Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange

Paulo Goes, Chenhui Guo, and Mingfeng Lin

{pgoes, chguo, mingfeng}@email.arizona.edu
Department of Management Information Systems
Eller College of Management
University of Arizona

Abstract

UGC (User-generated content) websites routinely deploy incentive hierarchies, where users achieve increasingly higher status in the community after achieving increasingly more difficult goals, to motivate users to contribute. Yet the existing empirical literature remains largely unclear whether such hierarchies are indeed effective in inducing user contributions. We gathered data from a large online crowd-based knowledge exchange to answer this question, and drew on the goal setting theory to study users' contributions before and after they reach consecutive levels of a vertical incentive hierarchy. We found evidence that even though these "glory"-based incentives may motivate users to contribute more before the goals are reached, user contribution levels dropped significantly after that. In other words, the cumulative effect appears only temporary. Our results hence highlight some unintended and heretofore undocumented effects of incentive hierarchies, and have important implications for business models that rely on user contributions, such as knowledge exchange and crowdsourcing, as well as the broader phenomenon of "gamification" in other contexts.

Key Words: online knowledge exchange; motivation; incentive hierarchy; goal setting; prospect theory

Version: March 22, 2014

1. Introduction

From ecommerce websites to online communities, many websites today depend heavily on user-generated contents (henceforth UGC); they regularly rely on voluntary contributions from a fluid membership to generate sufficient contents, and these contents serve to attract new members and retain old ones. The reach of such sites is only limited by imagination, ranging from wikis, blogs, social networking sites, consumer review sites, online games, to question-and-answer platforms. Regardless of the nature of contents, a fundamental issue for such websites is virtually the same: How do we motivate these users to continue participating in the site, and voluntarily contribute new content?

This issue is challenging due to the fundamental public goods problem inherent in user-generated content: one need not contribute their effort to enjoy the benefit of these UGC sites. A classic example is product reviews (Chen et al. 2010). From the perspective of a review writer, he or she expends the costs (time and efforts) but may not fully realize the benefits of providing the reviews. The "public goods" nature of user-generated contents will, as economic theories predict, lead to an *undersupply* of such contents. While the long memory of the "Internet" can capture and accumulate some whimsical moments of kindness—and thereby allowing more people to benefit from someone else’s efforts—the provision of timely information or contents can be highly challenging to sustain the growth of UGC sites.

Naturally, the key to resolve this "public goods" problem is to increase the rewards for contributing users so that they can better "internalize" some benefits of their contributions. Several aspects of such internalization can emerge endogenously. In the context of open-source software development, for example, Roberts et al. (2006) showed that even if users are not explicitly compensated for their efforts on one site, they may still be able to build up their profile or reputation to improve their offline employment prospects. Furthermore, Hann, Roberts and Slaughter (2013) found that merit-based ranking in online communities is associated with increase in offline earnings, especially for certain types of offline employment. Unfortunately, however, such endogenous internalization may not be available for all user-generated content sites, and it may only be useful for some extremely well-known members of the site. More importantly, from the perspective of site managers, this is beyond their control.

For these reasons, more and more UGC sites adopt "incentive hierarchies," where users can accumulate points for their voluntary contributions, and be awarded various accolades as their cumulative points pass increasingly higher thresholds. These mechanisms seek to help users internalize at least some benefits of providing "public" goods by recognizing them in front of their peers on the site, and by bestowing certain glory, honor, or bragging rights. Similar systems can be found in a variety of contexts (e.g. Hann et al. 2013).

However, despite its popularity in practice, there has been little systematic empirical research on whether such incentive hierarchies actually induce user contributions. The goal of this paper is to empirically investigate this question in the context of an online knowledge exchange, where users can ask technology-related questions and seek answers from their peer members. The asking members (henceforth "*askers*") can choose one or more answers as the solution, and those who provide the answers (henceforth "*answerers*") receive award points based on that. Once the answerer accumulates a sufficient number of points, their designation on the site will be elevated from “regular member” to “master” or other titles. We investigate whether this "meritocracy" type of incentive hierarchy actually induces users to more voluntarily contribute to the community, in the form of answering more questions posted by peers.

Since each level of the incentive hierarchy are intended as "goals" for users, we draw on the goal-setting theories in management to examine its effects on user contributions. In particular, goal-setting theories predict that user behaviors will be different *before* and *after* they reach each goal. They are likely to exert more effort before they reach goals than after (Heath et al. 1999). We therefore examine the effect of the goals before and after the user reaches them in our context. Specifically, the research questions that we seek to address in this paper are:

1. *How do users’ contribution levels change* before *they reach goals? And*
2. *How do users’ contribution levels change* after *they reach goals?*

We conduct a large number of empirical tests to examine the effect of the goal, or distance from the goal threshold, on user contributions. We track a random sample of users over time and use panel-data methods to study this effect using a regression-discontinuity design and a distance-based model. We then complement this by exploiting an exogenous change on the site, a time at which the incentive hierarchy was originally introduced, and examine how this exogenously imposed hierarchical rankings (which led to exogenous variations in each user's distance to goals) affected user behavior. We find evidence that even though hierarchies may motivate users to contribute more *before* the goals are reached, their effort levels drop significantly *after* that. Hence, the positive effect of goals in the incentive hierarchies appears only temporary.

Our study is one of the first to apply goal-setting theories to study the effect of incentive hierarchies in a user-generated content site, and to document several surprising findings. Our findings challenge the common wisdom that incentive hierarchies are effective in inducing user contributions; rather, after goals are reached, users exhibit signs of complacency and are not motivated to make further progress. These findings have important implications not only for crowd-based knowledge sharing or question-and-answer websites, but also for other situations where similar incentive hierarchies are used, such as the recent phenomenon of "gamification" in a wide range of contexts. Our study also contributes to the empirical literature on goal-setting by examining the effects of several consecutive goals, rather than just one goal, as has been the case in prior literature.

The rest of this paper is organized as follows. Section 2 reviews related literature to derive hypotheses. Section 3 provides details on the research context, and in Section 4 we describe our data and model specifications. Empirical results, robustness checks and alternative specifications are presented in Section 5. In Section 6, we discuss the implications and limitations of this study and conclude the paper.

2. Theoretical Background and Hypotheses

2.1. Literature Review

This section reviews related literature for our study. We begin by reviewing two streams of literature, i.e. goal pursuit and prospect theories of goal behaviors, as they directly inform our empirical tests. We then discuss some recent, ongoing research work by other researchers on related topics, and how our study differentiates from theirs.

