Designing Incentives for Online Question and Answer Forums

Shaili Jain
School of Engineering and
Applied Sciences
Harvard University
Cambridge, MA 02138 USA
shailij@eecs.harvard.edu

Yiling Chen
School of Engineering and
Applied Sciences
Harvard University
Cambridge, MA 02138 USA
yiling@eecs.harvard.edu

David C. Parkes
School of Engineering and
Applied Sciences
Harvard University
Cambridge, MA 02138 USA
parkes@eecs.harvard.edu

ABSTRACT

In this paper, we provide a simple game-theoretic model of an online question and answer forum. We focus on factual questions in which user responses aggregate while a question remains open. Each user has a unique piece of information and can decide when to report this information. The asker prefers to receive information sooner rather than later, and will stop the process when satisfied with the cumulative value of the posted information. We consider two distinct cases: a *complements* case, in which each successive piece of information is worth more to the asker than the previous one; and a *substitutes* case, in which each successive piece of information is worth less than the previous one. A best-answer scoring rule is adopted to model Yahoo! Answers, and is effective for substitutes information, where it isolates an equilibrium in which all users respond in the first round. But we find that this rule is ineffective for complements information, isolating instead an equilibrium in which all users respond in the final round. In addressing this, we demonstrate that an approval-voting scoring rule and a proportional-share scoring rule can enable the most efficient equilibrium with complements information, under certain conditions, by providing incentives for early responders as well as the user who submits the final answer.

Categories and Subject Descriptors

H.5.3 [**Information Interfaces and Presentation (e.g. HCI)**]: Group and Organizational Interfaces; J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Design, Economics, Theory

1. INTRODUCTION

Yahoo! Answers is a question and answer forum where users can post questions or answer questions on wide va-

riety of topics. Yahoo! Answers has 25 categories ranging from 'Computers & Internet' to 'Travel' to 'Family & Relationships' to 'Health.' Users may post discussion questions, factual questions or polls. In Yahoo! Answers, people do not exchange money for answers to questions. Participation in Yahoo! Answers is encouraged through an elaborate point system, and with leaderboards and top-contributor designations to encourage participants to accumulate more points.

In this paper, we provide a game-theoretic model of behavior for online question and answer forums such as Yahoo! Answers. We focus on modeling the one-shot question and answer game. Additionally, we focus on modeling factual questions, such as "What are the main causes of the current housing crisis?", rather than discussion questions (e.g. "What is your favorite movie of all time?"). Factual questions have been demonstrated to have a higher archival value than discussion questions [5].

Our interest is in understanding the structure of the equilibria in a model that captures some qualitative features of these environments, and especially in considering the effect of alternate scoring rules on the quality of these equilibria for the asker. In the model that we propose, each user has a unique piece of information that is relevant to a question and can decide when to report this information. As information is reported it is aggregated into the responses, so that the value to the asker monotonically improves while the question remains open. In the case that multiple pieces of information are simultaneously revealed in the final round, we assume that the asker is able to combine the information.

In considering the interactions between the information, we consider two distinct cases: a *complements* case in which each successive piece of information is worth more to the asker than the previous one; and a *substitutes* case in which each successive piece of information is worth less. The asker prefers to receive information sooner rather than later, and will stop the process when satisfied with the cumulative value of the posted information.

We first analyze the equilibrium for a *best-answer scoring rule*, that is designed to model the current Yahoo! Answers environment. Upon stopping the process, the asker assigns one point to the best answer among all responses. As we assume that every user has the ability to combine her own information with information that has already been revealed, under the best-answer scoring rule the asker assigns the point to the most recent answer (breaking ties at random in the case that multiple answers are received in the most recent round). We find that this scoring rule is ef-

fective in isolating the efficient equilibrium in the case of substitutes information, in which all information is posted in the very first round. On the other hand, the best-answer rule is ineffective for complements information, where it instead isolates the least efficient equilibrium in which every user posts information in the very last round.

In addressing this problem, we consider two alternative scoring rules. The first is an *approval-voting scoring rule*, in which the asker assigns one point to each of the best $k > 1$ answers. In our setting, this means that the asker assigns one point to each of the most recent set of k answers (with ties broken at random if more than k answers were received in the most recent round, or more than $k - k_1$ users in the penultimate round with k_1 in the most recent round, and so on). With this scoring rule, we find that it is now possible to have the most efficient equilibrium outcome for complements information, with certain restrictions on the asker's valuation function. This scoring rule also retains the efficient equilibrium in the case of substitutes information. However, the downside of this rule is that it also retains the least efficient equilibrium for complements information and introduces the least efficient equilibrium for substitutes information, again under certain restrictions on the asker's valuation function. An interesting feature of this scoring rule is the tunable parameter k, which represents the trade-off between the benefit of this scoring rule for the case of complements information and the disadvantage for the case of substitutes information.

The second scoring rule we propose is the *proportional-share scoring rule*, in which the asker assigns some share of the available points in proportion to the marginal value contributed by a user in the round in which the user submits information. With this scoring rule, we find that it is now possible to support the efficient equilibrium outcome (i.e., the equilibrium in which all answers are received in the first round) for complements information, again with certain restrictions on the asker's valuation function. This scoring rule also retains the efficient equilibrium as the unique outcome in the case of substitutes information. On the other hand, while the efficient equilibrium is unique across all 'pooling' equilibrium in which users respond in the same round (for certain restrictions on the valuation function), we are unable to rule out 'separating' equilibrium in which users respond in different rounds for complements valuations under the proportional-share rule.

The approval-voting and proportional-share rules differ in the informational requirements that they place on the asker. Approval voting does not require any additional information beyond that required in the best-answer rule, i.e. just a signal as to when the asker is happy with the most recent answer(s). On the other hand, the proportional-share rule requires the asker, upon stopping the process, to associate a numerical score with the total information value as the answers aggregate across rounds. It is an interesting open question to understand whether scoring rules can be designed that preclude inefficient equilibrium for complements valuations without requiring this additional information from the asker.

2. YAHOO! ANSWERS AND Q&A FORUMS

In Yahoo! Answers, users maintain a tally of points and are penalized for asking questions, rewarded for logging onto the system each day, rewarded for answering a question and

Activity	Points
first time user	+100 points
log into the system once per day	+1 point
ask a question	-5 points
answer a question	+2 points
have your answer chosen as best answer	+10 points
pick a best answer for your question	+3 points

Table 1: The Points Scheme in Yahoo! Answers

rewarded heavily for having answers selected as a best answer. The complete point system is given in Table 1.

Based on the number of points a user accumulates, the user receives a "level" designation, with there being seven levels in total. The higher the level, the greater privileges a user will get in terms of the number of questions she can ask per day, the number of questions she can answer per day, etc. All users have a profile where the number of points the user has, the level, and the percentage of best answers is clearly displayed. Perhaps this is the most visible information displayed about a user.

