizers of $q^{-A}c(q)$; there is a unique equilibrium (K,q^*) for each minimizer q^*.*

PROOF. Rewrite the payoff $\pi(q)$ as follows:

$$\pi(q,q^*,K) = V(q,q^*,K) - c(q) = q^A(b(q^*,K) - \psi(q)),$$

where $b = \frac{M}{K{q^*}^A}(1-(1-{q^*}^A)^K)$ in the simultaneous model and $b = \frac{M}{K}$ in the sequential model is independent of q and depends only on the number K and quality choice q^* of *other* contributors.

To ensure that q^* is a best response, it is enough to ensure that for all q, $\pi(q) \le \pi(q^*)$, or

$$\pi(q,q^*,K) = q^A(b(q^*,K)-\psi(q)) \le {q^*}^A(b(q^*,K)-\psi(q^*)) = 0,$$

where the last equality uses the zero profit condition, *i.e.*, (K,q^*) must be such that $\pi(q^*,K)=0$.

Equilibrium Construction. Set

$$q^* = \arg\min \psi(q) = \arg\min q^{-A}c(q), \quad (2)$$

and let K be such that[8]

$$b(q^*,K) - \psi(q^*) = 0.$$

Then, since $b(q^*,K) = \psi(q^*) = \min\psi(q)$, $\psi(q) \ge b$ for all q, or $b(q^*,K) - \psi(q) \le 0$. So (K,q^*) is such that q^* is a best response for each participant, and also these values satisfy the zero profit condition. So (K,q^*) constructed above is a free entry Nash equilibrium in pure strategies.

In the sequential model, K is given as

$$K_{\text{seq}} = \frac{M{q^*}^A}{c(q^*)}. \quad (3)$$

In the simultaneous model, the equilibrium level of participation K is given by the solution of

$$\frac{M}{K_{\text{sim}}}(1-(1-{q^*}^A)^{K_{\text{sim}}}) - c(q^*) = 0 \quad (4)$$

(note that $q^* = \arg\min q^{-A}c(q)$ is strictly greater than 0 since the value of $q^{-A}c(q)$at 0 is ∞ $(c(0)>0)$.

Uniqueness. To see that no other qualities can arise in equilibrium except for minimizers of $q^{-A}c(q)$, observe that for q^* to be a best response, we must have $b \le \psi(q)$ since $q^A \ge 0$. To satisfy the zero profit condition, we must have either $q^*=0$ or $b=\psi(q^*)$. But with $q^*=0$, the payoff $\pi(q^*) = 0 - c(0) < 0$ since $c(0)>0$ irrespective of K. So $q^*=0$ is not an equilibrium quality, since not participating (which has payoff zero) is always a profitable deviation. The second alternative, $b=\psi(q^*)$ corresponds to q^* minimizing $q^{-A}c(q)$.

Finally, to see that K is uniquely determined by q^*,(3) makes it evident that q^* uniquely determines K_{seq} in the sequential model (note that K_{seq} is unique even if there are multiple minimizers of $q^{-A}c(q)$). In the simultaneous model,

[8]Note that the values of M, A may be such that the nearest integer less than K is 0, which corresponds to no participants in equilibrium.

we use Proposition 4.1 below: since $\frac{(1-x^K)}{K}$ is a (strictly) decreasing function of K for $0 < x < 1$, there is a unique solution K_{sim} to (4) given q^*. □

We include the proof of the following property for completeness.

PROPOSITION 4.1. *The function $\frac{1-x^K}{K}$ is a decreasing function of K for any x with $0 < x < 1$.*

PROOF.

$$\frac{\partial}{\partial K}\frac{1-x^K}{K} = \frac{-1}{K^2}\left(1 - x^K + x^K \log(x)\right).$$

The term $(1 - x^K + x^K \log(x))$ is nonnegative, because $1 - x^K + x^K \log(x) \to 0$ as $x \to 1$ and is decreasing as a function of x: the partial derivative with respect to x

$$\begin{aligned}\frac{\partial}{\partial x}\left(1 - x^K + x^K \log(x)\right) &= Kx^{K-1}\left(-1 + K\log(x) + 1\right)\\ &= K^2 x^{K-1}\log(x) < 0\end{aligned}$$

for $0 < x < 1$. □

4.2 Monotonicity

We now investigate the monotonicity properties of the quality and quantity of contributions in equilibrium as a function of A, the number of tests a contribution must pass in order to qualify for display. We have the following theorems.