2.1.1. Goal Pursuit Literature

Psychologists have been investigating goal pursuit behaviors for more than half a century (e.g. Hull 1932, Miller 1944, Earley et al. 1989, Louro et al. 2007, Koo and Fishbach 2008, Liberman and Förster 2008, Bonezzi et al. 2011). One of the earliest empirical studies in this extensive literature (Hull 1932) reported experimental findings that rats ran faster when they were closer to the food, providing one of the earliest evidence that distance to goals affects effort levels. Such goal pursuit behaviors readily extend to human beings, since we are able to anticipate consequences of goals and exert effort to achieve goals (Lewin et al. 1944). Interestingly, many later studies in this literature focus on the presence or absence of goals, rather than the distance from them. For example, Locke's goal-setting theory (1967, 1968) suggested that individuals achieve higher performance in the presence of goals (for a review, see Locke and Latham 2002). Empirical evidence obtained from a wide range of contexts such as negotiation, driving, logging, and reading consistently showed that specific goals lead to higher performance, even if the goals are challenging (Tubbs 1986, Mento et al. 1987, Locke and Latham 2002). Further, researchers verified that the "specific, challenging goals" has a stronger effect on performance than the "do-your-best" setting, while goal difficulty has a positive influence on level of effort and performance (Tubbs 1986, Mento et al. 1987, Locke and Latham 1990).

While the above literature yielded many interesting insights, it is the original goal-gradient hypothesis (Hull 1932) that bears more relevance to our research questions, since we are interested in the distance from goals. Researchers have empirically examined the effect of goal proximity in several contexts such as loyalty reward programs and sales tasks (Kivetz et al. 2006, Cheema and Bagchi 2011, Zhang and Huang 2010, Huang and Zhang 2011). However, perhaps with the exception of Bonezzi et al. (2011), few studies considered individual motivations *after* goal attainment. We explicitly address this question using data from an online knowledge exchange.

2.1.2. Prospect Theory of Goal Behavior

Rooted in prospect theory in economic literature (Kahneman and Tversky 1979, Tversky and Kahneman 1992), the "reference point framework" provides another explanation for goal pursuit behavior. Heath et al. (1999) built a prospect theory model of goal pursuit. They demonstrated that when there exists a reference point (such as an assigned goal) that separates gains from losses, an individual tends to distort the value of an outcome psychologically (Heath et al. 1999). Figure 1 shows that the value function in the prospect theory is S-shaped and nonlinear near the reference point. The curve is steeper in the loss region than it is in the gain region. The slope is maximized at the reference point, for it is increasing in the loss region and decreasing in the gain region.

[Insert Figure 1 about here]

The reference point framework of goal behavior illustrates that motivation is positively correlated with marginal value of progress (i.e. slope of value function). Heath et al. (1999) explained that individual has different emotions toward success (gain) and failure (loss), and is likely to exhibit loss aversion. In psychology, goals can serve as categorical cutoff points that separate regions into different levels. As a result, an individual will make counterfactual comparisons near these cutoffs (Medvec and Savitsky 1997). The perception that "what might have been if I reached the cutoff point?" will affect an individual's response to actual performance. Individuals who have just reached one level are more satisfied than those who are in the same level but just miss the next level.[1] Following that logic, we infer that individuals who are close to achieving the goal (before goal attainment) perceive a greater marginal value of progress, thus exerting more effort. In contrast, individuals who have surpassed the goal (after goal attainment) perceive a lower marginal value of progress, therefore exerting less effort.

[1] For instance, students who get 90.0 or above can get "A" in a course, while students who get from 80.0 to 89.9 can get "B." Paradoxically, students with 89 are less satisfied than those with 81, although eventually both groups of students get "B" in the course. Similarly, Medvec et al. (1995) reported the phenomenon that Olympic athletes who win silver medals were less satisfied than those who win bronze medals.

“Goals” have long been an implicit component of the prospect theory. In fact, behaviors in this literature often occur due to the uncertainty associated with prospect of attaining goals. The original study of prospect theory focused on decision making when comparing risky financial portfolios or gambles (Kahneman and Tversky 1979), and suggested that goals may change an individual’s perception of success likelihood. Larrick et al. (2009) demonstrated that specific goals induce risk-taking behavior in negotiation and decision-making tasks. Based on the same framework, Wu et al. (2008) highlighted that “moderately” difficult goals encourage individuals to “exceed the goal, but only by a very little,” whereas extremely difficult goals hurt performance, especially for low-ability individuals.

More broadly, the theory of reference point and loss aversion has been applied to many other situations such as competition, motivation and performance (Berger and Pope 2011, Pope and Schweitzer 2011, Pope and Simonsohn 2011, Koop and Johnson 2012). In economics, the theory of reference-dependent preference implies that individuals would behave differently from what neoclassical economic theory predicts (Bateman et al. 1997). In particular, with explicit reference points (e.g. income target, sales goal), individuals tend to exhibit loss aversion in consumption (Bateman et al. 1997), trading (Genesove and Mayer 2001), and labor supply (Farber 2008, Crawford and Meng 2011). Our study builds on these prior studies to examine the effect of goals as a reference point in the context of UGC sites.

2.1.3. Incentive Programs on UGC sites

We now briefly review the small but growing literature on user contributions in response to website designated goals. To the best of our knowledge, only two studies presented in recent conferences are related to the topic that we study, and they both use data from StackOverflow.com, a free, online question-and-answer website (different from our context) where users earn various badges on different aspects of their activities. Li et al. (2012) identified a short-term positive effect of winning new badges, i.e. users contribute more if they obtained new badges in the previous period. Another study (Anderson et al. 2013) argued that users increase efforts to obtain badges, and plotted user activities before and after that as empirical evidence. Our goal in this paper is to contribute to this emerging area by drawing on a new theoretical perspective that is highly relevant to this new empirical context, and providing more comprehensive and systematic empirical evidence. Our study is also based on a different website than theirs. More importantly, whereas these papers study a horizontally differentiated set of goals (i.e. different badges for different aspects of website activities), we study the effect of consecutive goals along a unified, vertical incentive hierarchy.

2.2. Hypotheses Development

Drawing on the theoretical literature reviewed in the previous subsection, we study the effect of incentive hierarchy as a set of platform-assigned goals. We propose our hypotheses under the "reference point" framework (Heath et al. 1999, Larrick et al. 2009, Wu et al. 2008).

