Crowdsourcing with All-pay Auctions: a Field Experiment on Taskcn[*]

Tracy Xiao Liu[†] Jiang Yang[‡] Lada A. Adamic[§] Yan Chen[¶]

September 16, 2013

Abstract

To explore the effects of different incentives on crowdsourcing participation and submission quality, we conduct a randomized field experiment on Taskcn, a large Chinese crowdsourcing site using mechanisms with features of an all-pay auction. In our study, we systematically vary the size of the reward, as well as the presence of a soft reserve, or early high-quality submission. We find that a higher reward induces significantly more submissions and submissions of higher quality. In comparison, we find that high-quality users are significantly less likely to enter tasks where a high quality solution has already been submitted, resulting in lower overall quality in subsequent submissions in such soft reserve treatments.

Keywords: crowdsourcing, field experiment, all-pay auctions

JEL Classification: C93, D44

[*]We thank Eytan Adar, Teck-Hua Ho, Jeff MacKie-Mason, John Morgan, Paul Resnick, Rahul Sami, Ella Segev, Aner Sela, Jeff Smith, Neslihan Uhler, Lixin Ye and seminar participants at Arkansas, Chapman, Essex, Florida State, Michigan, Münich, Ohio State, National University of Singapore, Penn State, UCL, UC-Santa Barbara, Zürich, the 2010 International ESA meetings (Copenhagen), the ACM EC'11 Workshop on Crowdsourcing and User Generated Content (San Jose, CA), and the 2012 NSF/NBER Decentralization Conference (Caltech) for helpful discussions and comments, and Lei Shi for excellent research assistance. The financial support from the National Science Foundation through grant no. SES-0962492 and IIS-0948639 is gratefully acknowledged.

[†]Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China. Email: tracy.umich@gmail.com.

[‡]IBM CHQ, 1 New Orchard Road, Armonk, NY 10504-1722. Email: yangjiang78@gmail.com.

[§]School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109-2112. Email: ladamic@umich.edu.

[¶]School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109-2112. Email: yanchen@umich.edu.

1 Introduction

The Internet has transformed how work is done, from allowing geographically-dispersed workers to collaborate to enabling task solutions to be globally crowdsourced (Howe 2006, Howe 2008, Kleeman, Voss and Rieder 2008). The term crowdsourcing typically refers to the open solicitation of effort on a well-defined task to a community (crowd) to obtain a submitted solution before a deadline. Crowdsoucing has become an increasingly popular choice for tasks, such as translation, programming, website design and open innovation. Various crowdsourcing mechanisms have been used in practice, including voluntary contribution, monopoly and contests. In this study, we focus on a family of contest mechanisms with features of an all-pay auction.

Well-known crowdsourcing sites, such as Taskcn in China and TopCoder in the United States, have adopted variants of contests as their reward mechanisms. In the simplest form of this contest, a requester posts a task and respective reward; any user can then submit a solution to the task. Since every user who submits a solution expends effort, regardless of whether she wins, this simplest form of contest mechanism shares many features of a first-price all-pay auction, where everyone expends effort, but only the winner receives a reward. We subsequently model it as an all-pay auction. To our knowledge, our study is among the earliest field experiments to explore the effect of the reward level and reserve quality on participation and submission quality in such a competitive setting.

In addition to allowing for competition, crowdsourcing sites experiment with other features of the contest mechanisms. On Taskcn, for example, sequential all-pay auctions, where late entrants can observe the content of earlier submissions, used to be the only exchange mechanism. Recently, users were given the ability to password-protect their solutions.[1] Theoretically, if all users password-protect their solutions, a sequential all-pay auction is transformed into a simultaneous all-pay auction. On the other hand, if only a fraction of users password-protect their solutions, the contest becomes a hybrid sequential/simultaneous all-pay auction. By contrast, on TopCoder, every submission is sealed. The two sites also differ in their user reputation systems. On Taskcn, for every 100 CNY a contestant wins, she accrues 1 credit. On TopCoder, the platform calculates a skill rating for each participant on the basis of her past performance in contests (Boudreau and Lakhani 2012). This skill rating can influence her reputation and thus her career path as a software developer. In each system, design features that influence participant motivation can include monetary rewards, reputation rewards, or the opportunity to compete or collaborate. Given the options available, an evaluation of the various design features in contest mechanisms can potentially inform and thus improve the design and quality outcome of crowdsourcing mechanisms.

To evaluate the effects of both reward size and early high-quality submission (i.e., a soft reserve) on

[1] Taskcn uses two methods to protect solution content. One is to use a pre-paid service provided by the site; the other is to submit a solution with password protection and send the password to the requester by email.

overall participation levels and submission quality, we conduct a randomized field experiment on Taskcn. We choose Taskcn because we are interested in the sequential features of site, which enable us to explore the effects of early high-quality submissions. In our field experiment, we post different translation and programming tasks on Taskcn. The tasks are of similar difficulty, but the reward is exogenously varied. In addition, for a subset of tasks, we pose as a user and submit a high quality solution early in the contest. Unlike earlier field experiments on Google Answers (Chen, Ho and Kim 2010) and Amazon's Mechanical Turk (Mason and Watts 2009), in the competitive setting of Taskcn, we find significant reward effects on both participation levels and submission quality, which is consistent with our theoretical predictions. However, we also find that experienced users respond to our experimental treatments differently from inexperienced ones. Specifically, experienced users are more likely to select tasks with a high reward than inexperienced users. Furthermore, they are less likely to select a task where a high quality solution has already been posted. As a result, our reserve treatments result in significantly lower average submission quality than those without a reserve. While prior empirical papers have investigated the impact of prize amount on tournament outcomes (Ehrenberg and Bognanno 1990), to our knowledge, no one has exploited the impact of soft reserves on outcomes, even though many tournaments are in actuality sequential. Therefore, this paper deepens our understanding of the basic mechanisms of real world contests.

2 Field Setting: Taskcn

Since the crowdsourcing site Taskcn (`http://www.taskcn.com/`) was founded in 2006, it has become one of the most widely used online labor markets in China. On Taskcn, a requester first fills out an online request form with the task title, the reward amount(s), the closing date for submissions, and the number of submissions that will be selected as winners. When the closing date is reached, the site sends a notice to the requester who posts the task, asking her to select the best solution(s) among all the submissions. The requester can also choose the best solution(s) before the closing date. In this case, users are informed that a solution has been selected and the task is closed. Once the task is closed, the winner receives 80% of the reward and the site retains 20% of the reward as a transaction fee. As of August 24, 2010, Taskcn had accumulated 39,371 tasks, with rewards totaling 27,924,800 CNY (about 4.1 million USD).[2] Of the 2,871,391 registered users on Taskcn, 243,418 have won at least one reward.

To inform our field experiment, we first crawled and analyzed the full set of tasks posted on Taskcn from its inception in 2006 to March 2009. As of the time of our crawl, tasks were divided into 15 categories, including requests for graphic, logo and web designs; translations; business names and slogan suggestions; and computer coding. Note that challenging tasks, such as those involving graphic design and website

[2] The exchange rate between the US dollar and the Chinese yuan was 1 USD = 6.8 CNY in both 2009 and 2010.

building, have the highest average rewards (graphic design: 385 CNY; web building: 460 CNY) as they require higher levels of expertise, whereas tasks asking for translations, or name and slogan suggestions offer lower average rewards (translation: 137 CNY; name/slogan: 170 CNY). In addition, most tasks (76.5%) select only one submission to win the reward.

Within the site, each ongoing task displays continually updated information on the number of users who have registered for the task and the number of submissions. Unless protected, each solution can be viewed by all users. In August 2008, Taskcn began offering a solution protection program, which hides the content of one's submission from other users. To protect a submission, a user must enroll in the password protection program and pay a fee.[3] Password-protected submissions are displayed to the requester ahead of other submissions. As an alternative solution protection option, many users on Taskcn protect their solution content by submitting an encrypted solution and sending the password to the requester. The solution protection options make the contest mechanism on Taskcn a hybrid simultaneous/sequential all-pay auction.

Once on the site, after reading a task specification and any unprotected submitted solutions, a user can decide whether to register for a task and submit a solution before the closing date. A user can also view the number of credits accrued by previous submitters. The number of credits corresponds to the hundreds of CNY a user has won by competing in previous tasks, and may signal either expertise or likelihood of winning. Even after a user registers for a task, she may decide not to submit a solution. Furthermore, there is no filter to prevent low quality solutions.

Given Taskcn's design, it is of interest to understand how users respond to different incentives induced by design features. For example, one key question is whether a higher reward induces more submissions and submissions of higher quality. Another question revolves around the impact of an early high quality submission on the quality of subsequent submissions. We also examine whether certain types of tasks are more likely to elicit password-protected solutions, as well as whether experienced and inexperienced users respond differently to incentives.

3 Literature Review

Our study is closely related to the large body of economic literature comprised of studies of contests (Tullock 1980), rank-order tournaments (Lazear and Rosen 1981) and all-pay auctions (Nalebuff and Stiglitz 1983, Dasgupta 1986, Hillman and Riley 1989). In each of these mechanisms, competing agents have the opportunity to expend scarce resources to affect the probability of winning prizes. However, they differ in how agent expenditure is translated into the probability of winning.

To illustrate the similarities and differences across the three types of models, we use a nested formulation

[3] The fee for the password-protection program ranges from 90 CNY for three months to 300 CNY for a year.