THEOREM 4.2. *Fix M. Then, the equilibrium quality $q^*(A)$ is an increasing function of A.*

PROOF. Consider A_1, A_2, with $A_1 < A_2$, and let $q_i = q^*(A_i)$ be the equilibrium quality at A_i. Since q_i minimizes $q^{-A_i}c(q)$, we have

$$q_1^{-A_1}c(q_1) \le q_2^{-A_1}c(q_2). \quad (5)$$

$$q_2^{-A_2}c(q_2) \le q_1^{-A_2}c(q_1). \quad (6)$$

Rearranging and combining these inequalities gives us

$$\left(\frac{q_1}{q_2}\right)^{-A_1} \le \frac{c(q_2)}{c(q_1)} \le \left(\frac{q_1}{q_2}\right)^{-A_2}.$$

That is,

$$\left(\frac{q_1}{q_2}\right)^{A_2 - A_1} \le 1.$$

Since $A_1 < A_2$, $A_2 - A_1 > 0$, which means we must have $\frac{q_1}{q_2} \le 1$ for the above inequality to hold. That is, $q_1 \le q_2$, proving that $q^*(A)$ increases monotonically with A. □

Next we prove that the expected amount of content a viewer sees in equilibrium decreases with A in both the simultaneous and sequential models.

THEOREM 4.3. *Fix M. The expected number of contributions displayed in equilibrium, conditional on this number being nonzero, is a decreasing function of A.*

PROOF. *Sequential Model.* The equilibrium number of qualifying contributions in the sequential model is exactly K (*i.e.*, it is not a random variable). From (3), the equilibrium participation $K = Mq^{*A}/c(q^*)$, where q^* is the minimizer of $q^{-A}c(q)$. So to prove monotonicity of K, it is enough to prove the monotonicity of $f(A) = \min_q q^{-A}c(q)$ with A.

First, observe that for a fixed q, $q^{-A}c(q)$ increases with A. Again, let $A_2 > A_1$. Now, with $q_i = q^*(A_i)$,

$$q_1^{-A_1}c(q_1) \le q_2^{-A_1}c(q_2) \le q_2^{-A_2}c(q_2),$$

where the first inequality follows from optimality of q_1, and the second inequality uses that $q^{-A}c(q)$ increases with A. Therefore, $f(A) = \min_q q^{-A}c(q)$ is an increasing function of A, which implies that $K = M/f(A)$ is decreasing in A.

Simultaneous Model. The number of participants in equilibrium is K, but the number that qualify for display is a random number between 0 and K. The expected number of qualifying contributions, K_q, conditional on seeing at least one contribution, is

$$E[K_q \mid K_q > 0] = \frac{Kq^{*A}}{1 - (1 - q^{*A})^K},$$

since the probability of 0 qualifying contributions is $(1 - q^{*A})^K$. We can write this as

$$\frac{Kq^{*A}}{1 - (1 - q^{*A})^K} = \frac{M}{q^{*-A}c(q^*)}$$

from the zero profit condition. But we just proved that $q^{*-A}c(q^*)$ is an increasing function of A, which means the left-hand side is a decreasing function of A as required. □

4.3 Asymptotics

We now investigate asymptotic equilibrium behavior– how do the equilibrium quality and participation behave in the limit as $M \to \infty$, *i.e.*, what is the best we can hope to achieve with this mechanism? Since q^* is a function of A alone (*i.e.*, the increased benefit from having a larger viewer population will not influence the quality produced in equilibrium), the strictness of the test A will need to be increased to increase the equilibrium quality. We first show that letting $A \to \infty$ indeed causes $q^* \to 1$, that is, the elimination mechanism $\mathcal{M}_e$ indeed achieves what the proportional mechanism could not, namely eliciting optimal qualities in the limit. Next, we show that this can be done while eliciting a high level of participation as $M \to \infty$: in the sequential model we can achieve $K \to \infty$ with $q^* \to 1$, while in the simultaneous model, we show that the relevant quantity, which is the expected quality, can also be made to approach 1 as $M \to \infty$.