The key factor that makes a reference point important is the difference in the emotions of success and failure (Louro et al. 2007). Loss aversion based on prospect theory implies that individuals will perceive more negative emotion from losses, than positive emotion from gains. Therefore, the marginal benefit of one unit of performance increase should be higher before attaining the goal than after (Heath et al. 1999, Wu et al. 2008), and so is effort level. When incentive hierarchies exist, users who have exceeded thresholds can keep their title (designation) even if they do not contribute any further. Consequently, at least when they are close to the threshold, their optimal decision is to "exceed the threshold, but only by a very little" (Wu et al. 2008). On the other hand, after achieving certain goals, the users will be prompted by the system to attempt the next goal. However, they may have doubts about the attainability of new goals, and hesitate to make progress due to the sheer distance from the next threshold (Zhang and Huang 2010). We hypothesize that users who have just exceeded the threshold will *reduce* their effort level; therefore, effort levels are higher before winning the ranks than after. This indicates that the short-term effect of achieving goals should be negative.

Hypothesis 1: *Contributors of user-generated-content (UGC) websites will reduce levels of effort once they reach goals in a hierarchy.*

The goal-gradient hypothesis posits that motivation increases in proximity to goals (Hull 1932, Kivetz et al. 2006). The reference point framework yields a similar behavioral prediction on the relationship between distance to reference point (i.e. progress toward goal) and motivation. Specifically, the value function of prospect theory exhibits convexity in the loss region and concavity in the gain region, leading to the diminishing sensitivity property. Since the slope of the value function is steeper when closer to the threshold, the theory predicts that motivation should be higher when individuals are closer to the reference point (Heath et al. 1999). This also implies that the effect of progress to goal disappears when individuals are far away from the thresholds. Bonezzi et al. (2011) documented that individuals tend to "get stuck in the middle" of goal pursuit, because individuals in the middle are most distant from reference points at both beginning and end parts of goal pursuit. Their psychological model illustrates a U-shaped goal-gradient, pointing out that the goal effect diminishes as absolute distance to threshold increases. Following this logic, we propose two hypotheses regarding tendency of effort levels, one before achieving goals; the other, after achieving goals.

Hypothesis 2A: *Contributors of user-generated-content (UGC) websites accelerate their effort before they reach the goals (before goal attainment). In other words, for a given length of time, they contribute increasingly more contents as their distance toward the goal decreases.*

Hypothesis 2B: *Contributors of user-generated-content (UGC) websites decelerate their effort after they reach the goals (after goal attainment). In other words, for a given length of time, they contribute increasingly fewer contents as their distance past the goal increases.*

We next describe the website where we gathered the data.

3. Research Context

Our research context is a popular online knowledge exchange, where members ask questions and other community members provide answers, and the answers can be viewed by other members of the site. Launched in the late 1990s, the website is a world-wide question answering platform mostly for IT-related questions. As of early 2012, more than 100,000 users had contributed to the platform's knowledge repository by providing detailed solutions to more than 3 million questions raised by peer members of the site.

In a typical question answering process on this site (illustrated in Figure 2), an asker posts a technical question with assigned points (between 20 and 500 points). The number of points is determined by the asker based on level of difficulty and urgency. After answerers submit their comments, askers can select one or multiple comments as accepted solutions (with grading “A” “B” or “C”) and allocate reward points to answerers. The number of final reward points that the answerers earn is determined by share of assigned points and grading.[2] This process is largely comparable to other online Q&A sites.

[Insert Figure 2 about here]

On this website, users are allowed to accumulate these reward points to obtain ranks in the hierarchy by answering questions or writing articles. Overall, the reward points from posting articles are negligible, for very few users have ever posted any articles. The platform provides two types of incentive hierarchies. The overall performance hierarchy accounts for total contributions to the market, while topic-specific hierarchies refer to users’ expertise in certain topics, such as Windows XP or Perl Programming. When numeric points exceed some predefined thresholds, users will obtain increasingly higher ranks such as “masters” at the first level (50,000 points), “gurus” at the second level (150,000 points), another title at the third level (300,000 points), and so on. The user’s rank in the overall performance hierarchy is shown on the users’ profile avatar, which appears next to the username on all comments that they made on the site, as well as their profile page. On the other hand, numeric total-point values and ranks in topic-specific hierarchies are only visible on the user’s profile page. The overall performance hierarchy is much more conspicuous in the community than total-point values or topic-specific hierarchies, and much more likely to affect user behavior. We therefore focus on the overall hierarchy in this study.

4. Data and Methodology

In this section, we describe the dataset that we gathered for this study, and the empirical methods that we use to test our hypotheses.

[2] Grade “A” means the winner will be awarded points multiplied by a factor of 4; “B” by a factor of 3; and “C”, 2. For example, if a winner gets 80% share from a 500-point question and gets a grade of “A,” she can earn 500×0.8×4=1600 points.

4.1. Data

The platform provides a list of all the users who have earned reward points by providing accepted solutions. We obtained the complete list of answerers, which contains 117,174 users who have at least one reward point as of March 26th, 2012. From this list, we drew a random sample of 2,000 individuals[3] and collected their complete activity history between their first day on the site (when they registered) and March 26th, 2012.

We then constructed an unbalanced panel dataset from these answerers and their activities, where each observation records activities of each user in each week. As is common in online communities, not all users are constantly active. For those who stopped contributing after a certain date, their total-point values do not change, although the website retained their records. To account for such attritions, we dropped individual-period pairs four weeks after each individual's last observed action (in terms of signing-up, answering or asking). This operation mitigated possible estimation bias caused by inactive periods. In our main analysis[4], we included only individual-period pairs after the introduction of the incentive hierarchy program in early 2007, so users who became inactive before early 2007 were dropped from the original sample as well.

4.2. Main Variables and Summary Statistics

Dependent variables. The main outcome that we are interested in is the level of effort that users exert, and we study how the user's distance from hierarchical thresholds affects their efforts. We measure user effort in several ways. The first measurement is the number of questions that a user submitted their initial answers to in each period (i.e. *Number of Questions Attempted*), no matter whether the answers are accepted as correct or not. The platform allows answerers to come back to question pages and make additional comments at any time before the asker closes questions. It is reasonable to assume that the cost of submitting the first comment is the highest, because it involves the cost of reading and understanding

[3] Upon inspection we found that the website recorded one user's signup date incorrectly (it was earlier than the launch of the website), so we dropped that user from our analysis.
[4] In an alternative specification that we describe later in the paper, we analyze individual-period pairs before the exogenous change of 2007 (the introduction of the overall site hierarchy).

the question. We consider additional comments as follow-up discussions for the same solution, so most of the effort is made at the first comment. Therefore, the outcome variable does not take into account questions that users had already posted answers to in previous periods.

We also consider two alternative measurements of user efforts. The first one aggregates the number of points that askers assigned to the questions that the user attempted in that period, or *Number of Points Attempted*. Since ranks in the goal-related hierarchy are defined by whether users' total-points have reached predefined thresholds, this provides a proper measure of actual effort toward the goals. The other measurement is "*Number of Questions Solved*," which captures the number of questions that the user was able to successfully solve in a given period. These measurements are used in various specifications.