(see Dechenaux, Kovenock and Sheremeta 2012). Suppose that contestant i expends effort, e_i. Let the cost of her effort be $c(e_i)$, and the output of her effort be $q_i = e_i + \varepsilon_i$, where ε_i is a random variable drawn from a common distribution. Player i's probability of winning the contest is therefore given by the following contest success function:

$$p_i(q_i, q_{-i}) = \frac{q_i^r}{\sum_{j=1}^{n} q_j^r}, \tag{1}$$

where r is a sensitivity parameter. Note that a simple version of a Tullock contest can be obtained when there is no noise in the performance function, or $\varepsilon_i = 0$, with a linear cost function $c(e_i) = e_i$, and a probabilistic winner determination, $r \in [0, \infty)$. Likewise, a simple version of the all-pay auction can be obtained when there is no noise in the performance function, or $\varepsilon_i = 0$, with a linear cost function, $c(e_i) = e_i$, and no uncertainty in the winner determination, $r = \infty$. Finally, a simple rank-order tournament can be obtained when there is noise in the performance function, $q_i = e_i + \varepsilon_i$, with an identical cost function $c(e_i) = c(e)$, and no uncertainty in winner determination, $r = \infty$. Therefore, in a Tullock contest, the agent with the best performance is not necessarily the winner, whereas in both all-pay auctions and rank-order tournaments, the agent with the best performance wins. Note that an all-pay auction assumes the effort and output equivalence, whereas a rank-order tournament assumes that effort translates noisily to the output. We refer the reader to Konrad (2009) for a review of the relevant theoretical literature and Dechenaux, Kovenock and Sheremeta (2012) for a survey of the experimental literature.

Recent extensions of the above classical theoretical framework have also been applied to the design of innovation contests. For example, Terwiesch and Xu (2008) provide a categorization of different innovation tasks and a corresponding theoretical analysis. In their framework, tasks can be categorized based on the relative importance of expertise and the degree of uncertainty in the performance function. Specifically, agent performance in *expertise-based projects* is driven primarily by the level of expertise in the domain area and the level of contestant effort, with little uncertainty in the outcome. Examples of expertise-based tasks include translations and well-specified simple programming tasks. In comparison, *ideation* and *trial-and-error projects* involve some degree of uncertainty in the performance. Examples of such tasks include logo design. In a simultaneous innovation contest, Terwiesch and Xu (2008) demonstrate that, while the equilibrium effort decreases with the number of participants in an expertise-based project, the benefit of increased participation, or diversity, can mitigate its negative effect on the average effort level from participants in ideation or trial-and-error projects.

The theoretical framework of Terwiesch and Xu (2008) provides a useful lens for examining the design features of the best-known crowdsourcing sites using contests. Using their framework, we first examine sites that use solely simultaneous contests. We then apply it to the sequential/simultaneous hybrid structure made possible in the Taskcn community.

Two sites that use simultaneous contests are InnoCentive and TopCoder. On InnoCentive, problems are posted from diverse industries including aerospace, biotechnology and pharmaceuticals. Most problems have been attempted unsuccessfully by internal scientists. Therefore, the problems posted to the community are typically challenging, with an important uncertainty component in the performance function. In an empirical study of 166 scientific challenges posted on InnoCentive, Jeppesen and Lakhani (2010) find that both technical and social marginality play an important role in explaining individual success in specific problem-solving. The positive effect of diversity in solving problems with a significant uncertainty component is consistent with the predictions of Terwiesch and Xu (2008) for ideation or trial-and-error projects.

Another well-known contest-based crowdsourcing site, TopCoder.com, uses simultaneous contests to source software development tasks. Using historical data from TopCoder, Archak (2010) finds that reward level is a significant determinant of solution quality. Furthermore, he finds that highly-rated contestants tend to sign up early in the registration phase, thus deterring the entry of other contestants. In an empirical analysis of the effects of competition within TopCoder, Boudreau, Lacetera and Lakhani (2011) find that, while the average solution quality for easier tasks decreases with a larger number of competitors, the average solution quality for challenging tasks increases with greater competition. If more challenging tasks involve more uncertainty in performance, this empirical finding is again consistent with the predictions of Terwiesch and Xu (2008). Finally, in a recent field experiment on TopCoder, Boudreau and Lakhani (2012) find a significant effect of sorting (based on taste for competition), which can be explained by higher effort being expended by those who prefer competition, rather than unobserved skills.

In comparison to InnoCentive and TopCoder, Taskcn hosts a large number of expertise-based projects, ideation and trial-and-error projects. In a study using data crawled from Taskcn, Yang, Adamic and Ackerman (2008a) find a low correlation between reward size and the number of submissions. Importantly, using human coders for a random sample of 157 tasks, the authors find a positive and significant correlation between reward size and the level of skill required for the corresponding task, indicating that reward size is *endogenously* related to task difficulty. This difference in required skill may impact participation levels. Therefore, to investigate the causality between reward and contestant behavior, it is important to *exogenously* vary the reward level while controlling for task difficulty. In another study, DiPalantino and Vojnovic (2009) construct a theoretical all-pay auction model for crowdsourcing. Using a subsample of Taskcn data, they find that participation rates increase with reward at a decreasing rate, consistent with their theoretical prediction. However, neither study explores the impact of reward level on submission quality. Thus, our study contributes to the research on crowdsourcing by investigating both participation levels and solution quality using a randomized field experiment.

As mentioned, compared to the studies reviewed above, our study represents the first randomized field experiment on a contest-based crowdsourcing site. By exogenously varying the reward level and the presence of a soft reserve, we can more precisely evaluate the reward and reserve effects on both participation levels and solution quality, while preserving the realism of a natural field setting (Harrison and List 2004).

In our study, we use only expertise-based projects, such as translation and simple programming tasks, where each task is well defined, and its evaluation is straightforward and objective. Our choice of tasks implies that uncertainty in performance plays a relatively minor role. In our theoretical benchmark presented in Section 4, we make the simplifying assumption that there is no uncertainty in either the performance function ($\varepsilon_i = 0$) or the winner determination ($r = \infty$). That is, we simplify the model to the case of an all-pay auction.

Table 1: All-Pay Auction Literature: Theoretical Studies and Laboratory Experiments

	Simultaneous All-Pay Auctions	
	Theory	Laboratory Experiments
Complete Information	Baye, Kovenock and de Vries (1996) Bertoletti (2010) Anderson, Goeree and Holt (1998)	Potters, de Vries and van Winden (1998) Davis and Reilly (1998) Gneezy and Smorodinsky (2006) Lugovskyy, Puzzello and Tucker (2010) Liu (2011)
Incomplete Information	Amann and Leininger (1996) Krishna and Morgan (1997) Fibich, Gavious and Sela (2006) DiPalantino and Vojnovic (2009)	Noussair and Silver (2006)
	Sequential All-Pay Auctions	
	Theory	Laboratory Experiments
Complete Info.	Konrad and Leininger (2007)	Liu (2011)
Incomplete Info.	Segev and Sela (2012)	

Table 1 summarizes the theoretical and experimental studies relating to all-pay auctions, organized by the timing of bids and the relevant information structures. Within this area of research, Baye et al. (1996) provide a theoretical characterization of the mixed strategy Nash equilibrium for a simultaneous all-pay auction under complete information. Bertoletti (2010) extends this model to investigate the role of a reserve price and finds that a strict reserve price increases allocation efficiency. In an incomplete information setting, Krishna and Morgan (1997) and Amann and Leininger (1996) characterize the Bayesian Nash equilibrium separately under different informational assumptions.[4] While the previous studies all focus on a single

[4]Krishna and Morgan's model (1997) assumes that in a n-player game, each agent's signal is affiliated and symmetrically distributed, whereas Amann and Leininger (1996) consider a two-player incomplete information all-pay auction with an asymmetric

auction, DiPalantino and Vojnovic (2009) investigate a multiple all-pay auction model, where contestants choose between tasks with different rewards. In their study, DiPalantino and Vojnovic (2009) show that a higher reward increases participation levels. However, as mentioned, they do not examine the effect of reward on submission quality.

In addition to the theoretical literature, a number of laboratory experiments test the predictions of simultaneous all-pay auction models (Table 1, right column). Under complete information, most studies find that players overbid relative to the risk neutral Nash equilibrium predictions in early rounds, but then learn to reduce their bids with experience (Davis and Reilly 1998, Gneezy and Smorodinsky 2006, Lugovskyy et al. 2010, Liu 2011). One exception to this finding is Potters et al. (1998), who find bidding behavior consistent with Nash equilibrium predictions.[5] Rent overdissipation as a result of overbidding can be (partially) explained by a logit equilibrium (Anderson et al. 1998). In comparison, in an incomplete information and independent private value environment, Noussair and Silver (2006) find that revenue exceeds the risk-neutral Bayesian Nash equilibrium prediction, due to aggressive bidding by players with high valuations and passive bidding by those with low valuations. Both findings of overbidding and behavioral heterogeneity among different types of players are consistent with risk aversion (Fibich et al. 2006).

Compared to research on simultaneous all-pay auctions, fewer studies investigate sequential all-pay auctions. Relevant to our study, in a complete information sequential all-pay auction model with endogenous entry, Konrad and Leininger (2007) characterize the subgame perfect Nash equilibrium, where players with the lowest bidding cost enter late, while others randomize between early and late entry. Extending this work to an incomplete information sequential all-pay auction setting, Segev and Sela (2012) demonstrate that giving a head start to preceding players improves contestant effort. Furthermore, in a laboratory test of the Konrad and Leininger (2007) model, Liu (2011) finds that players learn to enter late in all treatments.

It is worth noting that there is also a growing literature comparing all-pay auctions with other mechanisms in the fundraising context, which has a public good component, differentiating it from our study. We refer the reader to Carpenter, Matthews and Schirm (2010) for a summary of this literature and the references therein.