THEOREM 4.4. *The equilibrium quality q^* approaches one as A approaches infinity, i.e., $\lim_{A\to\infty} q^*(A) = 1$.*

PROOF. First, we note that $q^*(A) = \arg\min q^{-A}c(q)$ must be an interior minimum, *i.e.*, $0 < q^*(A) < 1$, since $q^{-A}c(q)$ is infinity at both $q = 0$ and $q = 1$ (recall that $c(0) > 1$ and $c(1) = \infty$). Since $q^*(A)$ is an interior minimum of a smooth function, we have the first-order condition

$$-Aq^{*-A-1}c(q^*) + q^{*-A}c'(q^*) = 0.$$

Rearranging, the following relation between A and q^{*A} must always hold, in particular as $A \to \infty$:

$$A = \frac{q^*c'(q^*)}{c(q^*)}.$$

Now, we have $\lim_{q\to 1} q^* \frac{c'(q^*)}{c(q^*)} = \infty$: since $\lim_{q\to 1} c(q) = \infty$, $\lim_{q\to 1} \log c(q) = \infty$ as well. The derivative of a function cannot be finite where the function tends to infinity, so $\lim_{q\to 1}(\log c)' = \frac{c'(q^*)}{c(q^*)} \to \infty$ as well.

Also, this relation cannot be satisfied by any limit for q^* other than 1: by assumption, $c'(q)$ is finite for every $q < 1$, and the denominator $c(q) > 1$ everywhere, since $c(0) > 1$ and c is increasing. Therefore $\lim_{q\to q_0} q^* \frac{c'(q^*)}{c(q^*)}$ is finite for every $q_0 < 1$. □

Next we investigate the asymptotic *quantity* of display content in equilibrium. The trade-off between the number of qualifying contributions and q^* for a fixed M means that we must check whether this number is indeed large enough when we send $q^* \to 1$.

In the sequential model, there is no randomness associated with the number of displayed contributions– so if there is an equilibrium with $K \geq 1$, then there will be exactly K displayed contributions. While we do not prove the result here, an argument very similar to the one given below for expected quality shows that it is possible to choose A so that the equilibrium quality approaches 1 alongwith the number of qualifying contributions K approaching infinity.

In the simultaneous model, however, $K > 1$ is not enough to ensure that a nonzero number of contributions will eventually be displayed. Here, we would like to ensure not only that equilibrium quality approaches 1, but that the *expected* quality, which we define as the equilibrium quality times the probability of seeing at least one qualifying contribution, can be made to approach 1 as well. The next theorem answers the question of whether there exists a way to increase A with M such that we can achieve both objectives simultaneously.

THEOREM 4.5. *There exists a choice of $A(M)$ such that the expected quality in the simultaneous model, $q^*(1-(1-{q^*}^A)^K) \to 1$ as $M \to \infty$.*

PROOF. For the expected quality to tend to 1 as $M \to \infty$, we need the term $(1-{q^*}^A)^K$ to converge to 0 with M. For this, it is enough to require that for some $0 < \alpha < 1$, we have $(1-{q^*}^i)^K \leq \alpha^i$. Taking logarithms on both sides, it is enough that $K = K(i, M)$ satisfies

$$K \geq \frac{\log \alpha^i}{\log(1-{q^*}^i)}.$$

Suppose $A(M_0) = i$ for some M_0 is such that this inequality holds for the corresponding equilibrium values of K and q.