Independent variables. The key independent variable that we are interested in is the user's distance from hierarchical thresholds. A natural way of measuring this distance is to compute the difference between the number of points required of the certain hierarchical threshold (such as 50,000 points for the first rank) and the total points that each user had accumulated at the beginning of each period (TotalPoint). However, this metric does not reflect an important feature of this website: there can be a significant delay between the time that answerers submitted their solutions, and the time that askers accepted them. To meaningfully calculate the distance from goals as a motivator (or de-motivator) on user contribution, we need to include the points that the users can reasonably *expect* to receive from the solutions that they submitted previously, even if their solutions have not yet been accepted, in addition to the points that the answers had actually be awarded[5]. In Figure 3, we conceptually illustrate the formation of "incoming points," caused by the time delay between submission and acceptance (of solutions).

[Insert Figure 3 about here]

[5] When we discard these "incoming" points, we find that users' effort levels start to decrease prior to reaching the goal, which lends further support to using this modified distance metric by taking into account these incoming points. Please see Appendix A for details (to be made available online).

To further illustrate this point, suppose an answerer had accumulated 40,000 points at the beginning of a time period(t), and suppose the threshold of the first level on the incentive hierarchy is 50,000 points. A “naive” measurement of distance would be $40{,}000 - 50{,}000 = -10{,}000$ points. However, the user may had submitted solutions to 5 questions in the previous period $(t - 1)$ but the askers of those questions had not yet picked the “accepted” solution. Meanwhile, from his or her past experience on this site, this answerer could expect that at least half of the answers that he or she provided would be accepted by the asker. This is possible not only because of the answerer’s experience, but also due to the technical nature of the site. Hence, the “perceived” distance that would affect the answerer’s behavior in the next period (t) was not the naive -10,000 points, but in fact shorter, because additional points (which we call “incoming” points) would come from the solutions that he or she had provided in the previous period. As a simplified example, assuming a 50% success rate in the answerer’s historical performance, and a maximum 2,000 points per question to be awarded, the distance that would more realistically affect answerer behaviors should be $-10{,}000 + 2{,}000 \times 5 \times ½ = -5{,}000$ points.

More generally, in our analysis we measure the answerer’s distance from threshold as a *Modified Distance*:

$$Modified\ Distance = TotalPoints + IncomingPoints - Threshold$$

The “incoming” points are the points that users can reasonably expect to receive from those opening questions that they had submitted solutions to in the previous period. In turn, incoming points are calculated by multiplying the total number of possible points with the historical “success” rate of the answerer (in terms of probabilities of being selected as the right answer) and “quality” of the answer (in terms of the grade received):

$$IncomingPoints_{it} = SuccessRate_{it} \times Quality_{it} \times \sum_{j \in set\ of\ openning\ questions} AssignedPoint_{ijt}$$

Specifically, $SuccessRate_{it}$ denotes user i’s perception of success rate at time t, which is updated by the user at each period. The metric $Quality_{it} = \frac{1}{J}\sum_{j \in set\ of\ solved\ questions} Share_{ijt} \times Grade_{ijt}$ captures the variation in share of assigned points and grade of previously solved questions, which also varies across user and time. $\sum_{j=1}^{J} AssignedPoint_{ijt}$ denotes the sum of points assigned by askers to user i's opening questions at time t.

In addition to the measurement of distance, we further add several control variables such as *Tenure* (number of weeks since registration) and *Ask Count* (number of questions asked). The descriptions of main variables used in our empirical analysis are presented in Table 1.

[Insert Table 1 about here]

Table 2 provides summary statistics for these variables in our main analysis. It shows that 994 individuals were used for the main analysis, and the unbalanced panel dataset was comprised of 104,622 individual-period pairs. The mean weekly number of questions attempted by users is 0.549, the mean number of question points attempted is 0.237 thousand, and the mean number of questions solved is 0.314. As to ranks in hierarchy, 14.8% of individual-period pairs reached the first level goal, while 8% reached the second level. These indicate that the incentive hierarchy can be challenging to most users, comparable to goals in the goal-setting literature. Similar to other online communities, a large proportion of questions are answered by a small number of users. The average incoming point is 1.197 thousand, which is roughly 1.56% of average total points.

[Insert Table 2 about here]

In Figure 4, we plot the mean values of the main outcome variable, *Number of Questions Attempted*, at different total-point values. Our analysis focuses on the regions close to thresholds of goals, namely 50K and 150K points. At first glance, effort level shows a significant decline after reaching the first threshold, and it is relatively high near the second threshold. In the next subsection, we discuss the empirical model specifications that we use to test the hypotheses.

[Insert Figure 4 about here]

4.3. Model Specification and Estimation

To test our hypotheses, we built an empirical model specifying outcome variable as a function of relative distance (proximity) to goals and other covariates. Our aim was to identify a sudden change at reference point (for H1), and investigate tendency of effort levels near thresholds of goals (for H2A and H2B). Inspired by identification in regression discontinuity design or regression kink design (Thistlewaite and Campbell 1960, Card et al. 2009), we proposed the following parametric polynomial model:

$$y_{it} = \alpha_i + (1 - Goal_{it}) \cdot \sum_{p=1}^{\bar{p}} \beta_{1,p} Distance_{it}^{p} + Goal_{it} \cdot \sum_{p=1}^{\bar{p}} \beta_{2,p} Distance_{it}^{p} + \beta_3 Goal_{it}$$

$$+ \gamma_1 LogAskCount_{it} + \gamma_2 LogTenure_{it} + \delta YearDummies + \varepsilon_{it}$$

The outcome variable y_{it} above can be the number of questions attempted, number of points attempted, or number of questions solved by user *i* at time *t*. $Goal_{it}$ is an indicator whether user *i* reaches the first level or second level goal at time *t*. The coefficient of the indicator variable $Goal_{it}$ provides an estimate for the causal effect of "achieving goal" in a regression discontinuity fashion (Thistlewaite and Campbell 1960)[6] for testing H1. On the other hand, when we test H2A and H2B, we focus on measurements of distance to goal rather than a binary indicator.