Finally, a four-page summary of the results of our current paper appears in a conference proceeding (Liu, Yang, Adamic and Chen 2011). In the four-page summary, we include a condensed version of the introduction, a two-paragraph summary of our theoretical framework without any proofs, a summary of our experimental design, a statement of the first four hypotheses, and a summary of our results 1 to 6, without any tables or figures as supporting evidence. Thus, the current paper extends the logic and justification of

value distribution.

[5]The combination of several design features might explain the results in Potters et al. (1998), including a small group size ($n = 2$), stranger matching, a relatively large number of periods (30), and a per-period endowment rather than a lump sum provided at the beginning of the experiment.

the results presented in the summary.

Compared to the existing literature on all-pay auctions, we conduct a field experiment on Taskcn, where features of sequential and simultaneous all-pay auctions coexist. As such, our results have the potential to inform the design of all-pay auctions for crowdsourcing sites.

4 Theoretical Framework

In this section, we outline the theoretical framework we use to derive our comparative statics results, which serve as the basis for our experimental design and hypotheses. In doing so, we follow the model in Segev and Sela (2012), extending their model to incorporate the effects of a reward and a reserve price on bidding strategies in sequential and simultaneous all-pay auctions.

In our model, a single task is crowdsourced through an all-pay auction. The reward for the task is $v \geq 1$. There are n users, each differing in ability. Let $a_i \geq 0$ be user i's ability, which is her private information. User abilities are i.i.d. draws from the interval [0,1] according to the cumulative distribution function, $F(x)$, which is common knowledge. For user i, a submission of quality q_i costs q_i/a_i, indicating that it is less costly for a high ability user to submit a solution of a given quality than a low ability user. User i's expected payoff is thus $\left[v \prod_{j\neq i} F_j(q_j < q_i) - \frac{q_i}{a_i}\right]$. The user with the best quality solution wins the reward; all users incur time and effort in preparing their solutions.

To examine the effects of a reserve on participation levels and submission quality, we include a reserve quality, $q_0 \geq 0$. In this case, user i wins a reward equal to v if and only if the quality of her submission is the highest among the submissions *and* if it is at least as high as the reserve, i.e., $q_i \geq \max\{q_j, q_0\}, \forall j \neq i$.

In what follows, we separately characterize the comparative statics results for the sequential and simultaneous all-pay auctions under incomplete information. For the sequential case (Section 4.1), Propositions 1 through 3 also require the assumption that the ability distribution function is from the family, $F(x) = x^c$, where $0 < c < 1$. In comparison, for the simultaneous case (Section 4.2), for Propositions 4 through 6, we assume that $H_i(x) = \prod_{j\neq i} F(x) = F^{n-1}(x)$ is strictly concave and that $H_i(0) = 0$. However, we do not assume that $F(x) = x^c$. All proofs and examples are relegated to Appendix A of the Online Appendices.

Our comparative statics concern the effects of reward and reserve on participation levels and submission quality. While our model assumes exogenous participation, i.e., each user i submits a solution with quality $q_i \geq 0$, we measure *participation level* (1) theoretically as the *ex ante* likelihood that a user submits a solution of positive quality, $P_i(q_i > 0)$; and (2) empirically as the number of submissions of positive quality. In comparison, our definition of submission quality is standard, measured theoretically by the expected submission quality, $Q_i(q_i)$, and empirically by the submission quality evaluated by trained raters.

4.1 Sequential All-pay Auctions under Incomplete Information

When users cannot protect their solutions, the competitive process on Taskcn approximates a sequential all-pay auction, where solutions are submitted sequentially and the best solution is selected as the winner. Following Segev and Sela (2012), we first characterize the subgame perfect equilibria of a sequential all-pay auction under incomplete information.

In a sequential auction, each of n users enters the auction sequentially. In period i, where $1 \leq i \leq n$, user i submits a solution with quality, $q_i \geq 0$, after observing previous submissions. For technical reasons, we assume that ties are broken in favor of the late entrant.[6] Using backward induction, we characterize the equilibrium bidding functions of users n through 1 to derive the following comparative statics.

Proposition 1 (Reward Effect on Participation Level)**.** *In a sequential all-pay auction under incomplete information, without a reserve, a higher reward has no effect on the likelihood that user i submits a solution of positive quality. In comparison, with a positive reserve, a higher reward strictly increases the likelihood that user i submits a solution of positive quality.*

Proposition 1 indicates that we expect reward size to have a non-negative effect on user participation. Intuitively, a user's likelihood of participation *ex ante* depends on both the reward size and the highest quality submissions before hers. When the reward size increases, the highest quality among earlier submissions also increases. With a zero reserve and risk neutrality, these two effects cancel each other out and there will be no effect. In comparison, with a positive reserve, the reward effect on participation dominates the reward effect from the increase of the highest quality among earlier submissions, resulting in a strict increase in a user's likelihood of participation.

Note that a requester's satisfaction with the auction outcome depends more on the quality versus the quantity of submissions. This leads to our next proposition.

Proposition 2 (Reward Effect on Expected Submission Quality)**.** *In a sequential all-pay auction under incomplete information, a higher reward increases user i's expected submission quality.*

Proposition 2 indicates that we expect reward size to have a positive effect on the expected submission quality. In Appendix A, we present a two-player example (Example 1) with closed-form solutions for the quality and likelihood of submissions, as well as the average and highest quality.

We now examine the effect of a positive reserve on participation levels. The following proposition parallels the equivalent reserve price effect on participation in winner-pay auctions, where a positive reserve price excludes bidders with low values (Krishna 2009).

[6] This is a technical assumption to derive strict subgame perfect equilibria instead of ϵ-equilibria.

Proposition 3 (Reserve Effect on Participation Level)**.** *In a sequential all-pay auction under incomplete information, a higher reserve quality decreases the likelihood that a user submits a solution with positive quality.*

Intuitively, the higher the reserve quality, the less likely it is that a user with low ability will participate in the auction, since participation requires time and effort. In Appendix A, we present Example 2, a continuation of Example 1, to demonstrate the relevant comparative statics with respect to reserve quality.

As we do not have a general solution for the optimal reserve quality, we present a numerical example to illustrate the effects of reserve quality on the expected highest and average quality, respectively, in Appendix A.

4.2 Simultaneous All-pay Auctions under Incomplete Information

In this subsection, we investigate the case when all solutions are submitted with password protection. In this scenario, the competitive process is best approximated by a simultaneous all-pay auction, where users do not see others' solutions before submitting their own. The crowdsourcing process on TopCoder is an example of a simultaneous all-pay auction. We can thus derive comparative statics for simultaneous all-pay auctions under incomplete information to examine the effects of reward size and reserve quality.

Proposition 4 (Reward Effect on Participation Level)**.** *In a simultaneous all-pay auction under incomplete information, without a reserve, a higher reward has no effect on the likelihood that user i submits a solution of positive quality. In comparison, with a positive reserve, a higher reward strictly increases the likelihood that user i submits a solution of positive quality.*

Proposition 5 (Reward Effect on Expected Submission Quality)**.** *In a simultaneous all-pay auction under incomplete information, a higher reward increases the expected submission quality.*

Proposition 6 (Reserve Effect on Participation Level)**.** *In a simultaneous all-pay auction under incomplete information, a higher reserve quality decreases the likelihood that a user submits a solution with positive quality.*

Unlike the sequential auction, every user in a simultaneous all-pay auction is symmetric *ex ante*. In Appendix A, we present numerical examples to illustrate the effects of reserve quality on the expected quality for each player in a simultaneous all-pay auction.

In sum, we have separately characterized the reward and reserve effects on participation and submission quality under sequential and simultaneous all-pay auctions, respectively. We find that reward and reserve quality have similar effects on both participation levels and submission quality under each auction format.[7] While these characterizations provide benchmarks for our experimental design and hypotheses, in reality, most all-pay auctions on Taskcn are hybrid sequential/simultaneous auctions, where participants endogenously determine whether to password protect their solutions. Two other features of the field not captured by our theoretical models are endogenous entry timing and the choice among multiple auctions, each of which is modeled by Konrad and Leininger (2007) and DiPalantino and Vojnovic (2009), respectively.[8] A more realistic model which incorporates endogenous auction selection, endogenous entry and endogenous choice among multiple auctions is left for future work. Nonetheless, our experiment provides a useful framework with which to study the effect of reward level and reserve presence on both participation levels and submission quality.

5 Experimental Design

In this section, we outline our experimental design. We use a 2×3 factorial design to investigate the reward and reserve quality effects on user behavior on Taskcn. Specifically, we investigate whether tasks with a higher reward attract more submissions and generate solutions of a higher quality. We are also interested in determining whether an early high-quality solution which functions as a soft reserve will deter the entry of low quality solutions, especially if it is posted by a user with a history of winning.

5.1 Task Selection: Translation and Programming

In this study, we focus on translation and programming tasks for our field experiment, as such tasks are well defined, and the nature of the respective solutions is fairly standard and objective. Thus, our tasks are close to the expertise-based projects, where performance is driven primarily by level of expertise in the domain area and contestant effort, with little uncertainty in the outcome (Terwiesch and Xu 2008).

Our translation tasks fall into two categories: personal statements collected from Chinese graduate students at the University of Michigan and company introductions downloaded from Chinese websites. We choose these two categories as they are sufficiently challenging, each requiring a high level of language

[7] We are not aware of any systematic comparison of these two all-pay auction mechanisms under incomplete information. Under the assumption of no-reserve, Jian and Liu (2013) characterize the expected highest quality for the n-player sequential all-pay auctions and compare it with that in simultaneous all-pay auctions. When $n \leq 4$, they prove that the expected highest quality in simultaneous all-pay auctions is higher than that in sequential all-pay auctions.