Recall that for a fixed A, $K \to \infty$ as $M \to \infty$. We let M increase while holding A fixed at i, until the M_1 where K is large enough so that

$$K(M_1, i+1, q(i+1)) \geq \frac{\log \alpha^{i+1}}{\log(1-{q^*}^{(i+1)})}.$$

We then set $A(M_1) = i+1$, and $A(M) = i$ for all $M_0 \leq M < M_1$.

By construction, $\lim_{M\to\infty}(1-{q^*}^{A(M)})^K \to 0$, and the expected quality converges to one as required. □

How much contribution effort is wasted? Requiring each contribution to pass A tests, which happens with probability q^A, means that there will be a number of wasted contributions. Specifically, to generate K successful contributions, the expected number of participants required is the expected number of coin tosses before getting K heads, where each coin comes up heads with probability q^A– this expected value is K/q^A. So the fraction of generated content that is actually displayed is q^A. This is the same fraction of generated content that is displayed in an equilibrium in the simultaneous model as well, where the number of participants is exactly K (in contrast with an expected number of participants K/q^A in the sequential model), and the expected number of qualifying contributions is Kq^A (in contrast with exactly K in the sequential model). How does this ratio q^A behave in the limit as $A \to \infty$? Naturally, the answer must depend on the properties of the cost function c. The following result characterizes q^A in terms of the cost function c.

THEOREM 4.6. *Suppose we can find b such that $\lim_{q\to 1} \frac{qc'(q)}{c(q)} - z(1-q)^{-b} \to 0$ for some $0 < z < \infty$. Then*

- $\lim_{A\to\infty} q^A \to 1$ *if* $b < 1$;
- $\lim_{A\to\infty} q^A \to e^{-z}$ *if* $b = 1$; *and*
- $\lim_{A\to\infty} q^A \to 0$ *if* $b > 1$.

PROOF. Recall that A and q^* are related by $A = \frac{q^* c'(q^*)}{c(q^*)}$. Therefore,

$$\log(q^A) = A \log q = \frac{q^* c'(q^*)}{c(q^*)} \log q.$$

Since $\lim_{A\to\infty} q = 1$, we have

$$\begin{aligned}
\lim_{A\to\infty} \log(q^A) &= \lim_{q\to 1} \frac{q^* c'(q^*)}{c(q^*)} \log q \\
&= \lim_{q\to 1} z(1-q)^{-b} \log(q) \\
&= \lim_{q\to 1} \frac{z \log q}{(1-q)^b} \\
&= \lim_{q\to 1} \frac{z}{-bq(1-q)^{b-1}},
\end{aligned}$$

where the last step uses L'Hospital's rule since both the numerator and denominator tend to zero as $q \to 1$.

Now, if $b > 1$, $\lim_{q\to 1} q(1-q)^{b-1} = 0$, so that $\lim_{A\to\infty} \log q^A = -\infty$, or $q^A \to 0$. If $b = 1$, $\lim_{q\to 1} q(1-q)^{b-1} = 1$, and $\log q^A \to -(z/b) = -z$, so $q^A \to e^{-z}$, which is some constant between zero and one. Finally, if $b < 1$, $\lim_{q\to 1} \frac{z}{-bq(1-q)^{b-1}} = 0$, so that $q^A \to 1$. □

Consider again the function $c(q) = c(1-q)^{-r}$. For this function, $c'(q) = rc(1-q)^{-r-1}$, so that $qc'(q)/c(q) = rq(1-q)^{-1}$. Therefore $z = r$, $b = 1$, and $q^A \to e^{-k}$. So a constant fraction of the contributions are displayed, but also a constant fraction of the contributions are wasted. A natural question is whether it is possible to achieve optimal quality, along with the limiting fraction of waste converging to zero. We next answer this question in the affirmative.