The variable $Distance_{it}$ is the specific distance metric to the goal for user *i* at time *t*. We introduced two polynomial functions and multiplied them by indicator $1 - Goal_{it}$ or $Goal_{it}$ to distinguish between loss region and gain region. This model allows asymmetric impact of distance to goal on the two sides of the reference point, as our model predicts that marginal benefit of progress is asymmetric in two regions. If an individual is on the left of threshold, the expression $Goal_{it} \cdot \sum_{p=1}^{\bar{p}} \beta_{2,p} Distance_{it}^{p}$ equals to zero, while the expression $(1 - Goal_{it}) \cdot \sum_{p=1}^{\bar{p}} \beta_{1,p} Distance_{it}^{p}$ appears, and *vice versa*. Thus, parameters of each polynomial expression represent goal pursuit behavior in one region (either loss or gain). We estimated first and second order polynomial models,[7] as well as baseline models without polynomial

---

[6] Following the "local linear regression" identification strategy for regression discontinuity design, researchers often apply narrow bins for distance to goal (Imbens and Lemieux 2008). However, this strategy is not applicable in this study, as an RD design with narrower bins requires a much larger sample size.

[7] Higher order polynomial models require a large sample size, so we stopped at 2nd order.

functions. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were calculated for model selection.

$LogTenure_{it}$, $LogAskPoint_{it}$, and year dummies were incorporated as control variables. We included $LogTenure_{it}$ in the model to control for individual's seniority in the community. On the one hand, more senior users may know more about "answering questions efficiently." On the other hand, as time passes, users may get bored with question answering and reduce levels of effort. $LogAskPoint_{it}$ reflects the cost of participation in the knowledge exchange, for users who ask questions are probably willing to answer questions. The calendar year dummies were also included. We included the α_i terms, or individual fixed effects, to control for time-invariant individual heterogeneity, and assumed that ε_{it} is normally distributed with mean zero for estimation purposes.

The parameters of interest include fixed effects α_i, β parameters , γ parameters, and δ parameters. Because the outcome variable can be expressed in a linear function of observables given parameters and an additively separable disturbance, all the parameters can be obtained by applying the within estimator. In order to check tendency hypotheses H2A and H2B, we calculated the conditional expectation function of the outcome variable given different values of distance to goal for continuity specifications. The constant was calculated by fixing other covariates at their means. The standard errors (for calculating 95% confidence intervals) can be derived by using the delta method. The conditional expectation function is:

$$E[Y|Distance] = Constant + 1_{\{Distance \leq 0\}} \cdot \sum_{p=1}^{\bar{p}} \beta_{1,p} Distance^p + 1_{\{Distance > 0\}} \cdot \sum_{p=1}^{\bar{p}} \beta_{2,p} Distance^p$$

We also specified the marginal effect function at different values of distance for continuity specifications. Because the curve of conditional expectation function is non-differentiable at the threshold, there is a jump in marginal effect at the same place. The marginal effect function was obtained by calculating the first-order derivative of conditional expectation function with respect to distance (with standard errors calculated).

$$\frac{\partial E[Y|Distance]}{\partial Distance} = 1_{\{Distance \leq 0\}} \cdot \sum_{p=1}^{\bar{p}} p \cdot \beta_{1,p} Distance^{p-1} + 1_{\{Distance > 0\}} \cdot \sum_{p=1}^{\bar{p}} p \cdot \beta_{2,p} Distance^{p-1}$$

5. Results

Before we turn to the our results, some conceptual issues should be clarified. First, notice that our empirical model describes users' responses to incentive hierarchy both before and after reaching goals. The classical goal-gradient hypothesis suggests that goals motivate users before attainment (Kivetz et al. 2006), but it seldom discusses the after-goal situation. The reference point framework enriches literature by extending to user efforts after goal attainment, and accordingly, we hypothesize that individuals should be de-motivated after achieving goals. When testing H2A and H2B, we tracked activities of users who ultimately crossed the threshold in the hierarchy. In other words, we were not only testing short-term impact of "achieving goals," but also investigating the impact of distance to goals on level of effort. Finally, compared to previous literature on goal pursuit, our research focuses on a novel "hierarchical goal" setting, where another goal exists after one level is achieved.

5.1. Effect of First Level Goal

We first analyzed the effect of the first level goal by selecting individual-period pairs with modified total-point measures ranging from 0 to 150K. Notice 50K is the threshold for the first level, while 150K is the threshold for the second level goal. There are 974 individuals forming an unbalanced panel data with 96,267 individual-period pairs. Both information criteria (Akaike Information Criterion and Bayesian Information Criterion) support quadratic models[8]. We present our results in Table 3.

[Insert Table 3 about here]

The RD design provides an estimate for the causal effect of reaching the first level goal. In the linear and quadratic model with discontinuity, the coefficients of regressor $Goal_{it}$ are significantly negative (-1.8479, p-value<0.001, and -1.2359, p-value<0.001 respectively), suggesting that the local

[8] The likelihood ratio (LR) test prefers quadratic models over linear models.

effect of “achieving goal” is negative. In terms of practical significance, on average, users reduce their level of effort by 1.85 or 1.24 question per week after they reach the first level goal. Hence, H1 is supported. On the other hand, the continuity specification with quadratic distance provides finer details about the users’ effort level. In Figure 5, the left graph illustrates the conditional expectation function of outcome given different distance (note that the horizontal axis is the modified total points) to the goal and the right graph shows the marginal effect function of one unit increase in distance (progress). The dashed vertical line denotes the threshold of goal. To quantify the goal effect, we select several values of distance and calculate conditional expectations and marginal effects with standard errors in parentheses in Table 4.

[Insert Figure 5 about here]
[Insert Table 4 about here]

From Figure 5, the conditional expectation exhibits a peak near the goal; users first increase before goal attainment and then reduce their effort level after that. Average level of effort monotonically increases before reaching the goal. The average level of effort is 0.08 questions per week when total-point is 0, and it is 1.44 questions when total-point is 49K (1K prior to the threshold). By contrast, level of effort first decreases then increases *after* attainment of the goal. Quantitatively, when total-point is 51K (beyond the threshold by 1K) average level of effort is 1.40 questions, while it is 0.60 questions when total-point is 90K (reduced roughly by 57.1%). After that, effort level increases dramatically, possibly due to the effect of second level goal.

In general, marginal effect of distance to goal is significant (it is insignificant at 45K or 49K) and varies by the distance from the first level goal. Tendency hypotheses H2A and H2B predict that level of effort should be higher near the reference point. They are supported in the given interval (between 0K and 90K). Findings on other covariates such as $LogAskPoint_{it}$ and $LogTenure_{it}$ also offer some interesting insights into user behaviors: users who ask more questions in a given period are also likely to answer more in that period, and users who had been on the site longer tend to exert lower levels of effort. This may reflect user attrition or a more selective behavior in answering questions, or both.