[8] While the theoretical framework of Konrad and Leininger (2007) allows endogenous entry timing, it is under complete information, a feature which cannot be justified in Taskcn. In comparison, DiPalantino and Vojnovic (2009) only examine simultaneous auctions, whereas we are interested in the sequential feature in Taskcn contests. For these reasons, we choose not to adopt their frameworks.

skill and effort compared to other translation documents, such as resumes. In Appendix B, we provide an example of a personal statement and an example of a company introduction, as well as a complete list of Taskcn IDs and URLs for all the translation tasks used in our experiment.

For our programming tasks, we construct 28 different programming problems, including 14 Javascript and 14 Perl tasks. None of our programming tasks is searchable and each has a practical use. A complete list of the programming tasks is provided in Appendix B. One example of such a task reads: "Website needs a password security checking function. Show input characters as encoded dots when user types password. Generate an information bar to indicate the security level of the password, considering these factors: (1) length of the password; (2) mixture of numbers and characters; (3) mixture of upper and lower case letters; (4) mixture of other symbols. Please provide source code and html for testing." The functionality and thus quality of such programming tasks can be assessed by qualified programmers.

Table 2: Summary Statistics about Tasks on Taskcn from 2006 to March 27, 2009

	Reward (in CNY)			# of Submissions		
	Median	Mean	SD	Median	Mean	SD
Translation	100	137	164	42	109	163
Programming	100	176	378	6	10	17

To prepare for our field experiment, we crawled all the tasks on Taskcn posted from its inception in 2006 to March 27, 2009. Table 2 presents summary statistics (median, mean and standard deviation) for these two types of tasks. Note that, while translation and programming tasks have the same median reward on the site, the former generate a higher median number of submissions (possibly due to the ability to submit a machine-generated solution).

5.2 Treatments

Using the reward information provided in Table 2, we choose two reward levels for our tasks, 100 CNY and 300 CNY, based on the following considerations. First, using the median reward for our low reward treatments guarantees a certain amount of participation, whereas our high-reward level, 300 CNY, corresponds to the 90^{th} percentile of the posted tasks in these two categories. Second, the two reward levels have a monetarily salient difference and therefore allow us to test for differences across treatment levels.

As translation tasks have a relatively large number of submissions on Taskcn (Table 2), we investigate whether the early entry of a high quality submission influences participation levels, similar to the effect of a reserve price in an auction. Thus, for each reward level, we vary the reserve conditions, including

No-Reserve, Reserve-without-Credit, and Reserve-with-Credit.[9] The two reserve conditions differ only in whether the user posting the high quality solution has credits from previous wins. In the Reserve-without-Credit treatments, each early submission is posted by a user without a winning history on the site, whereas in the Reserve-with-Credit treatments, our submissions are posted by a user with four credits. To ensure the quality of the translations used in the reserve treatments, we ask a bilingual student (the owner of the personal statement when applicable) to provide the first round of English translations, and a native English speaker to provide a second round. To determine the quality of the reserve or any early submission, a user will need to read the translation.

Table 3: Number of Tasks by Experimental Treatment

	No-Reserve	Reserve-without-Credit	Reserve-with-Credit
Low-Reward	Programming (14)		
(100 CNY)	Translation (20)	Translation (20)	Translation (20)
High-Reward	Programming (14)		
(300 CNY)	Translation (20)	Translation (20)	Translation (20)

Table 3 summarizes our six treatments. The number in brackets indicates the number of distinct tasks posted in a treatment. A total of 120 translation (28 programming) tasks are randomly assigned to six (two) treatments. Thus the full 2×3 factorial design is applied to translation tasks, while programming tasks are used to check for the robustness of any reward effects. We use a greater number of translation tasks in the field experiment in part because of the relative difficulty in generating unique, plausible, and comparable programming tasks.

5.3 Experimental Procedure

Between June 3 and 22, 2009, we posted 148 tasks on Taskcn. We posted eight tasks per day (one translation and one programming task from each treatment) so as not to drastically increase the total number of tasks posted daily on the site.[10]

Each task was posted for seven days, with an indication that one winner would receive the entire reward. To avoid reputation effects from the requester side, we created a new user account for each task. After a task

[9] Recall that users earn 1 credit whenever they earn 100 CNY on the site. We created our own user account and obtained winning credits by winning tasks before the launch of our experiment.

[10] From January to March 2009, the average number of new tasks posted on the site per day is 12. Since each task is open between one week to a month, and all open tasks are listed together, users may select from among dozens to hundreds of tasks at any given time.

was posted, any user could participate and submit a solution within seven days. At the end of the seven-day period, we selected a winner for each task, excluding our reserve submissions.[11] We did not explicitly announce any tie-breaking rule for our tasks.

During our experiment, 949 users participated in the translation tasks, submitting a total of 3671 solutions, and 82 users participated in the programming tasks, submitting a total of 134 solutions. Table 4 presents the summary statistics of user credits among our participants.

Table 4: Summary Statistics for User Credits

	Mean	Median	Min	Max	Standard Deviation
Translation	0.43	0	0	96	4
Programming	4	0	0	62	11

In addition to the number of submissions, participants also vary in their password protection behavior between these two types of tasks. We find that 8% of the translation and 53% of the programming solutions are submitted with password protection. This difference in the proportion of password-protected submissions per task is statistically significant ($p < 0.01$, permutation test, two-sided).

5.4 Rating Procedure

To determine submission quality, we recruited raters from the graduate student population at the University of Michigan to evaluate each submission. These raters were blind to our research hypotheses. Our rating procedures follow standard practice in content analysis (Krippendorff 2003). To evaluate the translation submissions, we proceeded in two stages. First, we recruited three bilingual Chinese students to independently judge whether a submission was machine-translated. If two of them agreed that a submission was machine-translated, we categorized it as a machine translation. We then recruited nine bilingual Chinese students, whom we randomly assigned into three rating groups. For this stage, all valid translations plus one randomly-selected machine translation for each task were independently evaluated by three raters.[12] Raters for translation tasks each had scored above 600 on the TOEFL. To evaluate the programming submissions, we recruited three Chinese students, each with an undergraduate degree in computer science and several years of web programming experience. We conducted training and rating sessions for all our raters. Raters

[11] We find that the average quality of the winning solutions (4.33) is not significantly different from that of our reserve submissions (4.36), based on the evaluation of raters blind to the research design and hypotheses ($p = 0.40$, one-sided Wilcoxon signed-rank test).

[12] Note that the machine translations were not marked in the second stage. Thus, this procedure provides an additional consistency check for our raters.

within each rating group independently evaluated the same set of task-submission pairs. Details of the rating procedures and instructions can be found in Appendix C.

Table 5: Rating Task Quantities and Inter-rater Reliabilities (ICC[3,3])

	Group	# Tasks	# Submissions	Task Difficulty	Submission Quality
Translation	1	43	265	0.62	0.90
	2	35	215	0.88	0.88
	3	42	284	0.72	0.68
Programming	1	28	108	0.55	0.77

From October 2009 to February 2010, we conducted 45 rating sessions at the University of Michigan School of Information Laboratory. Each session lasted no more than two hours. Students were paid a flat fee of $15 per hour to compensate them for their time. We used intra-class correlation coefficients, ICC[3,3], to measure inter-rater reliability.

Table 5 presents the number of rating tasks and the inter-rater reliability for each rating group. The last two columns present the inter-rater reliability for each rating group. Good to excellent reliability is observed for all rating groups, thus increasing our confidence in our rater evaluations of solutions.[13] Additionally, machine translations are rated as having significantly lower quality than other valid translations in the second stage,[14] providing further evidence of rating consistency between the first- and second-stage raters. In our subsequent analysis, we use the median evaluation for the task difficulty and the overall submission quality.[15]

6 Results

Of the 120 translation and 28 programming tasks posted, we received at least one submission for every task. On average, each translation (programming) task received 1830 (1211) views, 46 (9) registrations and 31 (5) submissions. Although it might at first appear that participation is several times greater for translation tasks relative to programming tasks, most of the submissions we received for the translation tasks are machine-generated. The average number of valid translations per task (5) is equal to that of the

[13]In general, values above 0.75 represent excellent reliability, values between 0.40 and 0.75 represent fair to good reliability, and values below 0.40 represent poor reliability.

[14]On a 1-7 Likert scale, the average median quality of machine and valid translations is 2 and 5, respectively. Using the average median quality per task as one observation, we find that this quality difference is significant at the 1% level ($p < 0.01$, one-sided Wilcoxon signed-rank test).

[15]Task difficulty is measured by the median evaluation for questions 1(d) in translation and 1(b) in programming, whereas overall submission quality is measured by the median evaluation for questions 3 in translation and 2(d) in programming. See Appendix C for rating instructions.

solutions to programming tasks. Of the submissions received, 8% (53%) of the translation (programming) solutions are password protected, making them hybrid sequential/simultaneous all-pay auctions.

A total of 949 (82) unique users participated in our translation (programming) tasks.[16] We categorize the participants based on their prior winning experience. We define *experienced users* as those who have won at least 100 CNY (with at least one reputation credit) prior to our experiment, whereas we define *inexperienced users* as those who have not.[17] Table 6 reports the summary statistics of participants by credits won.[18] Specifically, we find that 4% (27%) of the participants in the translation (programming) tasks are experienced users.

Table 6: The Percentage of Each User Type in the Experiment

Task		Number of Users	Percentage	Median Credit	Mean Credit
Translation	Experienced Users	42	4	3	10
	Inexperienced Users	907	96	0	0
Programming	Experienced Users	22	27	5	10
	Inexperienced Users	60	73	0	0

We now present our results in two subsections. In Section 6.1, we present our main results related to our theoretical predictions and addressed directly by our experimental design. In Section 6.2, we present our secondary results.