4.4 Elimination with Restricted Entry

In many online settings, such as comments on blogs or news articles, or product or restaurant reviews, limiting entry by closing or disallowing further contributions may not be, practically speaking, desirable. But there are online settings where restricting entry may indeed be feasible, for instance, question-answer forums sometimes close questions to

new answers beyond a certain date. Below we show that restricting contributions when a certain number of *qualifying* contributions has been reached (where qualification is determined by the elimination mechanism as before) can lead to very desirable properties: it can allow us to send $q^* \to 1$ while sending the fraction of wasted contributions to zero.

Mechanism $\mathcal{M}_r$. We now consider a modified version of the elimination mechanism, M_r, which *restricts* entry– we continue to subject each unit of content to A tests, but limit K to a restricted number of contributors, that is, prevent free entry. (Note that simply restricting entry without elimination does not achieve anything since share of attention remains independent of quality, which means equilibrium qualities remain 0 exactly as in the baseline mechanism.) We analyze this under the sequential entry model. With $A = 1$, each contributor gets a payoff

$$\pi(q) = \frac{M}{K}q - c(q).$$

The equilibrium quality chooses q^* to maximize $\pi(q)$, which is independent of other contributors' choices and depends only on M, K. So

$$\frac{M}{K} = c'(q^*),$$

so that $q^* \to 1$ as M diverges provided M/K diverges as well, since $\lim_{q\to 1} c'(q) \to \infty$, and $\lim_{q\to q_0} c'(q)$ is finite for every $q_0 < 1$. Therefore, even with just one test, the equilibrium quality under restricted entry converges to 1 as $M \to \infty$, for any constant number of allowed contributions K. (Of course, with $A = 0$ the quality $q^* = 0$ as well.) The expected number of wasted contributions is $K(1-q^*)$, which converges to 0 for any constant K (and can also be made to go to 0 for increasing K provided the target K grows slowly enough).

5. A MORE GENERAL ATTENTION MODEL

Our analysis so far assumes that the attention per contribution when there are M viewers and K displayed contributions is M/K. This assumption is accurate when each viewer views exactly one unit of content, or views Y units of content where $Y < K$ (*i.e.*, no viewer has more attention available than the amount of content displayed), or alternately, spends some amount of time t on the page that is independent of the number of displayed contributions.

A more general model of viewing behavior is the following. There is a distribution of types in the viewer population, where a fraction p_i of viewers views $\min(i, K)$ units of content. That is, a viewer in the ith segment has enough attention to consume i contributions (such as comments, reviews), and will do so, unless fewer are displayed in which case she consumes all of it[9]. By definition, the p_i are a distribution, *i.e.*, $\sum_{i=1}^{\infty} p_i = 1$. When there are M viewers, the total amount of available attention in this model of viewing behavior, when K units are displayed is

$$f(M, K) = M \sum_{i=1}^{\infty} \min(i, K)p_i.$$

[9]In general, of course, this would also depend on the contribution quality, but our argument will work provided we use the distribution for the lowest quality.

We note that another natural model of viewing behavior, where a viewer continues to the next contribution with probability δ after reading a contribution, generates a special case of this model with p_i proportional to δ^{i-1}.

The total amount of available attention from M viewers is now a *function* of the number of displayed contributions K (as before, using random permutations, we ensure that this attention is distributed equally across all displayed content). We will show that our equilibrium analysis continues to hold for this more general model of viewing behavior, and in fact, also holds for any model of viewer attention as long as the total attention as a function of the number of viewers and displayed contributions, $f(M, K)$, satisfies $f(M, K)/K \geq M/K$ for each K.

THEOREM 5.1. *Suppose $f(M, K)/K > M/K$. Then the equilibrium quality q^* remains unchanged, and the equilibrium participation K is at least as large as that when each contribution's share of attention is M/K.*