In summary, our analysis of the first level goal in the hierarchy on an unbalanced panel of users demonstrates that users reduce their efforts once they achieve the goal. They increase effort level before reaching the threshold, but decrease after that, lending support to our hypotheses.

5.2. Effect of Second Level Goal

We next consider the second level goal by using individual-period pairs with modified total-point ranging from 50K (first level threshold) to 300K (third level threshold). We constructed an unbalanced panel dataset with 134 individuals and 10,180 individual-period pairs. Compared with the first-level goal, the second-level goal is much more difficult for most users, so emotional response driven by reference point should be stronger. Furthermore, having obtained some experience with the first level goal, users can be more aware of the benefit of obtaining ranks in the hierarchy. Users who have never reached the first level goal were automatically excluded, as no observation is available on their behavior in this point range. We estimated a similar set of models and present the results in Table 5.

[Insert Table 5 about here]

Consistent with the result of the first level goal, the discontinuity specification indicates that the effect of "achieving goal" is significantly negative, with a coefficient of -2.0848 (p-value<0.001) in linear distance measure and -2.4907 (p-value<0.001) in quadratic distance measure. Therefore, H1 is supported in the second level of the incentive hierarchy as well. Based on continuity specification with quadratic distance measure, we illustrate the graphical result in Figure 6, as well as conditional expectations and marginal effects at various distance values in Table 6.

[Insert Figure 6 about here]
[Insert Table 6 about here]

The graphical result in Figure 6 provides additional, strong evidence for goal pursuit behavior. The conditional expectation increases in the loss region, while it dramatically decreases after attaining the goal. In Table 6, when total-point is 100K, the average number of questions attempted per week is 4.30, and it increases to 5.10 when total-point is 149K (roughly by 18.6%), which supports H2A. Then after goal attainment, average number of attempted questions decreases from 5.08 (when total-point equals to

151K) to 3.49 (when total-point is 200K), roughly by 31.3%. Therefore, H2B is supported as well. Moreover, marginal effect is significant, and consistent with our hypotheses (positive before reaching the goal and negative after). In addition, not surprisingly, effects of $LogAskPoint_{it}$ and $LogTenure_{it}$ are consistent to the result of first level goal. To sum up, all our hypotheses of goal pursuit are supported for the second level goal.

5.3. Robustness Check: Time-based Distance Metric

In the main analysis, we measured users' distance-to-goal using the difference between current modified total-point and goal threshold, which incorporates expected earnings from open questions. An alternative approach, as used in Anderson et al. (2013), is to use the number of time periods before or after achieving goals as the distance measure (i.e. distance in time). To test the robustness of our findings, we made several changes in our specifications:

$$y_{it} = \alpha_i + (1 - Goal_{it}) \cdot \sum_{p=1}^{\bar{p}} \beta_{1,p} DistanceInTime_{it}^{p} + Goal_{it} \cdot \sum_{p=1}^{\bar{p}} \beta_{2,p} DistanceInTime_{it}^{p}$$

$$+ \beta_3 Goal_{it} + \gamma_1 LogAskCount_{it} + \delta YearDummies + \varepsilon_{it}$$

The regressor $LogTenure_{it}$ was discarded due to multicollinearity with time distance. The result is consistent with our previous findings, demonstrating that all hypotheses (H1, H2A, and H2B) are supported for both first and second level goals using this time-based distance (See Tables A3 and A4 in Appendix B, to be made available online). The quantitative difference in test results (from the main analysis) may be attributed to the difference in measuring distance to goals, as the literature suggests that different perceptions of goal proximity can lead to different goal pursuit outcomes (Cheema and Bagchi 2011, Liberman and Förster, 2008).

Although time-based distance specification does not require the calculation of expected incoming points, and yields consistent results, we retain the point-based distance in the main specification for two important reasons. First, the time-based distance metric assumes that users can perfectly anticipate the *exact* time that they will reach goals. This is more difficult to justify in our research context than the point-based distance metric, because in addition to the difference in cumulative points, this approach further assumes that the answerer knows how soon the askers will accept their answers, the share of assigned points, the grade or evaluation of their answers, the "supply" of questions in future periods, as well as their competition with other answerers. Second, the time-based distance metric cannot be computed for users who are eventually unable to attain the goals, so this specification reduces the sample size and the statistical power of our tests. For these reasons, we retain the point-based distance from goals in our main analysis.

5.4. Alternative Tests for H1: Exogeneity Due to the Introduction of Hierarchy

To further test H1, we exploit a natural experiment that took place on this website. In early 2007, the overall performance hierarchy was introduced to the community, creating an exogenous environmental change. This "shock" provided an identification opportunity for impact of obtaining ranks in the hierarchy: users who had accumulated more than 50K points suddenly received a rank, whereas those who had less than those points did not. In effect, this exogenous change randomly assigned the new titles (ranks) to some users, but not the others. We exploit this change in two ways.

First, we created a subsample (subsample #1) that contained users who had more than 50K points before the 2007 event. For these users, we conducted simple within-individual (FE) estimations to obtain an estimate for the effect of hierarchy. In the model that we estimated, the regressor $AfterShock_{it}$ is an indicator that equals to 1 if the period is after the 2007 shock. To ensure consistency, we also applied pooled OLS and Quasi-Poisson[9] specifications. In all specifications, year dummies were dropped due to multicollinearity, while in the fixed effects model, variable $LogBioLength_i$ is not identifiable, since it was time-invariant.

$$y_{it} = \beta_0 + \beta_1 AfterShock_{it} + \gamma_1 LogTenure_{it} + \gamma_2 LogAskCount_{it} + \gamma_3 LogBioLength_i + \varepsilon_{it}$$

The coefficient of $AfterShock_{it}$ in fixed effects model is -0.9819 and is significant (In OLS it is -0.8454, while in Quasi-Poisson it is -0.4177. The coefficients are statistically significant in all cases.)

[9] Quasi-Poisson regression is more appropriate than Poisson regression, since it adjusts standard errors to overcome over-dispersion.

(See Table 7). In other words, users who suddenly received a rank in the overall performance hierarchy when the exogenous shock occurred *reduced* their level of effort by 0.98 questions per week (marginal effect of fixed effect model) or by 34.31% (marginal effect of Quasi-Poisson model). This is consistent with our previous result in the polynomial models, that "achieving goal" has a negative impact on level of effort. This analysis focuses on users who had more than 50K points before the shock, to allow for within-user estimates.