In addition to the point system, leaderboards and top contributor designations encourage users to accumulate more points. For each category and sub-category, the "top contributor" is displayed at the top of the page. Likewise, for each category and sub-category, there is a leaderboard of the top ten users. It has been shown that points are a factor in motivating users to participate in points-based question and answer forums [14]. Note, though, that points are not needed to ask questions because users can always create new identities and obtain 100 new points!

While we believe that this is the first game-theoretic investigation of question and answer forums, there do exist a number of existing empirical studies, there have been a number of empirical studies devoted to understanding how users participate in such forums [14, 19, 1]. Nam et al. [14] aim to understand the underlying motivations for why users participate in question and answer forums. They study the Naver Knowledge-iN (KiN) system, the largest question and answer community in South Korea. They give a survey to 26 users of KiN and find that altruism along with selfish reasons (e.g. points, promoting personal businesses, maintaining personal knowledge, etc.) are top motivations for participating in KiN. These authors also show that the average number of answers to a question increases as the expected reward increases.[1] Indeed, others observe that higher rewards lead to increased participation. Yang et al. [19] find that higher monetary rewards for a solution to a task on *Taskcn*, a popular web-based knowledge-sharing market in China, attract more views for the task and increased task participation. Harper et al. [6] compare different question and answer forums and find that monetary rewards and increased rewards lead to higher quality answers.

Perhaps in support of the fact that users can build upon previous answers and receive credit for the aggregated information, Nam et al. [14] present an interesting result: In the "C++" forum, the next-to-last question is chosen as a best answer 51% of the time and the last question 69% of the time [14].[2] These authors get virtually identical results for

---

[1] KiN allows an asker to increase the reward for a question by up to 100 additional points.

[2] KiN allows for selecting more than one best answer.

the "Java" forum and similar results for the "Singer" forum.

Making the case that users are behaving strategically, Yang et al. [19] also point out that users learn to select tasks where they are competing against few opponents, to increase their chances of winning. Users also, over time, select tasks with higher expected rewards.

Yang et al. [19] also study user behavior over time and find that most users become inactive after only a few submissions. They also find that there is a very small core of successful users who manage to increase their win percentage over time. This core group accounts for 20% of the winning solutions on Taskcn. Nam et al. [14] also observe a heavy tailed distribution in terms of user contribution and that many users drop out after a few contributions. This is a general theme across many websites with user generated content [18].

In an attempt to improve the quality of information on Yahoo! Answers, Harper et al. [5] focus on distinguishing *factual questions* in Yahoo! Answers from *discussion questions*. Harper et al. point out that factual questions have a higher archival value than discussion questions, which indeed is our motivation for focusing on factual questions. Harper et al. then use a number of classifiers to group questions as either factual questions or discussion questions. Adamic et al. [1] also observe the difference between factual forums and discussion forums and find that factual forums tend to have a smaller number of responses, while each response is relatively long, whereas discussion forums tends to have more responses that are shorter in length. In addition, these authors find that discussion forums tend to have a greater amount of overlap between users who ask questions and users who ask questions than factual forums.

Other research in question and answer forums has focused on determining experts in such communities [20, 11, 12], while others have observed the redundancy in such systems and attempt to retrieve semantically similar questions [10]. Many studies notice the varying quality of user generated content, and aim to retrieve high quality content [2, 4] and improve content quality [8]. Yet other studies try to predict certain properties of Q&A forums (e.g., the likelihood an answer will be chosen as a best answer [1] or asker satisfaction [13]).

Online question and answer forums fit into the larger realm of *peer production* systems [3], which is a term used to refer to decentralized system of users that contribute to a system to achieve a global goal, without receiving monetary compensation for their work. Examples of peer production systems include Wikipedia, YouTube, and human computation systems [16, 17]. Prior work has presented a game-theoretic analysis of human computation systems, specifically the PhotoSlap game [7] and the ESP game [9]. There have also been a number of empirical studies analyzing user contribution to various peer production systems. It has been shown that there exists strong regularities among a wide range of peer production systems [18], namely power laws in terms of contribution to such systems. Pouwelse et al. [15] also provide a nice empirical survey of the growth of many peer production systems over the past decade.

3. OUR MODEL

We focus on modeling how users participate in answering a single question posted by asker. We assume that the question is on a particular topic that has n pieces of dis-joint information. Denote $\mathcal{I} = \{I_1, I_2, \ldots, I_n\}$ as the information space of the question. There are n users in the system that can potentially answer the question. Each user $i \in \{1, \ldots, n\}$ possesses a unique piece of information I_i. Even though information is private, the fact that everyone possesses a piece of information out of n total pieces is common knowledge for all users and the asker.

The question-answering process has T rounds, unless the asker closes the question earlier. The users each make a decision about which round to participate in, and are able to observe the responses by other users before responding. When participating in round $t > 1$, every user has the ability to combine her own piece of information with all other pieces of information that have been revealed in previous answers, and submit an integrated answer. Each user, however, can only answer the question once. This restriction is also present in the Yahoo! Answers forum. At the end of each round, the asker decides whether or not to close the question.

We assume that users seek to maximize their expected score in answering the question and are thus selfish and motivated solely by points, not altruism. Because there is no cost in our model to answering a question, and because users seek to maximize their score, we can restrict attention to strategies in which a user will always submit an answer to the question. Moreover, a user will always choose to submit an answer that aggregates her information with previous information because this can only increase her score under all scoring rules that we consider. When multiple new answers are received in the same round, we assume that the asker is herself able to combine the information in these responses.

The asker is modeled with a valuation function, that maps sequences of information from information space $\mathcal{I}$ into a real number representing the asker's value for the cumulative information associated with the sequence of information. Let σ be a permutation of $\mathcal{I}$ and $\sigma_i(\mathcal{I})$ be the i-th element of σ. The asker's valuation function v satisfies the following properties:

P1: $v(I_0) = 0$, when $I_0 = \emptyset$;

P2: $v(I_1, I_2, \ldots, I_j) = v(\sigma_1(\mathcal{I}), \sigma_2(\mathcal{I}), \ldots, \sigma_j(\mathcal{I}))$ for all $1 \leq j \leq n$ and all permutations σ of $\mathcal{I}$;

P3: $v(\sigma_1(\mathcal{I}), \sigma_2(\mathcal{I}), \ldots, \sigma_j(\mathcal{I})) < v(\sigma_1(\mathcal{I}), \sigma_2(\mathcal{I}), \ldots, \sigma_{j+1}(\mathcal{I}))$ for all $1 \leq j < n$ and all permutations σ of $\mathcal{I}$.

From now on, we use $v(i) = v(I_1, I_2, \ldots, I_i)$ to denote the asker's valuation for *any* i pieces of information, not just those from agents $\{1, \ldots, i\}$. This is possible due to property P2 which requires that all information is equivalent in the sense that any i pieces of information generate the same value.