PROOF. For simplicity, consider the sequential model. The probability of being selected for display, q^A, is unchanged, as is the cost function $c(q)$. The payoff is modified to $\pi(q) = \frac{f(M,K)}{K}q^A - c(q)$. Since $f(M, K)$ is independent of q, the same equilibrium construction as in Theorem 4.1 can be used to show that equilibrium quality $q^* = \arg\min q^{-A}c(q)$. Therefore, the equilibrium quality is unchanged. The equilibrium K is determined by the zero profit condition $\frac{f(M,K)}{K} = {q^*}^{-A}c(q^*)$. Since $\frac{f(M,K)}{K} \geq \frac{M}{K}$ for each K, the value of K satisfying the zero profit condition above is larger than $\frac{M{q^*}^A}{c(q^*)}$, which is the equilibrium participation when the total available attention is exactly M. □

In the distribution model of viewing behavior, it is evident that $f(M, K) > M$ since $\sum_{i=1}^{\infty} ip_i > 1$. Therefore, $f(M, K)/K > M/K$. This immediately implies that as in our analysis so far, $\lim_{A\to\infty} q^* \to 1$, and this can be achieved while simultaneously sending $K \to \infty$. Note that the monotonicity results hold as well, by the same arguments as before.

6. OPTIMIZING FOR VIEWER WELFARE

We have so far studied the problem of incentivizing high-quality content from the perspective of strategic contributors, which allows addressing questions concerning high quality and participation. However, for a *fixed* number of viewers M, recall that there is a trade-off between quality and quantity that can be achieved by changing the number of tests A. How should we choose A, or equivalently, a point on this trade-off curve?

The answer to this question must come from a model of *viewer utility*, which specifies how viewers derive utility from a *set* of contributions. In general, the utility from a set of contributions can be an arbitrary, complicated function of a number of factors in addition to number and quality, such as the overlap amongst the contributions. Here we will illustrate this idea of choosing A based on the viewer's utility for an example viewer utility function.

Viewer Model. As an illustration, suppose the utility of consuming reviews $q_1, q_2, \ldots, q_K$ is $MK^{-\alpha}\sum_{i=1}^{K} q_i$, for some $0 < \alpha < 1$ and M users. Here, the viewer does derive a positive value from additional reviews, but with decreasing

returns to scale (which is necessary to prevent needing an infinite number of reviews). When α is close to 1, the user only cares about the average content quality and does not derive value from having a number of different contributions, so producing a single excellent contribution is more desirable than several average ones. This could be the case, for example, when a user asks for the proof of a theorem on the popular mathematics question-and-answer site MathOverflow– a few almost correct proofs are useless compared to a single actual proof. In contrast, using α close to 0 models situations where *diversity* is very valuable where the user would rather have a large number of reasonable contributions than a few near-perfect ones, such as for instance when asking for suggestions about things to do in New York.

The viewer-optimal outcome is the best that could be done for the viewers given their utility model, while respecting the individual rationality of contributors. Suppose a dictator can see the quality of content. To maximize viewer welfare, he would set a threshold quality q_1 and a number of reviewers K to maximize $MK^{1-\alpha}q_1$ subject to $M/K - c(q_1) \geq 0$. Clearly this solves with $M/K = c(q_1)$ so that $K = (M/c(q_1))$, and hence the viewer-optimal solution maximizes $Mq_1(M/c(q_1))^{1-\alpha}$. This solution chooses a threshold quality q_1 independent of M that minimizes $f(q_1) = q_1^{\frac{1}{1-\alpha}}c(q_1)$. The function $f(q_1)$ takes a minimum in $(0, 1)$ and has derivative

$$q_1^{\frac{1}{1-\alpha}-1}c(q_1)\left(\frac{q_1c'(q_1)}{c(q_1)} - \frac{1}{1-\alpha}\right).$$

In the free entry solution in the elimination mechanism with A tests, the equilibrium quality satisfies $\frac{q^*c'(q^*)}{c(q^*)} = A$. Setting the number of tests to be the closest integer A to $\frac{1}{1-\alpha}$ produces an equilibrium quality close to the quality that maximizes viewer welfare (if there is a unique solution to the first-order condition)[10]. That is, the model of viewer welfare now allows us to answer the question of what is the "optimal" number of tests A: it is the one that induces the optimal utility for users. We see that as α decreases down to zero (marginal returns become less and less diminishing, *i.e.*, additional contributions become more valuable) one needs less and less stringent tests; for $\alpha = 0$ (corresponding to the extreme of additive utility), a single test is optimal.