[Insert Table 7 about here]

The second approach uses a difference-in-differences model. We gathered users who had more than 50K points in the last periods to form a second subsample (subsample #2) with 164 individuals, and put all users into *full sample* with 1,999 individuals. We thus essentially assigned individual-period pairs with more than 50K points in a treatment group and others in a control group, as a natural experiment. We augmented the previous model by introducing an indicator "having total-point above threshold of 50K ($AboveTreshold_{it}$)" and an interaction term $AfterShock_{it} \times AboveTreshold_{it}$. If the coefficient of interaction term is significantly negative, it indicates that exceeding threshold has a stronger (negative) effect when a hierarchy program exists, thus verifying the negative impact of the sudden introduction of the hierarchy.

$$y_{it} = \beta_0 + \beta_1 AfterShock_{it} + \beta_2 AboveTreshold_{it} + \beta_3 AfterShock_{it} \times AboveTreshold_{it} \\ + \gamma_1 LogTenure_{it} + \gamma_2 LogAskCount_{it} + \gamma_3 LogBioLength_i + \varepsilon_{it}$$

In Table 7, coefficients of $AboveTreshold_{it}$ are significantly positive in all cases, indicating that before the shock, "having more than 50K points" is positively correlated with effort level. In sub-sample two, coefficient of interaction terms in fixed effects model is -2.4352 and significant, meaning that users reduce 2.44 questions per week more in the environment with a hierarchy program. The OLS and Quasi-Poisson specifications yield qualitatively consistent results (marginal effect is a reduction of 2.42 questions per week for OLS or 41.37% for Quasi-Poisson). In full sample specifications, the results are again consistent, with marginal effect equals to -2.44, -2.08 and -50.63%, in OLS, FE and Quasi-Poisson respectively. Figure 7 illustrates the mean outcome values of observations in treatment and control group for linear models (i.e. OLS and FE specifications).

[Insert Figure 7 about here]

5.5. Alternative Specifications

5.5.1. Alternative Dependent Variable: Number of Points Attempted

We previously adopted the *number of questions attempted* as the outcome variable. *Number of points attempted* may be another proper measure for level of effort, since users accumulate points to reach goals. Moreover, because askers assign different points to questions, users might endogenously select questions with more points to participate in. We replicated the main analysis with this new dependent variable for the first level goal (*Modified Distance* was adopted), and the results remain highly consistent (See Table A5 in Appendix C). We also conducted this specification for second level goal, and the result is again consistent. Because askers do not pay points to post questions on this website, most of questions are assigned with upper bound amount of points (500 points). Thus, there is no significant difference between the results based on those two alternative dependent variables.

5.5.2. Alternative Dependent Variable: Number of Questions Solved

In the analysis using *Original Distance* for the first level goal, we uncovered an inverted-U shaped relationship between distance to goal and level of effort before goal attainment (See Appendix A). One may argue that this happens because users learn to answer more efficiently. If so, users can reduce their levels of effort to keep earning the same amount of points. We ruled out this explanation by investigating the impact of original distance to first level goal on actual contribution level, which is measured as the number of questions solved (i.e. number of successful attempts; see Table A6 in Appendix C). Not surprisingly, the result is consistent with previous analyses focusing on number of questions attempted. Therefore, users reduce effort before gathering enough actual points for the first level goal, but not because they can keep the same contribution level by answering more efficiently.

5.5.3. Alternative Model: Cross-Sectional Analysis

Although the fixed effects model yields within-individual estimate, it does not take between-individual variation into consideration. To verify the between-individual effect of first level goal, we specified a cross-sectional model with modified distance and an indicator of attaining the first level goal as the main independent variables. In a short time window (we tried two-week and four-week windows) just after the shock, we investigated behavior of users who had less than 150K points. Based on different distributional assumptions, we conducted both pooled OLS and Quasi-Poisson regression specifications. We present OLS result in Table 8 and Quasi-Poisson result in Table 9.

$$y_i = \beta_0 + \beta_1(1 - FirstGoal_i) \times Distance_i + \beta_2 FirstGoal_i \times Distance_i + \beta_3 FirstGoal_i$$
$$+ \gamma_1 LogTenure_i + \gamma_2 LogAskCount_i + \gamma_3 LogBioLength_i + \varepsilon_i$$

[Insert Table 8 about here]
[Insert Table 9 about here]

The result indicates that correlation between effort level and modified distance to goal is positive (whether before or after goal attainment), which is different from the result of within-individual specifications. This is not surprising because users with higher effort levels are more likely to have higher total-point values. With indicator $FirstGoal_i$, the effect of "achieving goal" is significantly negative. This impact is also economically remarkable, because its marginal effect is about 3.69 questions (reduction) over two weeks, or 5.60 questions over four weeks. The result of Quasi-Poisson regression yields marginal effect of about 89% decline in two weeks and 86% in four weeks. In conclusion, cross-sectional analysis verifies the short term negative effect of achieving goals (i.e. H1 is supported), while the tendency hypotheses are not fully supported. The effects of control variables $LogTenure_i$ and $LogAskCount_i$ are again consistent with the results of within-individual models.

6. Discussion and Conclusion

In this study, we examined the effect of incentive hierarchy on user contribution in an online knowledge exchange by drawing on goal-setting theories (summarized in Table 10). We also showed the importance of taking *incoming* points into consideration. Our results indicate that individuals exert more effort before reaching goals, but lower effort level after that. While incentive hierarchies are intended to induce user efforts, they seem to be only doing so when users are close to the goal thresholds, and before they reach those goals. In other words, the overall impact appears to be temporary.

[Insert Table 10 about here]

Nonetheless, there are many things that designers of UGC sites can do to mitigate these unintended negative effects. Most notably, when goals are excessively difficult, individuals may exhibit stronger loss aversion, and their performance (and level of voluntary contribution) will suffer (Earley et al. 1989, Wu et al. 2008). Managers should better design the difficulty of k-th level rank to fit abilities of individuals who have reached the (k-1)-th level rank. This can be achieved by reducing the distance to the next level of goals, or rewarding more "points" for each unit of effort exerted. Such examples can be found in online games (where high-level players are allowed to work on more difficult tasks associated with more points, thus effectively reducing the actual work load for next level goals), or airline customer loyalty programs (where higher-level customers can earn more points for each purchase).

Our study can be extended in several directions for future research. In our research design, we attributed changes in effort level to changes in relative distance to goals. Although we controlled for many factors, levels of effort may still exhibit some natural trend even without the hierarchy. To check this, we examined trend of effort levels using individual-period pairs *before* the introduction of the hierarchy program. We find the average effort level is slightly decreasing as total-point increases near thresholds, and there is no significant jump (or "kink") at the thresholds (i.e. 50K and 150K; see Appendix D). This serves as a counterfactual analysis for our study, but future research with larger samples over longer periods of time can better address the trend issue in user contributions. Second, there remains some data limitations despite our best efforts. Our original random sample contained 2,000 users drawn from a master list of answerers provided by the website. However empirical tests on the effect of incentive hierarchies can only be tested on a subset of users, such as those who were ultimately able to attain those goals. Consequently, the empirical results are generated from unbalanced panels with relatively few individuals. Though statistical power is large enough for drawing conclusions, the small sample of users does not allow us to conduct more robust specifications (e.g. RD design with narrow bins). Finally, because very few users had attained higher levels in the incentive hierarchy beyond the first two, we cannot test our hypotheses for higher levels.