To define the stopping rule, we model the asker as drawing a *threshold value*, $\theta \sim U[0, v(n)]$, uniformly at random between 0 and the value of receiving all information in the information space $\mathcal{I}$ of the question. Once the asker's valuation for items received exceeds θ, the asker is satisfied and closes the question, and awards points to one or more users according to some specific scoring rule. The distribution on this threshold value is common knowledge to all users but only the askers knows the actual threshold value. That the question is closed once the threshold is exceeded models the intuition that the asker prefers to receive an answer sooner rather than later.

Given that an earlier answer is preferred to a later answer and that more information is better than less information, we can identify the *efficient outcome* as that in which every user responds in the very first round. This can be thought of as a *pooling equilibrium*, with every user coordinating on a particular round in which to submit information. Apart from a pooling equilibrium, we can also consider a *separating equilibrium*, in which pieces of information are received in distinct rounds. The *least* efficient outcome is that in which each user waits until the very last round to respond. Note that in all equilibria, the asker will certainly receive all information by the last round because it is costless for users to submit an answer. Therefore, a pooling equilibrium in which all users coordinate on the last round is the least efficient of all outcomes.

Depending on the nature of the question, the pieces of information related to the question may be complements or substitutes. For example, suppose the asker posts the question: "What should I do for a one-day visit to Boston?" The two pieces of information, "walk along the Freedom Trial" and "have lunch at Quincy Market (which is on the Freedom Trial)" are complements, because the value of knowing both pieces of information for the asker is higher than the sum of the values of only knowing a single piece of information. However, if the asker posts the question: "Where should I have lunch in Times Square?", the answers "Becco" and "Kodama" are substitutes for the asker, since the asker must choose between the two.

To model the different nature of these two questions and the associated information, we consider the case where the value of each successive piece of information received by the asker is of greater value than the previous one (the *complements* case) and the case where the value of each successive pierce of information received by the asker is less than the value of the previous one (the *substitutes* case). In the following definitions, let $\delta_j = v(j) - v(j-1)$.

DEFINITION 3.1. *In the complements case, the valuation function must satisfy $\delta_j < \delta_{j+1}$ for all $1 \leq j < n$.*

The substitutes case is defined analogously.

DEFINITION 3.2. *In the substitutes case, the valuation function must satisfy $\delta_j > \delta_{j+1}$ for all $1 \leq j < n$.*

Under these configurations, we provide the probability for the asker to close a question at time t. Let $b(t)$ denote the amount of information that the asker has at the end of each time round, where $1 \leq t \leq T$.

REMARK 3.1. *The probability of stopping after each round t, conditional on that the question has not been closed at round $t - 1$, is $\frac{v(b(t))}{v(n)}$ for both the complements and the substitutes cases. Furthermore, in the complements case, if $b(t) = i$, then $\frac{v(b(t))}{v(n)} < \frac{i}{n}$ for all i. Likewise, for the substitutes case, if $b(t) = i$, then $\frac{v(b(t))}{v(n)} > \frac{i}{n}$ for all i.*

PROOF. Conditional on that the question has not been closed, the probability of stopping at step t equals $P[\theta \leq v(b(t))]$. Since θ follows a uniform distribution, we have

$$P[\theta \leq v(b(t))] = \frac{v(b(t))}{v(n) - v(\emptyset)} = \frac{v(b(t))}{v(n)}.$$

We must have $v(b(t)) = \sum_{j=1}^{i} \delta_j$ and $v(n) = \sum_{j=1}^{n} \delta_j$. For the complements case, by Definition 3.1, $\delta_i < \delta_{i+1}$. We

have,

$$\frac{v(b(t))}{v(n)} = \frac{\sum_{j=1}^{i} \delta_j}{\sum_{j=1}^{n} \delta_j} = \frac{1}{1 + \frac{\sum_{j=i+1}^{n} \delta_j}{\sum_{j=1}^{i} \delta_j}} < \frac{1}{1 + \frac{(n-i)\delta_i}{i\delta_i}} = \frac{i}{n}.$$

For the substitutes case, by Definition 3.2, $\delta_i > \delta_{i+1}$. Hence the inequality in the above expression is reversed for the substitutes case. □

REMARK 3.2. *If $T = 2$, the probability of stopping in the first round is p where $p = \frac{v(i)}{v(n)}$ and i is the number of items received in the first round. The probability of stopping in the second round is $1 - p = \frac{v(n) - v(i)}{v(n)}$.*

PROOF. By Remark 3.1, the probability of stopping at the first round is $p = \frac{v(i)}{v(n)}$, and the probability of stopping at the second round, conditional on not stopping at the first round, is $\frac{v(n)}{v(n)} = 1$. Hence, the unconditional probability of stopping at the second round is $(1 - p) \times 1 = 1 - p$. □

REMARK 3.3. *If $T \geq 2$, the probability of stopping in the first round is p where $p = \frac{v(i)}{v(n)}$ and i is the number of items received in the first round. The probability of stopping in the second round is p' where $p' = \frac{v(i+j) - v(i)}{v(n)}$ and j is the number of items received in the second round, and so on. More generally, the probability of stopping in round k is q where $q = \frac{v(b(k)) - v(b(k-1))}{v(n)}$, where $b(k)$ is the set of information available at the end of round k.*

PROOF. The probability that the question was not closed at round $k - 1$ is $P[\theta > v(b(k-1))]$. Then conditional on the fact that the question was not closed at round $k - 1$, the probability of stopping at round k is $P[\theta < v(b(k))|\theta > v(b(k-1))]$. Hence, the unconditional probability of stopping at round k is the product of the two probabilities, which equals

$$P[v(b(k-1)) < \theta < v(b(k))] = \frac{v(b(k)) - v(b(k-1))}{v(n)}.$$

□

4. ANALYSIS OF BEST-ANSWER RULE

The best-answer rule models the scoring method currently used by Yahoo! Answers. In Yahoo! Answers, upon closing the question, the asker can select one answer as the best answer and the associated user is then awarded some fixed number of points. Without loss of generality, we normalize the number of points awarded to 1.[3]

When the asker closes the question because the value has reached the threshold, the asker selects the user that answered in the most recent round as the winner. When there are multiple answers provided in the final round, the asker uniformly picks one of them as the best answer. On one hand we see that users would prefer to wait so that the value of the answer that they submit is maximized since their information will be aggregated with earlier answers. But on the other hand, waiting too long could result in a missed opportunity because the question may be closed in an earlier stage.

---

[3]This is without loss of generality because we model only a single game, and thus the relative weight of points for being selected as the best answer vs. asking a question and so forth is irrelevant in our analysis.

This tradeoff between waiting to form better answers and participating before the question is closed is the key strategic consideration facing users. In the rest of this section, we explore this tradeoff and perform equilibrium analysis for both complements and substitutes cases. We use the notion of an *active round* in our analysis. A round is active if at least one user participates in that round, otherwise it is inactive. We will establish a clean separation for the complements and substitutes cases: the unique Nash equilibrium profile for complements valuations is the most inefficient outcome, while the unique Nash equilibrium profile for substitutes valuations is the most efficient outcome.