6.1 Treatment Effects

Before analyzing our results, we first check that our randomization of tasks across treatments works. Pairwise Kolmogorov-Smirnov tests comparing task difficulty across treatments yield $p > 0.10$ for both translation and programming tasks, indicating that the level of task difficulty is comparable across different treat-

[16]We treat each unique ID as a unique user, as the reputation system on the site encourages users to keep a single identity across tasks.

[17]We have used two alternative definitions of experienced users: winning ratio, and a guru score (Nam, Ackerman and Adamic 2009). *Winning ratio* is defined by the number of tasks a user wins divided by the total number of tasks a user participates in on the site. The *guru score* is defined by $g_i = \frac{\sum_{j=1}^{m_i} b_{ij} - x_i}{x_i}$, where $x_i = \sum_{j=1}^{m_i} \frac{1}{n_j}$ represents the probability that user i's submissions are chosen as the winner for each task if a requester randomly selects one submission as the winner; $b_{ij} = 1$ if user i provides the best answer for task j and 0 otherwise; m_i is the number of tasks user i participates in; and n_j is the total number of submissions for task j. The guru score takes into account the number of other users submitting a solution to a task and indicates whether a user's performance is better or worse than chance. Using the winning ratio or guru score as alternative measures of user experience in Section 6.2, we find that Result 7 remains robust, whereas the weakly significant portion of Results 5 and 6 are no longer significant.

[18]These summary statistics are computed based on field data from Taskcn from 2006 through June 2, 2009, the day before our experiment.

ments. In what follows, we evaluate the specific treatment effects on participation levels and submission quality.

We first examine whether different reward levels affect participation. Specifically, we separately examine the effect of reward level on both the total number of translation submissions and the number of valid translations. To qualify for a valid translation, a submission must be neither machine-translated nor copied from previous submissions. Similarly, we separate programming submissions into valid and invalid solutions. Of the 134 programming submissions, we find that 26 are invalid due to either incompleteness or copying from previous submissions. In both types of tasks, valid solutions involve a certain amount of effort in the preparation process, while invalid ones involve minimum effort. In our separate analyses, we find no significant difference between the reserve-with-credit and reserve-without-credit treatments in their effect on either participation or valid submission quality (participation: $p > 0.1$; quality: $p > 0.1$, one-sided permutation tests). Therefore, in subsequent analyses, we pool these two treatments into a single reserve treatment.

We first examine the reward effect on participation levels. Based on Propositions 1 and 4, we expect that a task with a higher reward should receive more submissions. While participation is measured theoretically by the likelihood that a user submits a solution of positive quality, empirically, we measure participation by the number of submissions. Implicitly, we treat every submission, including machine translation, as one with positive quality.

Hypothesis 1 (Reward Effect on Participation)**.** *A task with a high reward attracts more submissions than a task with a low reward.*

Figure 1 presents the reward effect on participation in both the translation (top panel) and programming tasks (bottom panel). For each type of task, we present separate participation data for the group of all submissions and the group of only valid submissions. The average number of submissions and standard errors for the high- and low-reward treatments are presented in each graph. We summarize the results below.

Result 1 (Reward Effect on Participation)**.** *Translation (programming) tasks in the high-reward treatments receive significantly more submissions compared to those in the low-reward treatments.*

Support. *Table 7 presents the summary statistics and treatment effects for both the translation and programming tasks. Specifically, we find that the average number of translation submissions per task is significantly higher in the high-reward than in the low-reward treatments (no-reserve:* $p = 0.017$*; reserve:* $p < 0.01$*, one-sided permutation tests). Furthermore, this difference is (weakly) significant for the subset of valid translations (no-reserve:* $p = 0.094$*; reserve:* $p < 0.01$*, one-sided permutation tests). For programming tasks, one-sided permutation tests yield* $p = 0.037$ *for all submissions and* $p = 0.051$ *for valid submissions.*

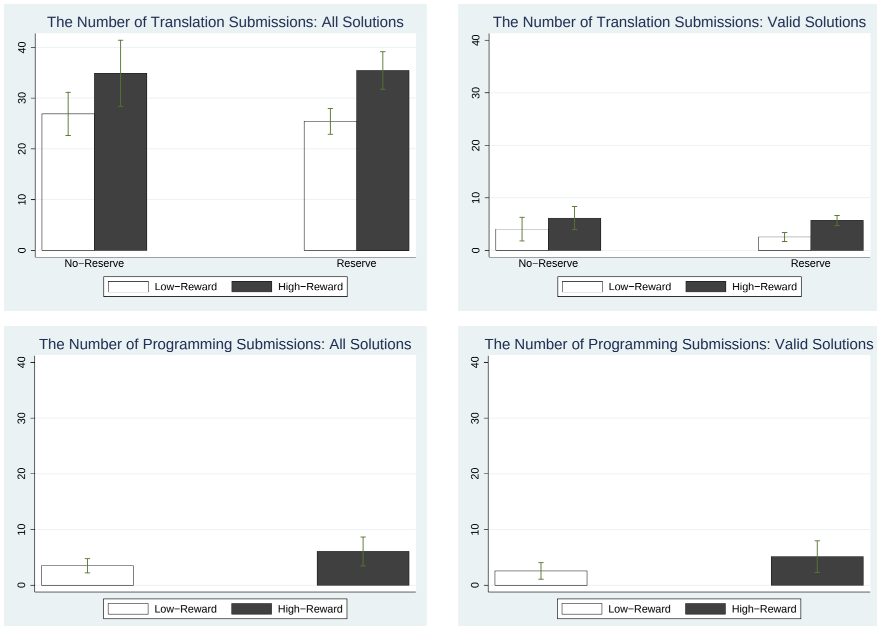

Figure 1: Reward Effect on Participation Level

Table 7: Treatment Effects on the Average Number of Submissions Per Task

All Solutions	Translation			Programming	
	No-Reserve	Reserve	*Reserve Effect*		All
High-Reward	35	35	$p = 0.445$	High-Reward	6
Low-Reward	27	25	$p = 0.263$	Low-Reward	4
Reward Effect	$p = 0.017$	$p = 0.000$		*Reward Effect*	$p = 0.037$
Valid Solutions	Translation			Programming	
	No-Reserve	Reserve	*Reserve Effect*		All
High-Reward	6	6	$p = 0.324$	High-Reward	5
Low-Reward	4	3	$p = 0.087$	Low-Reward	3
Reward Effect	$p = 0.094$	$p = 0.000$		*Reward Effect*	$p = 0.051$

By Result 1, we reject the null hypothesis in favor of Hypothesis 1. In other words, a higher reward induces more submissions. This result is consistent with our theoretical predictions in Propositions 1 and 4 only for the reserve case. In the absence of a reserve, both propositions predict that participation does not vary with reward size, which is not supported by our data. We note that the theoretical prediction relies on the risk neutral assumption, which is unlikely to be satisfied in the field. Furthermore, Result 1 is also consistent with other empirical findings on both the Taskcn (DiPalantino and Vojnovic 2009) and Topcoder sites (Archak 2010).

We now analyze the reserve effects on participation levels. Based on Propositions 3 and 6, we predict that an early high quality submission should decrease overall participation. Even though our reserve is not binding, we predict that users who cannot produce a translation with a higher quality will decline to participate. Thus, we expect less participation in the reserve treatments compared to the no-reserve treatments.

Hypothesis 2 (Reserve Effect on Participation)**.** *The number of submissions in the reserve treatments is lower than that in the no-reserve treatments.*

Summarizing all treatments, Table 8 reports three OLS regressions in a comparison of the relative effectiveness of the different treatments on participation levels for our translation tasks. The dependent variables are (1) the total number of solutions (2) the number of valid solutions and (3) the number of invalid solutions, respectively. Independent variables include the following (with omitted variables in parentheses): high-reward (low-reward), reserve (no-reserve), and task difficulty. In addition, we control for the task posting date in all three specifications. From Table 8, we see that the coefficient of the high-reward dummy is positive and significant at the 1% level in all three specifications, indicating a robust reward effect on participation when we control for other factors. Specifically, from low-reward to high-reward tasks, the average number of submissions increases by 10 for all solutions, 3 for valid solutions and 7 for invalid solutions. Furthermore, the coefficient of the reserve dummy is negative and significant in (2), indicating that a reserve submission deters the entry of other submissions for the subsample of valid entries. Finally, the coefficient for task difficulty is negative and significant, indicating that more difficult tasks receive fewer submissions. We summarize the reserve effect below.

Result 2 (Reserve Effect on Participation)**.** *While the overall number of submissions is not significantly different between the reserve and no-reserve treatments, the number of valid submissions is significantly lower in the reserve treatments, after controlling for task difficulty and posting date dummies.*

Support. *Column 4 in Table 7 reports the p-values for one-sided permutation tests for the effect of a reserve on participation for each treatment for both all solutions (upper panel) and the subset of valid solutions (lower panel). These results show that none of the effects is significant at the 10% level except for low-reward valid submissions (*$p = 0.087$*). In comparison, Table 8 reports the OLS regressions for participation.*

Table 8: OLS: Determinants of the Number of Submissions in Translation Tasks

Dependent Variable	# of Submissions (All)	# of Submissions (Valid)	# of Submissions (Invalid)
	(1)	(2)	(3)
High-Reward	9.700***	2.914***	6.785***
	(1.638)	(0.565)	(1.410)
Reserve	-1.380	-1.331**	-0.049
	(1.764)	(0.609)	(1.518)
Task Difficulty	-2.622***	-0.981***	-1.641**
	(0.954)	(0.329)	(0.821)
Constant	48.810***	13.340***	35.465***
	(6.049)	(2.088)	(5.208)
Observations	120	120	120
R^2	0.502	0.483	0.441

Notes: 1. Standard errors are in parentheses. 2. Significant at: * 10%; ** 5%; *** 1%.
3. Posting date dummies are controlled for.