We point out that exactly the same reasoning applies for the slightly more general contributor utility function is $MK^{-\beta} - c(q)$, with $1/(1-\alpha)$ being replaced by $\beta/(1-\alpha)$ as the choice for A.

7. DISCUSSION

In this paper, we introduced a game-theoretic framework to analyze and compare mechanisms for user generated content with a very simple model of utility for contributors, assuming nonstrategic viewers. There are a number of interesting directions for further work within this framework by relaxing various assumptions we make.

A natural direction to explore is different utility functions for contributors; we outline two specific ideas. First, some contributors might have negative cost of contribution for q less equal some threshold q_0, *i.e.*, they derive positive value from simply generating the content irrespective of exposure; this is relevant, for instance, in the context of reviews for very popular books and movies. How does the presence of such agents, who will always participate, affect which mechanisms have good equilibria?

A second interesting direction is asymmetric costs, which is relevant in the context of Q&A sites– an expert may have a much lower cost for contributing a high-quality answer than a nonexpert. The theory developed in this paper can be extended to asymmetric costs with some increase in notational complexity. Asymmetric costs generally open the door to inefficient equilibria, where a low cost contributor is deterred from contributing because of the contribution of a higher cost contributor. Such inefficiencies are especially likely in the sequential world, because lower cost contributors need not be the first to appear, but are also possible in the simultaneous environment. While the elimination mechanism can be used to induce high quality contributions, it treats every contribution that passes the test equally; it would be desirable to incorporate into the mechanism a provision that contributions which do better are shown more prominently. This relation of quality to quantity induces something akin to the proportional mechanism, although with a cutoff to insure a minimum quality. As we saw, though, the proportional mechanism is not very good at inducing quality; a 'top-P' type mechanism will in fact induce higher quality contributions by lower marginal cost suppliers, but then fails to take advantage of the higher quality by displaying it more prominently. Thus, our findings suggest a tradeoff between inducing higher quality, and taking advantage of the higher quality. It is an open problem to ascertain what mechanisms perform best in the asymmetric cost environment.

Finally, we note briefly that allowing malicious or strategic voters, who do exist in practice, introduces an additional dimension of complexity by creating a two-sided setting with strategic agents on both sides, since the quality of both contributions and voters is unknown. This problem might be best modeled and studied in a repeated setting which utilizes user and voter reputation.

Acknowledgments. We are grateful to Judd Antin, Joshua Gans and David Pennock for helpful comments and pointers to the literature.

8. REFERENCES

[1] E. Friedman, P. Resnick, R. Sami, *Manipulation-Resistant Reputation Systems*, Algorithmic Game Theory, Cambridge University Press, 2007.

[2] Y. Chen, S. Jain, D. Parkes, *Designing Incentives for Online Question and Answer Forums*, ACM EC, 2009.

[3] S. Jain, D. Parkes, *The Role of Game Theory in Human Computation Systems* (Position Paper), HCOMP 2009.

[4] S. Jain, D. Parkes, *A Game-Theoretic Analysis of the ESP Game*, Working Paper.

[5] A. Jøsang, R. Ismail, C. Boyd, *A Survey of Trust and Reputation Systems for Online Service Provision*, Decision Support Systems, 43(2), 618-644, March 2007.

[6] K. Ling, G. Beenen, X. Wang, K. Chang, D. Frankowski, P. Resnick, R. Kraut, *Using social psychology to motivate contributions to online communities*, Journal of Computer Mediated Communication, 10(4), 2005.

[7] S. Rafaeli, Y. Ariel, *Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia*, Psychological Aspects of Cyberspace, Cambridge University Press, 2008.

[10] This is the case if, for example, $f(q) = \frac{qc'(q)}{c(q)}$ is an increasing function of q (this is satisfied by $c(q) = c/(1-q)^b$).