4.1 Complements Case

For the complements case, the asker's valuation of having a collection of several pieces of information is higher than the sum of her valuations for individual pieces. The benefit of waiting to form a better answer is therefore relatively high. The following results show that the only Nash equilibrium for the complements case under the best answer rule, is that all users answer the question in the final round, just before the question will definitely close. This is the least efficient equilibrium, because the asker must wait to get an answer until the last possible round.

LEMMA 4.1. *Consider any strategy profile that involves all users playing in the same round. The only one of these strategy profiles that forms a pure-strategy Nash equilibrium is the one in which all users play last, for any valuation function satisfying the complements condition and under the best-answer rule.*

PROOF. Any strategy profile that involves all users playing in the same round, yields an expected payoff of $\frac{1}{n}$ to each user, because with probability $\frac{1}{n}$ their answer is selected as the best answer. Let the active round be t. When $t < T$, a user can deviate by participating in round $t + 1$. The probability that the question is closed at the end of round t is $p < \frac{n-1}{n}$, due to Remark 3.1. The deviating user earns an expected payoff $1 - p > \frac{1}{n}$. Thus, all users playing at round $t < T$ can not be a Nash equilibrium. Finally, consider the strategy profile consisting of all users participating in the Tth round. If user i deviates by going earlier, his expected payoff equals the probability that the question is closed before round T, which in this case is $\frac{v(1)}{v(n)} < \frac{1}{n}$ according to Remark 3.1. $\square$

THEOREM 4.2. *For any valuation function satisfying the complements condition, the unique pure-strategy Nash equilibrium under the best-answer rule is the least efficient strategy profile, in which all users participate in the last round.*

PROOF. Lemma 4.1 indicates if a strategy profile forms an equilibrium and it is not the strategy profile that involves all users playing in the last round, then the strategy profile must have more than one active round. Consider the first active round, call this round t. Suppose i users have played in the tth round. The probability that the question is closed at the end of the tth round is given by $p = \frac{v(i)}{v(n)}$. The expected payoff of a user who plays in round t is p/i. Consider the expected payoff of a user from round t who deviates to the next active round, call this round t'. Suppose other j users have played in round t'. The probability that the question is closed at the end of round t' under this deviation is $p' =$

$\frac{v(i+j)-v(i-1)}{v(n)}$ according to Remark 3.3. The expected payoff of the deviating user is now $p'/(j+1)$. It is easy to see that $p'/(j+1) > p/i$ under the complements condition. Thus any strategy profile that has more than one active round cannot be an equilibrium. $\square$

4.2 Substitutes Case

For the substitutes case, the asker's valuation of having a collection of several pieces of information is lower than the sum of her valuations for individual pieces. The benefit of waiting to form a better answer in this case is therefore relatively low. In contrast to the complements case, the only Nash equilibrium for the substitutes case under the best answer scoring rule, is that all users answer the question in the very first round. This is the most efficient equilibrium, because the asker will get all the answers without waiting.

LEMMA 4.3. *Consider any strategy profile that involves all users playing in the same round. The only one of these strategy profiles that forms a pure-strategy Nash equilibrium is the one in which all users play first, for any valuation function that satisfies the substitutes condition and under the best-answer rule.*

PROOF. Any strategy profile that involves all users playing in the same round, yields an expected payoff of $\frac{1}{n}$ to each user. Let the active round be t. When $t > 1$, a user can deviate by participating in round $t - 1$. The probability that the question is closed at the end of round $t-1$ is $p > \frac{1}{n}$, due to Remark 3.1. The deviating user earns an expected payoff $p > \frac{1}{n}$. Finally, consider the strategy profile consisting of all users participating in the 1st round. If a user deviates by going later, the probability that the question is closed after the first round is greater than $\frac{n-1}{n}$ according to Remark 3.1 and the user's expected payoff is less than $\frac{1}{n}$. $\square$

THEOREM 4.4. *For any valuation function satisfying the substitutes condition, the unique pure-strategy Nash equilibrium under the best-answer rule is the most efficient strategy profile, in which all users participate in the first round.*

PROOF. Lemma 4.3 indicates that if a strategy profile is in equilibrium and it is not the strategy profile that involves all users playing in the first round, then the strategy profile must have more than one active round. Consider the last active round, call this round t. Suppose i users have played in the tth round. This means that $n - i$ users played earlier. The expected payoff for a user who played in the last round is $(1-p)/i$, where $p = \frac{v(n-i)}{v(n)}$ is the probability that the question is closed before the last round was reached. Consider the value of p. We know from Remarks 3.1, 3.2, and 3.3 that the probability of stopping before the last round is reached is greater than $\frac{n-i}{n}$, so the expected payoff of participating in the last active round must be less than $\frac{1}{n}$. A user who participates by going in the last round can deviate by playing in the first active round. Assume that there are j users who play in the first active round, including the user who deviated. The probability that the question is closed at the end of the first round is $\frac{v(j)}{v(n)} > \frac{j}{n}$. So the expected payoff of participating in the first round is greater than $\frac{1}{n}$. $\square$

5. ANALYSIS OF APPROVAL VOTING

Under the best-answer rule, the unique equilibrium for the complements case is all users going last, which is inefficient.

It is possible that by changing the design of the scoring rule, we can induce a useful change in the behavior of users and in particular, enable a more efficient equilibrium. In this section, we consider an *approval-voting* scheme and analyze the equilibrium play of users under this rule.

Under the proposed approval-voting scheme, the asker assigns one point to each of $k > 1$ users, where $k < n$. The number of winners, k, is a design parameter. Note that if $k = 1$, this reduces to the best-answer rule of Yahoo! Answers. The Naver Knowledge-iN forum, in comparison, does allow for askers to select more than one best answer. In approval voting, the winners are the k most recent users to answer before the question is closed, with ties broken uniformly at random. In the special case in which the question is closed and more than k users respond in the most recent round, then a subset of k winners is selected uniformly at random. Similarly, when $k_1 < k$ users respond in the most recent round then each receive one point and a subset of $k - k_1$ users that responded in the previous round are also selected as winners, and so forth.

We consider the approval-voting scheme because it is simple and also because it seems possible that allowing users that responded earlier to receive points will facilitate additional equilibrium, by lessening the incentive in the complements case for every user to wait to the very last moment to respond to the question.