In this set of regressions, the coefficient of the Reserve dummy is negative and significant only for the valid entry subsample (specification 2).

By Result 2, we reject the null hypothesis in favor of Hypothesis 2 for valid submissions.

In addition to participation, we are interested in what factors may affect submission quality. For submission quality, based on Propositions 2 and 5, we expect that a task with a higher reward will attract higher quality submissions.

Hypothesis 3 (Reward Effect on Submission Quality)**.** *A task with a high reward will attract submissions of higher quality than a task with a low reward.*

To investigate this hypothesis, we use two outcome measures to evaluate submission quality: the quality of all submissions and the quality of the best solution for each task. For tasks such as programming, only the quality of the best solution may matter. However, for modularizeable tasks such as translations, the requester might care about the average quality of the submitted solutions, as different translations may be combined at the sentence or paragraph level. Thus, we examine the reward effect on both the average submission quality and the highest submission quality.

Table 9 presents the results from six OLS specifications which investigate factors affecting submission

Table 9: OLS: Determinants of Submission Quality for Translation Tasks

Dependent Variable	All Translations	Valid Translation Submissions				Invalid Translations
	(1) Quality	(2) Quality	(3) Quality	(4) Best Quality	(5) Best Quality	(6) Quality
High Reward	0.126	0.328***	-0.028	0.289*	-0.0319	0.090
	(0.119)	(0.118)	(0.134)	(0.165)	(0.261)	(0.134)
Reserve	0.119	-0.619***	-0.609***	-0.530***	-0.509**	0.244*
	(0.124)	(0.112)	(0.132)	(0.155)	(0.202)	(0.139)
Task Difficulty	-0.118*	0.131***	0.130**	0.073	0.166*	-0.159**
	(0.062)	(0.049)	(0.060)	(0.098)	(0.098)	(0.071)
Invalid Submission	-2.932***					
	(0.105)					
Constant	5.245***	4.194***	3.106***	5.706***	1.235	2.510***
	(0.347)	(0.249)	(0.462)	(0.435)	(0.892)	(0.380)
User Fixed Effects	No	No	Yes	No	Yes	No
Observations	3,671	533	533	178	178	3,138
R^2	0.628	0.181	0.710	0.342	0.757	0.305

Notes: 1. Robust standard errors in parentheses are clustered at the task level in specifications (1), (2), (4) and (6).
2. Significant at: * 10%; ** 5%; *** 1%. 3. Posting date dummies are controlled for.

quality.[19] The dependent variables are the quality of all translation submissions (1), all valid translation submissions (2 and 3), the best translation submissions (4 and 5), and the invalid translation submissions (6). The independent variables include the following (with omitted variables in parentheses): high-reward (low reward), reserve (no-reserve), task difficulty and posting date dummies. In addition, specification (1) includes an invalid-submission dummy. For specifications (1), (2), (4) and (6), we report pooled models with standard errors clustered at the task level. We find that the coefficient of the high-reward dummy is positive and significant in (2), and weakly significantly in (4), indicating a significant (marginal) reward effect on the average (best) valid submission quality. Furthermore, the coefficient of the reserve dummy is negative and significant in both specifications, indicating a negative reserve effect on the quality of valid submissions. By contrast, it is positive and marginally significant in (6), indicating a positive reserve effect on the quality of invalid submissions, likely due to copying the high quality reserve solution. The coefficient of task difficulty is positive and significant in (2), but negative and significant in (6), suggesting that a valid (invalid) submission for a more difficult task is more (less) likely to receive a higher rating. Lastly, the coefficient of the invalid-submission dummy is negative and significant in (1), suggesting that, on average, the quality of an invalid submission is rated 3 points lower than that of a valid submission. We summarize

[19]Ordered probit specifications yield similar results and are available from the authors upon request.

these results below.

Result 3 (Reward Effect on Submission Quality)**.** *The average (best) quality of valid translation submissions is significantly (weakly) higher in the high-reward treatments than in the low-reward treatments.*

Support. *In Table 9, the high-reward dummy is positive in both specifications (2) and (4). It is significant at the 1% level in (2), and 10% level in (4).*

By Result 3, we reject the null hypothesis in favor of Hypothesis 3. That is, a task with a high reward attracts submissions of higher quality than a task with a low reward. In comparison, we find that, while programming tasks in the high-reward treatment attract higher average quality submissions than those in the low-reward treatment, this difference is not statistically significant (the average quality of valid solutions is 3.89 vs. 3.79, $p = 0.340$; the average quality of best solutions is 5.00 vs. 4.78, $p = 0.379$, using one-sided permutation tests).

Lastly, as we do not have analytical solutions for the optimal reserve, we are agnostic to the effect of a reserve on submission quality.

Hypothesis 4 (Reserve Effect on Submission Quality)**.** *The average submission quality will be different between the reserve and no-reserve treatments.*

Result 4 (Reserve Effect on Submission Quality)**.** *The quality of valid and best translation submissions is significantly lower in the reserve treatments than in the no-reserve treatments.*

Support. *In Table 9, the reserve dummy is negative and significant at the 1% level in both specifications (2) and (4).*

Result 4 indicates that the presence of a reserve has a negative and significant effect on submission quality. While a fully rational user should submit a solution only when its quality exceeds that of any previous submission, our participants do not always follow this rule. This result could come from the fact that the quality of the reserve submission is very high (at the far end of the quality distribution). As a result, experienced users might stay away from tasks with a reserve. If all experienced users drop out, the submission quality will decrease. We will explore the sorting explanation in Section 6.2.

In summary, we find significant treatment effects of both reward size and a reserve. We next investigate whether these effects are driven by within-user variations. That is, we explore whether a user submits a better solution to a task with a higher reward. Following the literature, we call this the *incentive effect*. Alternatively, our treatment effects might be driven by a *sorting effect* where tasks with a higher reward may attract better users.

To address the issue of an incentive effect, we examine whether within-user variation in submission quality exists. As 43% (38%) of the users who submit a valid (best) solution participate in more than one task, we use fixed effects models for specifications (3) and (5) in Table 9 to investigate whether the estimation in the pooled model is driven by within-user variation in the submission quality over tasks. Using the fixed effects model, we find no significant reward effect on submission quality within each user. However, our reserve dummy remains negative and significant, indicating that each user produces a submission of relatively lower quality for tasks with a reserve, compared to those without a reserve. In the next subsection, we investigate the sorting effects.

6.2 Sorting Effects

In this subsection, we investigate the extent to which Results 3 and 4 in our study are driven by user entry decisions. Even though we do not incorporate choice among multiple tasks in our theoretical model, for reasons of analytical tractability, a large literature in personnel and labor economics suggests that sorting is an important factor in improving worker performance. Specifically, Lazear (2000a, 2000b) examines the sorting effect when a fixed-payment mechanism is replaced by a pay-for-performance scheme, such as piece-rate or tournament. In his empirical study of a large auto glass company, he finds that, a pay-for-performance scheme increases worker effort (the incentive effect) and encourages the entry of high-ability workers (the sorting effect) (Lazear 2000b). Subsequent laboratory experiments report a similar sorting effect in pay-for-performance schemes (Cadsby, Song and Tapon 2007, Eriksson and Villeval 2008, Eriksson, Teyssier and Villeval 2009, Dohmen and Falk 2011). Finally, in a field experiment conducted on TopCoder, Boudreau and Lakhani (2012) find that when workers are endogenously sorted by skill level, they perform significantly better than do unsorted workers. Since the task reward structure on Taskcn might be considered a special form of pay-for-performance scheme, we expect sorting may also play a role in our experiment.

In comparison to Section 6.1, where we derive our hypotheses from our theoretical model, our hypotheses in this section are based on either empirical or theoretical prior findings. In what follows, we investigate the extent to which sorting may explain the results we obtain in our pooled model in Section 6.1.

Hypothesis 5 (Reward Effect on Entry)**.** *Tasks with a high reward are more likely to attract high-quality users.*

To test this hypothesis, we analyze user entry decisions by type, computed from two perspectives: (1) submission quality exhibited within our experiment and (2) their winning history on the site prior to the start of our experiment. We first investigate entry decisions using submission quality exhibited within our experiment. To do so, we construct a two-stage model.[20] In the first stage, we regress submission quality on our user dummies. Consequently, the estimated coefficient for user i, $\hat{\mu}_i$, approximates user submission quality compared to that of the omitted user. Note that this measure of user quality might be determined by

[20]We thank Jeff Smith for suggesting this approach.

various factors, such as user ability, effort, or reputation.[21] In our second stage, we construct a new statistic, $\bar{\hat{\mu}}_t = \frac{1}{n_t}\sum_{i=1}^{n_t}\hat{\mu}_t$, that represents the average user submission quality per task. We then regress $\bar{\hat{\mu}}_t$ on the reward size of each task, the reserve dummy, task difficulty and our posting date dummies.

Table 10: OLS: Determinants of User Quality in Translation Tasks

Dependent Variable	Average User Quality Among Valid Solutions	Average User Quality Among Best Solutions
	(1)	(2)
High Reward	0.741***	1.677**
	(0.225)	(0.684)
Reserve	-0.515**	-0.977
	(0.244)	(0.619)
Task Difficulty	-0.013	-0.302
	(0.138)	(0.494)
Constant	-2.073***	0.799
	(0.693)	(2.001)
Observations	112	103
R^2	0.273	0.231

Notes: 1.Robust standard errors are in parentheses.

2. Significant at: * 10%; ** 5%; *** 1%.

3. Posting date dummies are controlled for.

4. Of our 120 translation tasks, 8 did not receive any valid submissions, while the best solution of each of 17 tasks is either a reserve or invalid. These tasks are dropped from (1) and (2), respectively.

Table 10 reports the results from two OLS specifications investigating the determinants of average user submission quality among (1) valid and (2) best translation submissions. In specification (1), we find that the coefficient of the high-reward dummy is positive and significant, indicating that a high-reward task attracts higher-quality users. In comparison, the coefficient of the reserve dummy is negative and significant, indicating that the average user quality in a task with a reserve is lower. For our sample of best solutions (2), the coefficient of the high-reward dummy is positive and significant, indicating that, among those users who

[21]Note also that a high quality user is someone whose average submission quality within our experiment is high, whereas an experienced user is someone who has earned one or more credits prior to the start of our experiment. The average submission quality of experienced users is 5.21, whereas that of the inexperienced users is 4.91. The difference is significant ($p = 0.037$, one-sided).

provide the best solutions, average user quality is significantly higher for a high-reward task compared to that for a low-reward task. In comparison, the coefficient of the reserve dummy is negative but insignificant ($p = 0.118$, two-sided), suggesting that the presence of a reserve does not significantly impact submission quality for our group of best users.

Having analyzed individual entry decisions based on user quality exhibited within our experiment, we now investigate entry decisions using each user's winning history prior to the start of our experiment. To do so, we first compute the median user credit per task for our sample. Considering all valid solutions for a task, we find that the average median user credit is higher in the high-reward treatment than that in the low-reward treatment. This difference is weakly significant in the no-reserve treatments.

Result 5 (Reward Effect on Entry)**.** *Average user quality among the groups of valid and best translations is significantly higher in the high-reward than in the low-reward treatments. Furthermore, the average median user credit is weakly higher in the high-reward-no-reserve than in the low-reward-no-reserve treatment.*

Support. *Table 10 reports the results from two OLS specifications investigating the determinants of average user submission quality in translation tasks. The coefficient for the high-reward dummy is positive and significant in both specifications. Using user credit prior to our experiment, we find that, in the no-reserve treatments, the average median user credit is 0.45 in the high-reward treatment, and 0.05 in the low-reward treatment. This difference is weakly significant ($p = 0.055$, one-sided permutation test). In comparison, for the reserve treatments, we find the same relationship but at an insignificant level (0.14 vs. 0.09, $p = 0.369$, one-sided permutation test).*

By Result 5, we reject the null in favor of Hypothesis 5. That is, translation tasks with a high reward are more likely to attract high-quality users. In comparison, programming tasks with a high reward also attract high-quality users, but at an insignificant level (valid solutions: 2.09 vs. 1.34, $p = 0.196$, one-sided permutation test). This latter result may be due to the smaller number of observations for our programming tasks.

Using similar analysis, we now summarize the reserve effects on user entry decisions, using user submission quality (Table 10) as well as user credits accumulated prior to our experiment. Using user credit history, we find that, among all valid solutions for a high-reward task, the average median user credit is weakly lower in our reserve treatment.

Hypothesis 6 (Reserve Effect on Entry)**.** *Tasks with a reserve are more likely to deter high-quality users.*

Result 6 (Reserve Effect on Entry)**.** *The average user quality among valid translations is significantly lower in the reserve than in the no-reserve treatments. Furthermore, the average median user credit is weakly lower in the reserve-high-reward than in the no-reserve-high-reward treatment.*

Support. *Table 10 reports the results of two OLS specifications investigating the determinants of user submission quality in translation tasks. The coefficient for the reserve dummy is negative and significant for specification (1). Using user credit prior to our experiment, we find that, in the high-reward treatment, the average median user credit is 0.14 in the reserve treatment and 0.45 in the no-reserve treatment. This difference is weakly significant ($p = 0.073$, one-sided permutation test). In comparison, for the low-reward treatments, the difference between the reserve and no-reserve treatments is not significant (0.05 vs. 0.09, p=0.545, one-sided permutation test).*

By Result 6, we reject the null in favor of Hypothesis 6. Overall, Result 6 indicates that an early high quality translation is more likely to deter other high-quality (experienced) users rather than low-quality (inexperienced) users. This differential entry response in the presence of a high quality reserve partially explains our finding that the reserve has a negative effect on subsequent submission quality (Result 4).

Lastly, following the theoretical predictions regarding entry timing in sequential all-pay auctions in Konrad and Leininger (2007), we investigate what factors may influence submission time in our study. In a previous study, Yang, Adamic and Ackerman (2008b) find a positive correlation between reward size and later submission on Taskcn. As reward level is endogenously determined in their naturally occurring field data, but exogenously determined in our experiment, we are able to separate the effects of reward size and task difficulty on submission timing.

Hypothesis 7 (Submission Timing)**.** *Experienced users will submit their solutions later than inexperienced ones.*

In Table 11, we report the results of four OLS specifications to investigate factors affecting the submission time for all translation submissions (specifications 1 and 2) as well as only those that are valid (specifications 3 and 4). To replicate the results from Yang et al. (2008b), specifications (1) and (3) include the high-reward dummy as our only independent variable. In comparison, specifications (2) and (4) include the following additional independent variables (with omitted variables in parentheses): reserve (no reserve), task difficulty, experienced users (inexperienced users) and solution protection (no protection). Our findings indicate that, when other variables are not controlled for, a high reward has a positive and significant effect on submission time. This result is consistent with the finding in Yang et al. (2008b). However, after controlling for task difficulty and user experience, this finding becomes insignificant for valid solutions, which indicates that the reward effect on submission timing for valid solutions can be decomposed into two effects. First, experienced users wait to submit solutions for high reward tasks, possibly due to strategic reasons or more effort. Second, more difficult tasks require more time to complete. We summarize these results below.

Result 7 (Submission Time)**.** *For the sample of valid translation submissions, experienced users submit their translations significantly later than do inexperienced ones, when we control for task difficulty.*

Table 11: Determinants of Submission Time for Translation Tasks

Dependent Variable	Submission Time (All)		Submission Time (Valid)	
	(1)	(2)	(3)	(4)
High-Reward	0.211***	0.138***	0.371*	0.242
	(0.039)	(0.043)	(0.188)	(0.195)
Valid Translation		1.237***		
		(0.107)		
Reserve		-0.031		-0.041
		(0.045)		(0.199)
Task Difficulty		0.020		0.205**
		(0.027)		(0.096)
Experienced User		0.113		0.724**
		(0.136)		(0.284)
Protected Solution		-0.097		-0.067
		(0.142)		(0.335)
Constant	0.567***	0.252*	1.423***	0.486
	(0.084)	(0.147)	(0.307)	(0.529)
Observations	3,515	3,515	485	485
R^2	0.014	0.095	0.054	0.078

Notes:

1. Standard errors in parentheses are clustered at the task level.

2. Significant at: * 10%; ** 5%; *** 1%. 3. Posting date dummies are controlled for.

4. Data on submission time were retrieved after the experiment. By then, Taskcn had deleted 156 of our submission pages, 48 of which were pages for valid solutions.

Support. *In specification (4) of Table 11, the coefficient of the experienced user dummy is positive and significant at the 5% level, indicating that experienced users submit their solutions later than do inexperienced ones. On average, experienced users submit their solutions 0.724 days later than inexperienced ones do.*

By Result 7, we reject the null in favor of Hypothesis 7. We further find that, among all solutions, high-reward task solutions are submitted 0.138 days later. Furthermore, a valid translation is submitted 1.237 days later than a machine-translation. Restricting our analysis to only valid submissions, we find that translations for a high-reward task are still submitted marginally significantly later than those for a low-reward task. However, after controlling for task difficulty, we find that experienced users submit their solutions 0.724 days later than inexperienced users, while the reward effect on submission time is no longer significant. Furthermore, the task difficulty coefficient is positive and significant, indicating that users take 0.205 days longer to submit a valid solution for each additional level of difficulty (on a 1-7 Likert scale).

In summary, we find significant reward effects on both participation levels and submission quality, suggesting that a monetary incentive is effective in inducing more submissions and better solutions, both of which are consistent with the predictions of our model. While our model does not incorporate choice among multiple tasks, we find significant sorting effects among experienced users. Specifically, a higher reward also attracts higher quality (more experienced) users. Furthermore, while the early entry of a high quality solution does not significantly affect the number of submissions in contrast to our model's prediction of a reduction in quantity, we find that solution quality dramatically decreases with the presence of a reserve, as it deters the entry of high quality (experienced) users. The latter is again a consequence of sorting, which is not incorporated into our model. Lastly, in addition to their entry decisions, experienced users also submit their solutions later than inexperienced users do, controlling for task difficulty. While entry timing is exogenous in our model, the late entry of experienced users is predicted in a model of endogenous timing (Konrad and Leininger 2007).

7 Discussion

Crowdsourcing continues to be an important problem-solving tool, utilized by individuals, non-profit and for-profit organizations alike. Consequently, evaluating the behavioral responses of various design features will help improve the performance of crowdsourcing institutions and thus increase user satisfaction. In this study, we examine the effect of different design features of a crowdsourcing site on participation levels, submission quality and user entry decisions. Conducting a field experiment on Taskcn, we find that a higher reward induces both greater participation and higher submission quality. Controlling for the existence of a reserve in the form of a high quality early submission, we find that a reserve lowers subsequent submission quality, as it preferentially deters the entry of experienced users. Experienced users also distinguish themselves from inexperienced ones by being more likely to select higher reward tasks over lower reward ones, and by submitting their solutions relatively later.

Through our field experiment, we are able to observe interesting patterns that likely would not have emerged had the experiment been conducted in a lab setting. Perhaps the most surprising finding of our experiment is that the entry decisions of high quality (experienced) users drive the reward and reserve effects on overall submission quality. Specifically, we find that a higher reward attracts more experienced users, while a high quality reserve deters them. The first finding is consistent with the sorting effect found in the labor economics literature, that a higher reward attracts better workers. However, our finding on the selection effect of a high quality reserve submission is new to this body of literature.