REMARK 5.1. *For any valuation function satisfying the complements condition, and for any strategy profile consisting of all users playing in the same round, a user cannot profitably deviate by playing in an earlier round under the approval-voting rule, for any $k > 1$.*

PROOF. If all users play in the same round, their expected payoff is $\frac{k}{n}$. If a user deviates by going earlier, she receives a payoff of one unit only if the question is closed after the first active round, which occurs with probability $< \frac{1}{n}$ for any valuation function satisfying the complements condition. □

REMARK 5.2. *For any valuation function that satisfies the condition $\frac{v(n-1)}{v(n)} \geq \frac{n-k}{n}$, and for any strategy profile consisting of all users playing in the same round, a user cannot profitably deviate by going in a later round, under the approval-voting rule for $k > 1$ winners.*

PROOF. If all users play in the same round, their expected payoff is $\frac{k}{n}$. If a user deviates by going later, she receives a payoff of one unit only if the question is not closed after the first active round. The question is closed after the first active round with probability $p = \frac{v(n-1)}{v(n)}$. The user who deviates, gets an expected payoff of $1 - p$. We need $1 - p \leq \frac{k}{n}$ or in other words, $p \geq \frac{n-k}{n}$. □

REMARK 5.3. *For any valuation function satisfying the substitutes condition, and for any strategy profile consisting of all users playing in the same round, a user cannot profitably deviate by playing in a later round under the approval-voting rule, for any $k > 1$.*

PROOF. If all users play in the same round, their expected payoff is $\frac{k}{n}$. If a user deviates by going later, she receives a payoff of one unit only if the question is not closed after the first active round, which occurs with probability less than $\frac{1}{n}$ for any valuation function satisfying the substitutes condition. □

REMARK 5.4. *For any valuation function that satisfies the condition $\frac{v(1)}{v(n)} \leq \frac{k}{n}$, and for any strategy profile consisting of all users playing in the same round, a user cannot profitably deviate by playing in an earlier round, under the approval-voting rule for $k > 1$ winners.*

PROOF. If all users play in the same round, their expected payoff is $\frac{k}{n}$. If a user deviates by going earlier, she receives a payoff of one unit only if the question is closed after the first active round, which occurs with probability $\frac{v(1)}{v(n)}$. This deviation is not profitable if and only if $\frac{v(1)}{v(n)} \leq \frac{k}{n}$. □

Lemmas 5.1 and 5.2 are two technical lemmas that together show that any strategy profile in which there are at least two active rounds cannot be a Nash equilibrium for any valuation function that satisfies the complements information criterion. The proof of these lemmas will appear in the full version of the paper.

LEMMA 5.1. *Any strategy profile that has at least two active rounds, where at least k users participate in the last active round, cannot be a pure-strategy Nash equilibrium for any valuation function satisfying the complements condition, under the approval-voting rule for $k > 1$ winners.*

LEMMA 5.2. *Any strategy profile that has at least two active rounds, where less than k users participate in the last active round, cannot be a pure-strategy Nash equilibrium for any valuation function satisfying the complements condition, under the approval-voting rule for $k > 1$ winners.*

THEOREM 5.3. *For any valuation function that satisfies the complements condition and $\frac{v(n-1)}{v(n)} \geq \frac{n-k}{n}$, all users playing in the same round is a pure-strategy Nash equilibrium, for any round, under the approval-voting rule for $k > 1$ winners. Moreover, these are the only pure-strategy Nash equilibria. For any valuation function that satisfies the complements condition and $\frac{v(n-1)}{v(n)} < \frac{n-k}{n}$, the only pure-strategy Nash equilibria is for all users to play in the last round.*

PROOF. From Lemmas 5.1 and 5.2, we know that any strategy profile that has two or more active rounds cannot be a Nash equilibrium. Therefore any equilibrium must have only one active round. We know from Remark 5.1, that if all users are going in the last round, a user cannot profitably deviate for any valuation function that satisfies the complements condition. Therefore, this strategy profile is a Nash equilibrium for any valuation function that satisfies the complements condition. We know from Remark 5.2, that when $\frac{v(n-1)}{v(n)} < \frac{n-k}{n}$ and all users are playing in the same round, a user can profitably deviate by going later. Thus, when $\frac{v(n-1)}{v(n)} < \frac{n-k}{n}$, any strategy profile where all users participate in a round that is not the last round cannot be a Nash equilibrium. We know from Remark 5.2, that when $\frac{v(n-1)}{v(n)} \geq \frac{n-k}{n}$ and all users are playing in the same round, a user cannot profitably deviate by going later. We know from Remark 5.1, that if all users are playing in the same round, a user cannot profitably deviate by going earlier. Thus, when $\frac{v(n-1)}{v(n)} > \frac{n-k}{n}$, any strategy profile where all users participate in a single round is a Nash equilibrium. □

Similar to the case of complements information, Lemmas 5.4 and 5.5 are two technical lemmas that together show

that any strategy profile in which there are at least two active rounds cannot be a pure-strategy Nash equilibrium for any substitutes valuation. The proof of these lemmas will appear in the full version of the paper.

LEMMA 5.4. *Any strategy profile that has at least two active rounds, where at least k users participate in the next-to-last active round, cannot be a pure-strategy Nash equilibrium for any valuation function satisfying the substitutes condition, under the approval-voting rule for $k > 1$ winners.*

LEMMA 5.5. *Any strategy profile that has at least two active rounds, where less than k users participate in the next-to-last active round, cannot be a pure-strategy Nash equilibrium for any valuation function satisfying the substitutes condition, under the approval-voting rule for $k > 1$ winners.*

THEOREM 5.6. *For any asker valuation function satisfying the substitutes condition and $\frac{v(1)}{v(n)} \leq \frac{k}{n}$, all users playing in the same round is a pure-strategy Nash equilibrium, for any round, under the approval-voting rule for $k > 1$ winners. Moreover, these are the only pure-strategy Nash equilibria. For any valuation function that satisfies the substitutes condition and $\frac{v(1)}{v(n)} > \frac{k}{n}$, the only pure-strategy Nash equilibrium is for all users to play in the first round.*

PROOF. From Lemmas 5.4 and 5.5, we know that any strategy profile that has two or more active rounds cannot be a Nash equilibrium. Therefore any equilibrium must have only one active round. We know from Remark 5.3, that if all users are going in the first round, a user cannot profitably deviate for any valuation function that satisfies the substitutes condition. Therefore, this strategy profile is a Nash equilibrium for any valuation function that satisfies the complements condition. We know from Remark 5.4, that when $\frac{v(1)}{v(n)} > \frac{k}{n}$ and all users are playing in the same round, a user can profitably deviate by going earlier. Thus, when $\frac{v(1)}{v(n)} > \frac{k}{n}$, any strategy profile where all users participate in a round that is not the first round cannot be a Nash equilibrium. We know from Remark 5.4, that when $\frac{v(1)}{v(n)} \leq \frac{k}{n}$ and all users are playing in the same round, a user cannot profitably deviate by going earlier. We know from Remark 5.3, that if all users are playing in the same round, a user cannot profitably deviate by going later. Thus, when $\frac{v(n-1)}{v(n)} > \frac{n-k}{n}$, any strategy profile where all users participate in a single round is a Nash equilibrium. □

In contrast to the best-answer rule, the approval-voting rule can enable the most efficient equilibrium outcome for the case of complementary information. However, it is not possible to isolate this as the only equilibrium. For substitutes, we see that approval-voting can sometimes isolate the most efficient equilibrium (as was the case for the best-answer rule.)

The number of winners, $k > 1$, is a tunable parameter in the approval-voting rule that changes the equilibrium structure. The larger k is, the more likely it is to enable the most efficient equilibrium for the complements case, however, the larger k is, the more likely it is to introduce the least efficient equilibrium for the substitutes case. In what follows, we introduce some special cases of complements and substitutes in order to study this tradeoff in a little more detail. We first turn to complements valuations.

DEFINITION 5.7. *Valuation function v satisfies additive complements if and only if, in addition to satisfying Definition 3.1, v satisfies $\delta_{j+1} = \delta_j + c$ with $c > 0$ for all $1 \leq j < n$.*

DEFINITION 5.8. *Valuation function satisfies multiplicative complements if and only if in addition to satisfying Definition 3.1, v satisfies $\delta_{j+1} = \delta_j \cdot c$ with $c > 1$ for all $1 \leq j < n$.*

Corollary 5.9 is a very positive result, that we enable the efficient equilibrium for any additive complements valuations and any value of n, in other words, any additive complements valuation and n pair.

COROLLARY 5.9. *For any valuation function satisfying the additive complements condition, all users playing in the same round is a pure-strategy Nash equilibrium for any round, under the approval-voting rule for $k > 1$ winners. These are the only Nash equilibria of the game.*

In contrast to the previous result, we find that the number of winners k must be relatively large, in order to enable the most efficient equilibrium.

COROLLARY 5.10. *For any valuation function satisfying the multiplicative complements condition, all users playing in the same round is a pure-strategy Nash equilibrium for any round, under the approval-voting rule for $k \geq \frac{n \cdot c^{n-1} \cdot (c-1)}{c^n - 1}$ winners. These are the only Nash equilibria of the game. If $k < \frac{n \cdot c^{n-1} \cdot (c-1)}{c^n - 1}$, then all users playing in the last round is a unique pure-strategy Nash equilibrium.*

DEFINITION 5.11. *Valuation function v satisfies additive substitutes if and only if, in addition to satisfying Definition 3.2, v satisfies $\delta_{j+1} = \delta_j + c$ with $c < 0$ for all $1 \leq j < n$.*

DEFINITION 5.12. *Valuation function v satisfies multiplicative substitutes if and only if in addition to satisfying Definition 3.2, v satisfies $\delta_{j+1} = \delta_j \cdot c$ with $c < 1$ for all $1 \leq j < n$.*

Recall that for the case of substitutes information, the relevant question is to understand when it is possible to isolate the efficient equilibrium from amongst the pooling equilibrium (as in the best-answer scoring rule). For additive substitutes we see that this is not possible for $k \geq 2$:

COROLLARY 5.13. *For any valuation function satisfying the additive substitutes condition, all users playing in the same round, for any round, is a pure-strategy Nash equilibria for any $k > 1$ winners in the approval-voting rule. These are the only Nash equilibria of the game.*

One would need to resort to the best-answer rule (equivalently, approval-voting with $k = 1$), to isolate the efficient equilibrium in this additive substitutes case. And, in picking a value of k there is a clear tradeoff to make between handling additive substitutes and additive complements. On the other hand, we obtain positive results for multiplicative substitutes valuations:

COROLLARY 5.14. *For any valuation function satisfying the multiplicative substitutes condition, all users playing in the first round is a unique pure-strategy Nash equilibrium under the approval-voting rule with $k < \frac{n \cdot (1-c)}{1 - c^n}$. Otherwise, all users playing in the same round, for any round, is a pure-strategy Nash equilibrium for $k \geq \frac{n \cdot (1-c)}{1 - c^n}$, and these are the only Nash equilibria of the game.*

	$c = 1.01$	$c = 2$	$c = 10$
$n = 5$	$k \geq 2$	$k \geq 3$	$k \geq 5$
$n = 10$	$k \geq 2$	$k \geq 6$	$k \geq 9$
$n = 50$	$k \geq 2$	$k \geq 25$	$k \geq 45$

Table 2: Necessary condition on the number of winners, $k > 1$, in approval-voting with *multiplicative complements* in order to enable the most efficient equilibrium.

	$c = 0.99$	$c = 0.5$	$c = 0.1$
$n = 5$	$-$	$k = 2$	$k \leq 4$
$n = 10$	$-$	$k \leq 5$	$k \leq 9$
$n = 50$	$-$	$k \leq 25$	$k \leq 45$

Table 3: Necessary condition on the number of winners, $k > 1$, in approval-voting with *multiplicative substitutes* in order to enable the most efficient equilibrium *as the unique equilibrium*. Entry '$-$' indicates that this is not possible.

In Tables 2 and 3 we illustrate the requirements for multiplicative complements and multiplicative substitutes. One can infer the following kind of difficulty with the approval voting rule: the requirement on k to allow for the most efficient equilibrium for the case of complements valuations tends to be at odds with the requirement on k to isolate the most efficient equilibrium as the unique equilibrium for the case of substitutes valuations. Under the proportional-share rule, proposed in the next section, we do not have this problem since in the case of substitutes valuations, the most efficient outcome is always a unique equilibrium. However, we cannot enable the most efficient equilibrium outcome for all complements valuations with the proportional share rule and so neither rule dominates the other.

6. ANALYSIS OF PROPORTIONAL-SHARE SCORING RULE

In this section, we consider the *proportional-share scoring rule*, and analyze the equilibrium behavior of users. In the proportional-share scoring rule, the asker is given a fixed number of points that she can distribute. Without loss of generality, we normalize the total number of points to distribute to 1 so that each user that participates gets some fraction of a point.

We assume that the asker distributes this point according to her valuation function. More specifically, suppose the question closes after $C \leq T$ active rounds, collects $k \leq n$ pieces of information in total, and at each active round $t \leq C$ there are n_t participants. In the proportional-share scoring rule, the asker distributes $\frac{v(b(1))}{v(k)}$ equally among the n_1 users participated in the active round 1, and, similarly, distributes $\frac{v(b(t)) - v(b(t-1))}{v(k)}$ to the n_t users that participated in active round $t > 1$, where $v(b(t))$ denotes the value of the items received at the end of round t.

In addition to being a natural scoring rule, with each user receiving credit in proportion to the marginal value contributed to the system in the period in which his or her answer is provided, we are interested in this rule because we want to explore whether or not it can remove the inefficient

equilibrium in the complements case. While the approval-voting rule was successful in introducing the efficient equilibrium (under certain conditions on the asker valuation), it was unable to isolate this as the only equilibrium. The proportional-share scoring rule is designed to provide more credit to early responders than the approval-voting rule in order to mitigate this problem.

We first present a lemma on the behavior of users when there is only one active round.