Our findings not only help to inform the design of crowdsourcing institutions, but also provides useful feedback to contest theory. While most existing theoretical models of all-pay auctions ignore entry decisions, a model with endogenous entry (DiPalantino and Vojnovic 2009) treats every user as fully rational, which cannot explain our reserve effects on submission quality.[22] Our results suggest that a more accurate theory for predicting behavior in the field should incorporate behavior of both naïve and sophisticated types. Naïve users submit low-cost computer-generated solutions irrespective of a reserve, while sophisticated users are more likely to choose tasks with a higher probability of winning, i.e., those without a high-quality reserve.[23] Lastly, Taskcn provides an example that the auction format is endogenously determined by user password protection behavior, ranging from a sequential (no password protection) to a simultaneous all-pay auction (one hundred percent password protection), with hybrid sequential/simultaneous in the middle. To our knowledge, this has not been modeled theoretically.

Future research could expand on our findings by studying the effect of password protection on participation level and submission quality.[24] Our finding that early high-quality submissions tend to deter subsequent high-quality submissions suggests that it may be desirable to have submissions password protected and to hide user experience level or identity.

References

Amann, Erwin and Wolfgang Leininger, "Asymmetric All-Pay Auctions with Incomplete Information: The Two-Player Case," *Games and Economic Behavior*, 1996, *14*, 1–18.

Anderson, Simon P., Jacob K. Goeree, and Charles A. Holt, "Rent Seeking with Bounded Rationality: An Analysis of the All-Pay Auction," *Journal of Political Economy*, 1998, *106* (4), 828–853.

Archak, Nikolay, "Money, Glory and Cheap Talk: Analyzing Strategic Behavior of Contestants in

[22]Morgan, Orzen and Sefton (2010) presents a theoretical model with endogenous participation in the Tullock contest.

[23]We thank an anonymous referee for this suggestion.

[24]We thank an anonymous referee for these suggestions.

Simultaneous Crowdsourcing Contests on TopCoder.com," in "Proceedings of the 19th international conference on World Wide Web" Raleigh, North Carolina 2010.

Baye, Michael R., Dan Kovenock, and Casper G. de Vries, "The All-pay Auction with Complete Information," *Economic Theory*, 1996, *8*, 291–305.

Bertoletti, Paolo, "On the reserve price in all-pay auctions with complete information and lobbying games," 2010. Manuscript.

Boudreau, Kevin J. and Karim R. Lakhani, "High Incentives, Sorting on Skills-or Just a Taste for Competition? Field Experimental Evidence from an Algorithm Design Contest," 2012.

Boudreau, Kevin J, Nicola Lacetera, and Karim R. Lakhani, "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis," *Management Science*, May 2011, *57* (5), 843–863.

Cadsby, Bram, Fei Song, and Francis Tapon, "Sorting and Incentive Effects of Pay-for-Performance: An Experimental Investigation," *The Academy of Management Journal*, 2007, *50* (2), 387–405.

Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm, "Tournaments and Office Politics: Evidence from a Real Effort Experiment," *American Economic Review*, 2010, *100* (1), 504–17.

Chen, Yan, Teck-Hua Ho, and Yong-Mi Kim, "Knowledge Market Design: A Field Experiment at Google Answers," *Journal of Public Economic Theory*, 2010, *12* (4), 641–664.

Dasgupta, Partha, "The Theory of Technological Competition," in J.E. Stiglitz and F. Mathewson, eds., *New Developments in the Analysis of Market Structures*, London: Macmillan, 1986, pp. 519–548.

Davis, Douglas D. and Robert J. Reilly, "Do Too Many Cooks Spoil the Stew? An Experimental Analysis of Rent-seeking and the Role of a Strategic Buyer," *Public Choice*, 1998, *95*, 89–115.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M. Sheremeta, "A Survey of Experimental Research on Contests, All-Pay Auctions and Tournaments," 2012. Chapman University Working Paper.

DiPalantino, Dominic and Milan Vojnovic, "Crowdsourcing and All-Pay Auctions," in "Proceedings of the 10th ACM conference on Electronic commerce" 2009.

Dohmen, Thomas and Armin Falk, "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender," *Amer. Econ. Rev.*, 2011, *101* (2), 556–590.

Ehrenberg, Ronald G. and Michael L. Bognanno, "Do Tournaments Have Incentive Effects?," *Journal of Political Economy*, 1990, *98* (6), pp. 1307–1324.

Eriksson, Tor and Marie-Claire Villeval, "Performance-Pay, Sorting and Social Motivation," *Journal of Economic Behavior & Organization*, 2008, *68* (2), 412–421.

__ , Sabrina Teyssier, and Marie-Claire Villeval, "Does Self-Selection Improve the Efficiency of Tournaments?," *Economic Inquiry*, 2009, *47* (3), 530–548.

Fibich, Gadi, Arieh Gavious, and Aner Sela, "All-pay Auctions with Risk-averse Players," *International Journal of Game Theory*, 2006, *34* (4), 583–599.

Gneezy, Uri and Rann Smorodinsky, "All-pay Auction: An Experimental Study," *Journal of Economic Behavior and Organization*, 2006, *61*, 255–275.

Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature*, December 2004, *42* (4), 1009–1055.

Hillman, Arye and John Riley, "Politically Contestable Rents and Transfers," *Economics and Politics*, 1989, *1*, 17–40.

Howe, Jeff, "The Rise of Crowdsourcing," *Wired*, 2006, p. 4.

__ , "Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business," *New York: Crown Business*, 2008.

Jeppesen, Lars Bo and Karim R. Lakhani, "Marginality and Problem Solving Effectiveness in Broadcast Search," *Organization Science*, 2010, *21* (5), 1016–1033.

Jian, Lian and Tracy Xiao Liu, "Competing Openly or Blindly in Crowdsourcing Contests," 2013. Manuscript.

Kleeman, F., G.G. Voss, and K. Rieder, "Un(der)paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing," *Science, Technology & Innovation Studies*, 2008, *4* (1), 5–26.

Konrad, Kai A., *Strategy and Dynamics in Contests*, New York, NY: Oxford University Press, 2009.

__ **and Wolfgang Leininger**, "The Generalized Stackelberg Equilibrium of the All-pay Auction with Complete Information," *Review of Economic Design*, 2007, *11* (2), 165–174.

Krippendorff, Klaus, *Content analysis: An introduction to its methodology*, 2nd ed., Thousand Oaks, CA: Sage Publications, 2003.

Krishna, Vijay, *Auction Theory*, 2nd edition ed., Waltham, Massachusetts: Academic Press, 2009.

__ **and John Morgan**, "An Analysis of the War of Attrition and the All-Pay Auction," *Journal of Economic Theory*, 1997, *72* (2), 343–362.

Lazear, Edward P., "Performance Pay and Productivity," *Amer. Econ. Rev.*, 2000, *90* (5), 1346–1361.

__ , "The Power of Incentives," *Amer. Econ. Rev.*, 2000, *90* (2), 410–414.

__ **and Sherwin Rosen**, "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, 1981, *89* (5), 841–864.

Liu, Tracy X., "Sequential vs. Simultaneous All-pay Auction: An Experimental Study," 2011. University of Michigan, Manuscript.

Liu, Tracy Xiao, Jiang Yang, Lada A. Adamic, and Yan Chen, "Crowdsourcing with all-pay auctions: A field experiment on Taskcn," *Proceedings of the American Society for Information Science and Technology*, 2011, *48* (1), 1–4.

Lugovskyy, Volodymyr, Daniela Puzzello, and Steven Tucker, "An Experimental Investigation of Overdissipation in the All Pay Auction," *European Economy Review*, 2010, *54* (8), 974–997.

Mason, Winter and Duncan J. Watts, "Financial Incentives and the 'Performance of Crowds'," in "KDD-HCOMP. 2009" Paris, France 2009.

Morgan, John, Henrik Orzen, and Martin Sefton, "Endogenous entry in contests," *Economic Theory*, 2010, pp. 1–29.

Nalebuff, Barry J. and Joseph E. Stiglitz, "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *The Bell Journal of Economics*, 1983, *14* (1), pp. 21–43.

Nam, Kevin K., Mark S. Ackerman, and Lada A. Adamic, "Questions in, Knowledge iN? A Study of Naver's Question Answering Community," in "CHI'2009: Proceeding of the twenty-seventh annual SIGCHI conference on Human factors in computing systems" 2009. Boston, MA.

Noussair, Charles and Jonathon Silver, "Behavior in All-Pay Auctions Under Incomplete Information," *Games and Economic Behavior*, 2006, *55* (1), 189–206.

Potters, Jan, Casper G. de Vries, and Frans van Winden, "An Experimental Examination of Rational Rent-Seeking," *European Journal of Political Economy*, 1998, *14*, 783–800.

Segev, Ella and Aner Sela, "Multi-Stage Sequential All-Pay Auctions," 2012. Manuscript.

Terwiesch, Christian and Yi Xu, "Innovation Contests, Open Innovation, and Multiagent Problem Solving," *Management Science*, 2008, *54* (9), 1529–1543.

Tullock, Gordon, *Efficient Rent Seeking*, College Station, Tx: Texas A&M University Press,

Yang, Jiang, Lada A. Adamic, and Mark S. Ackerman, "Competing to Share Expertise: the Taskcn Knowledge Sharing Community," in "Proceedings of International Conference on Weblogs and Social Media" Seattle 2008.

__ , __ , and __ , "Crowdsourcing and Knowledge Sharing: Strategic User Behavior on Taskcn," in "Proceedings of the 8th ACM Conference on Electronic Commerce" Chicago 2008.