LEMMA 6.1. *For any strategy profile in which all users play in the same round, and if $\frac{v(1)}{v(n)} \leq 1 - \sqrt{\frac{n-1}{n}}$, a user cannot profitably deviate by going in an earlier round under the proportional-share rule. For any strategy profile in which all users play in the same round, and if $\frac{v(n-1)}{v(n)} \geq 1 - \sqrt{\frac{1}{n}}$, a user cannot profitably deviate by going in a later round.*

PROOF. Consider the strategy profile consisting of all users going in the same round. The expected payoff of each user is $\frac{1}{n}$. The expected payoff of a user who deviates by playing in a later round is $(1-p) \cdot (1-p)$, where $p = \frac{v(n-1)}{v(n)}$. In order for this deviation not to be profitable, we need $(1-p)^2 \leq \frac{1}{n}$, or equivalently, $p \geq 1 - \sqrt{\frac{1}{n}}$. The expected payoff of a user who deviates by playing in an earlier round is $p + (1-p) \cdot p$, where $p = \frac{v(1)}{v(n)}$. In order for this deviation not to be profitable, we need $p + (1-p) \cdot p \leq \frac{1}{n}$, or equivalently, $p \leq 1 - \sqrt{\frac{n-1}{n}}$. $\square$

Applying Lemma 6.1, we get the following theorem, which characterizes the equilibrium structure when we restrict our attention to pooling equilibrium.

THEOREM 6.2. *Consider the proportional-sharing rule and pooling equilibria where all users participate in a single active round. If $\frac{v(n-1)}{v(n)} \geq 1 - \sqrt{\frac{1}{n}}$ and $\frac{v(1)}{v(n)} > 1 - \sqrt{\frac{n-1}{n}}$, the strategy profile consisting of all users going in the first round is a unique pure-strategy pooling Nash equilibrium. If $\frac{v(1)}{v(n)} \leq 1 - \sqrt{\frac{n-1}{n}}$ and $\frac{v(n-1)}{v(n)} < 1 - \sqrt{\frac{1}{n}}$, the strategy profile consisting of all users going in the last round is a unique pure-strategy pooling Nash equilibrium. If $\frac{v(1)}{v(n)} \leq 1 - \sqrt{\frac{n-1}{n}}$ and $\frac{v(n-1)}{v(n)} \geq 1 - \sqrt{\frac{1}{n}}$, then any strategy profile in which there is only one active round can be a pure-strategy pooling Nash equilibrium. Finally if $\frac{v(1)}{v(n)} > 1 - \sqrt{\frac{n-1}{n}}$ and $\frac{v(n-1)}{v(n)} < 1 - \sqrt{\frac{1}{n}}$, there is no pure-strategy pooling Nash equilibrium.*

We summarize the results of Theorem 6.2 in Table 4. Although Theorem 6.2 completely characterizes the equilibrium structure when we restrict attention to pooling equilibrium, we are unable to rule out the possibility of separating equilibria. We know that separating equilibria do exist, however, they appear to hold for a very narrow range of valuation functions.

Theorem 6.2 gives us a partial characterization of the equilibrium structure under special cases of complements valuations. We will return to this below. For now we return to substitutes valuations and see that we retain the same property as for the best-answer rule and isolate the efficient equilibrium as the only pure-strategy Nash equilibrium.

	$\frac{v(1)}{v(n)} > 1 - \sqrt{\frac{n-1}{n}}$	$\frac{v(1)}{v(n)} \le 1 - \sqrt{\frac{n-1}{n}}$
$\frac{v(n-1)}{v(n)} \ge 1 - \sqrt{\frac{1}{n}}$	all go first	any pooling
$\frac{v(n-1)}{v(n)} < 1 - \sqrt{\frac{1}{n}}$	no pooling	all go last

Table 4: Summary of partial equilibrium characterization results for the proportional-sharing rule. (Theorem 6.2)

THEOREM 6.3. *For any valuation function that satisfies the substitutes condition, the only pure-strategy Nash equilibrium under the proportional-share scoring rule is the strategy profile consisting of all n users going first.*

PROOF. Consider any strategy profile in which there are at least two active rounds. Suppose that j users play in the last active round. The expected payoff of a user who participates in the last active round is $(1-p) \cdot \frac{(1-p)}{j}$, where $p = \frac{v(n-j)}{v(n)}$. Consider the expected payoff of a user in the last active round who deviates by going in the first active round. Suppose that i users participate in the first active round, including the user who deviated. His expected payoff is at least $\frac{p'}{i}$, where $p' = \frac{v(i)}{v(n)}$. For any valuation function that satisfies the substitutes condition, we know that $\frac{p'}{i} \ge \frac{1-p}{j}$, so $\frac{p'}{i} > (1-p) \cdot \frac{(1-p)}{j}$. Thus any strategy profile in which there are at least two active rounds cannot be a Nash equilibrium. Consider any strategy profile in which there is only one active round. Lemma 6.1 tells us that if all users are going in the first round, no user has incentive to deviate if and only if the valuation function satisfies the condition: $\frac{v(n-1)}{v(n)} \ge 1 - \sqrt{\frac{1}{n}}$, which is always satisfied by any valuation function that satisfies the substitutes condition. Lemma 6.1 also tells us that if all users are going in the same round, that is not the first round, no user has incentive to deviate by going earlier if and only if the valuation function satisfies the condition: $\frac{v(1)}{v(n)} \le 1 - \sqrt{\frac{n-1}{n}}$. However, this condition is never satisfied by any valuation function that satisfies the substitutes condition, therefore any strategy profile in which all users play in the same round, that is not the first round, cannot be a Nash equilibrium. $\square$

Unlike the case of the approval-voting rule, it is not always possible to achieve the most efficient equilibrium for complements valuations. On the other hand, we do not need to worry about more inefficient equilibria being introduced for the case of substitutes information. To better understand the condition on positive results for complements valuations, we can again consider the special case of additive complements and multiplicative complements. Note that in this case there is no design parameter, and thus no explicit tradeoff that needs to be made between good performance across different valuation models. For additive complements, we see that the value of c needs to be quite small with respect to δ_1 to enable a good equilibrium. We see that given a fixed additive complements valuation, it may not be possible to have the most efficient equilibrium outcome, depending on the value of n. In other words, we cannot enable the most efficient equilibrium outcome for all valuation and n pairs, when the valuation satisfies additive complements.

COROLLARY 6.4. *Consider the proportional-share rule and $n \ge 4$. If the valuation function of the asker satisfies additive complements with $c < \frac{2 \cdot \delta_1}{\sqrt{n(n-1)}}$, the most efficient outcome of all users going in the first round is a pure-strategy Nash equilibrium. Moreover, this is the only strategy profile that has one active round that is a pure-strategy Nash equilibrium (ruling out other pooling equilibrium).*

The following corollary tells us that we can enable the most efficient equilibrium outcome for the case of multiplicative complements, however, we can only do so when the value of c and n are both relatively small. Again we see that, depending on the valuation function and n pair, it may not be possible to have the most efficient equilibrium outcome. This fact is also illustrated in Table 5.

COROLLARY 6.5. *Consider the proportional-share rule. If the valuation function of the asker satisfies multiplicative complements with $\frac{c^n - c}{c^n - 1} < \frac{\sqrt{n-1}}{\sqrt{n}}$ and $\frac{c^{n-1}(c-1)}{c^n - 1} \le \frac{1}{\sqrt{n}}$, the most efficient outcome of all users going in the first round is a pure-strategy Nash equilibrium. Moreover, this is the only strategy profile that has one active round that is a Nash equilibrium (ruling out other pooling equilibrium).*

We present in Table 5 some example values of n and c and whether the all going first equilibrium can be enabled for the multiplicative complements case.

	$c = 1.01$	$c = 2$	$c = 10$
$n = 5$	yes	no	no
$n = 10$	yes	no	no
$n = 50$	yes	no	no

Table 5: Examples on whether the most efficient equilibrium can be enabled for the case of *multiplicative complements*.

7. CONCLUSIONS

We believe that appropriate incentive design can help to improve the information quality in question and answer forums. In studying this, we have introduced a simple, game-theoretic model of a question and answer forum such as Yahoo! Answers. We analyze the best-answer scoring rule, which models that of Yahoo! Answers, and show that it is effective with information items that are substitutes but enables only the least efficient outcome, in which every user plays in the very last round, in the case of complements. In considering the effect of different scoring rules on the equilibrium structure of the game, we have identified two scoring rules that lead to efficiency-improving changes in the equilibrium. Specifically, the approval-voting rule can enable the most efficient equilibrium for any complements valuation and n pair, with an appropriately chosen value of k (the number of winners). On the other hand, for any substitutes valuation and n pair, the approval-voting rule can introduce the least efficient outcome as an equilibrium, depending on the value of k. The tunable parameter, k, enables a tradeoff between the benefit of this scoring rule for the case of complements information and the disadvantage of introducing this scoring rule for the case of substitutes information.

The proportional-share rule, in comparison, never introduces a less efficient equilibrium for the case of substitutes

valuations. Moreover, for certain valuations it is possible with the proportional-share rule to isolate the efficient equilibrium while ruling out all other pooling equilibrium for complements valuations. On the other hand, there are some complements valuations and n pairs for which the proportional-share rule does not allow the efficient equilibrium and we are not, in general, able to rule out additional separating equilibria (in which users respond in different rounds) for the case of complements valuations.

Taken altogether, while the approval-voting rule (for a small enough k) and the proportional-score rules seem to have more desirable properties than the best-answer rule, we do not yet have a clear ordering between our two new rules. In considering the appropriate rule, one must also remember that the approval-voting rule requires less information from the asker to allocate points while the proportional-share rule requires the asker to allocate a value to each answer (or to each set of answers, when multiple answers are received in the same round) in a sequence of answers.

Clearly there are a lot of avenues for future work. In addition to characterizing the complete equilibrium structure (including split equilibria) for complements valuations in the proportional-share rule, one could study variations on our simple model, such as answerers that have different valued pieces of information (from the asker side) and answerers that have overlapping information. Such extensions would remove symmetry of the answerers and move us towards a richer model. It would also be interesting to incorporate the fact that some users are partially motivated by altruistic reasons into our model. Another direction for future work would be to model the cost to the asker of combining information provided by multiple users in the same round, leading to the identification of scoring rules that promote "build" equilibrium where the user responses are optimally sequenced and build of each other. Finally, another extension is to consider information cascade effects, wherein one user's response triggers another user to recall a new piece of information that would have not been available if not triggered by the first user.

Acknowledgements

8. REFERENCES

[1] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Everyone knows something: Examining knowledge sharing on Yahoo answers. In *Proc. 17th Intl. Conf. on the World Wide Web (WWW)*, 2008.

[2] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *Proc. 1st ACM Intl. Conf. on Web Search and Data Mining (WSDM)*, 2008.

[3] Y. Benkler. Coase's Penguin, or, Linux and the Nature of the firm. *The Yale Law Journal*, 112, 2002.

[4] J. Bian, Y. Liu, E. Agichtein, and H. Zha. Finding the right facts in the crowd: Factoid question and answering over social media. In *Proc. 17th Intl. Conf. on the World Wide Web (WWW)*, 2008.

[5] F. M. Harper, D. Moy, and J. A. Konstan. Facts or friends? Distinguishing informational and conversational questions in social Q&A sites. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI)*, 2009.

[6] F. M. Harper, D. Raban, S. Rafaeli, and J. A. Konstan. Predictors of answer quality in online q&a sites. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI)*, 2008.

[7] C.-J. Ho, T.-H. Chang, and J. Y. jen Hsu. Photoslap: A multi-player online game for semantic annotation. In *Proc. 22nd Conf. on Artificial Intelligence (AAAI)*, 2007.

[8] G. Hsieh and S. Counts. mimir: A market-based real-time question and answer service. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI)*, 2009.

[9] S. Jain and D. C. Parkes. A game-theoretic analysis of games with a purpose. In *Proc. 4th Intl. Workshop on Internet and Network Economics (WINE)*, 2008.

[10] J. Jeon, W. B. Croft, and J. H. Lee. Finding similar questions in large question and answer archives. In *Proc. ACM Conf. on Information and Knowledge Management (CIKM)*, 2005.

[11] P. Jurczyk and E. Agichtein. Discovering authorities in question answer communities by using link analysis. In *Proc. ACM Conf. on Information and Knowledge Management (CIKM)*, 2007.

[12] P. Jurczyk and E. Agichtein. Hits on question answer portals: exploration of link analysis for author ranking. In *Proc. 30th ACM SIGIR Conf.*, 2007.

[13] Y. Liu, J. Bian, and E. Agichtein. Predicting information seeker satisfaction in community question answering. In *Proc. 31st ACM SIGIR Conf.*, 2008.

[14] K. K. Nam, M. S. Ackerman, and L. A. Adamic. Questions in, Knowledge iN?: A study of Naver's question answering community. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI)*, 2009.

[15] J. Pouwelse, P. Garbacki, D. Epema, and H. Sips. Pirates and Samaritans: A decade of measurements on peer production and their implications for net neutrality and copyright. *Telecommunications Policy Journal*, 32:701–712, 2008.

[16] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI)*, 2004.

[17] L. von Ahn and L. Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.

[18] D. M. Wilkinson. Strong regularities in online peer production. In *Proc. 9th ACM Conf. on Electronic Commerce (EC)*, 2008.

[19] J. Yang, L. A. Adamic, and M. S. Ackerman. Crowdsourcing and knowledge sharing: Strategic user behavior on Taskcn. In *Proc. 9th ACM Conf. on Electronic Commerce (EC)*, 2008.

[20] J. Zhang, M. Ackerman, and L. Adamic. Expertise networks in online communities: Structure and algorithms. In *Proc. 16th Intl. Conf. on the World Wide Web (WWW)*, 2007.