Despite these limitations, our study provides new empirical evidence on the effect of incentive hierarchy on users' contributing behavior at UGC websites. It also contributes to the IS literature by utilizing goal pursuit behavior in explaining motivation and contribution in online knowledge exchange. More broadly, it also informs the practice of other forms of user and customer interactions, such as crowdsourcing, open innovation contexts, and more broadly the recent trend of "gamification" in various contexts. Although the "public goods" problem of user contribution is less salient in some of those contexts since users are able to internalize more of their efforts, how incentive hierarchies interact with user motivations to induce their efforts can be a rich area for future research.

References

Anderson A, Huttenlocher D, Kleinberg J, Leskovec J (2013) Steering user behavior with badges. *Proceedings of the 22nd International Conference on World Wide Web*, (ACM, New York), 95-106.

Berger J, and Pope D (2011) Can losing lead to winning? *Management Science* 57(5): 817-827.

Bateman I, Munro A, Rhodes B, Starmer C, Sugden R (1997) A test of the theory of reference-dependent preferences. *Quarterly Journal of Economics* 112(2): 479-505.

Bonezzi A, Brendl CM, DeAngelis M (2011) Stuck in the middle: The psychophysics of goal pursuit. *Psychological Science* 22(5): 607-612.

Card D, Lee DS, Pei Z (2009) Quasi-experimental identification and estimation in the regression kink design. Working Paper, Princeton University (Version November 2009).

Cheema A, Bagchi R (2011) The effect of goal visualization on goal pursuit: Implications for consumers and managers. *Journal of Marketing* 75: 109-123.

Chen Y, Harper FM, Konstan J, Li SX (2010) Social comparisons and contributions to online communities: A filed experiment on MovieLens. *American Economic Review* 100(4): 1358-1398.

Crawford VP, Meng J (2011) New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review* 101: 1912-1932.

Earley PC, Connolly T, Ekegren G (1989) Goal, strategy development, and task Performance: Some limits on the efficacy of goal setting. *Journal of Applied Psychology* 74(1): 24-33.

Farber H (2008) Reference-dependent preferences and labor supply: The case of New York City taxi drivers. *American Economic Review* 98(3): 1069-1082.

Hann IH, Roberts JA, Slaughter SA (2013) All are not equal: An examination of the economic returns to different forms of participation in open source software communities. *Information Systems Research* 24(3): 520-538.

Genesove D, Mayer C (2001) Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics* 116(4): 1233-1260.

Heath C, Larrick RP, Wu G (1999) Goals as reference points. *Cognitive Psychology*, 38: 79-109.

Huang SC, Zhang Y (2011) Motivational consequences of perceived velocity in consumer goal pursuit. *Journal of Marketing Research* 48(6): 1045-1056.

Hull CL (1932) The goal-gradient hypothesis and maze learning. *Psychological Review* 39: 25-43.

Imbens G, Lemieux T (2008) Regression discontinuity design: A guide to practice. *Journal of Econometrics* 142(2): 615-636.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2): 263-292.

Kivetz R, Urminsky O, Zheng Y (2006) The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research* 43(1): 39-58.

Koo M, Fishbach A (2008) Dynamics of self-regulation: How (un)accomplished goal actions affect motivation. *Journal of Personality and Social Psychology* 94(2): 183-195.

Koop GJ, Johnson JG (2012) The use of multiple reference points in risky decision making. *Journal of Behavioral Decision Making* 25(1): 49-62.

Larrick RP, Heath C, Wu G (2009) Goal-induced risk taking in negotiation and decision making. *Social Cognition* 27(3): 342-364.

Li Z, Huang KW, Cavusoglu H (2012) Can we gamify voluntary contributions to online Q&A communities? Quantifying the impact of badges on user engagement. *Proc. of 2012 Workshop on Information Systems and Economics (WISE 2012)*, Orlando, Florida.

Liberman N, Förster J (2008) Expectancy, value and psychological distance: A new look at goal gradients. *Social Cognition* 26(5): 515-533.

Lewin K, Dembo T, Festinger L, and Sears PS (1944) Level of aspiration. Hunt JM, eds. *Personality and the Behavior Disorders*, Vol. 1 (Ronald Press, New York), 333-378.

Locke EA (1967) Motivational effects of knowledge of results: Knowledge or goal setting? *Journal of Applied Psychology* 51(4): 324-329.

Locke EA (1968) Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance* 3(2): 157-189.

Locke EA, Latham GP (1990) *A theory of goal setting and task performance* (Prentice Hall, Englewood Cliffs, NJ).

Locke EA, Latham GP (2002) Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* 57(9): 705-717.

Louro MJ, Pieters R, Zeelenberg M (2007) Dynamics of multiple-goal pursuit. *Journal of Personality and Social Psychology* 93(2): 174-193.

Medvec VH, Madey SF, Gilovich T (1995) When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology* 69(4): 603-610.

Medvec VH, Savitsky K (1997) When doing better means feeling worse: The effect of categorical cutoff points on counterfactual thinking and satisfaction. *Journal of Personality and Social Psychology* 72(6): 1284-1296.

Mento AJ, Steel RP, Karren RJ (1987) A meta-analytic study of the effects of goal setting on task performance: 1966-1984. *Organizational Behavior and Human Decision Processes* 39: 52-83.

Miller NE (1944) Experimental studies of conflict. Hunt JM, eds. *Personality and the Behavior Disorders*, Vol. 1 (Ronald Press, New York), 431-465.

Pope D, Schweitzer ME (2011) Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review* 101(1): 129-157.

Pope D, Simonsohn U (2011) Round numbers as goals: Evidence from baseball, SAT takers, and the lab. *Psychological Science* 22(1): 71-79.

Roberts JA, Han IH, Slaughter SA (2006) Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management Science* 52(7): 984-999.

Thistlewaite D, Campbell D (1960) Regression-discontinuity analysis: an alternative to the ex-post facto experiment. *Journal of Educational Psychology* 51: 309–317.

Tubbs ME (1986) Goal setting: A meta-analytic examination of the empirical evidence. *Journal of Applied Psychology* 71(3): 474-483.

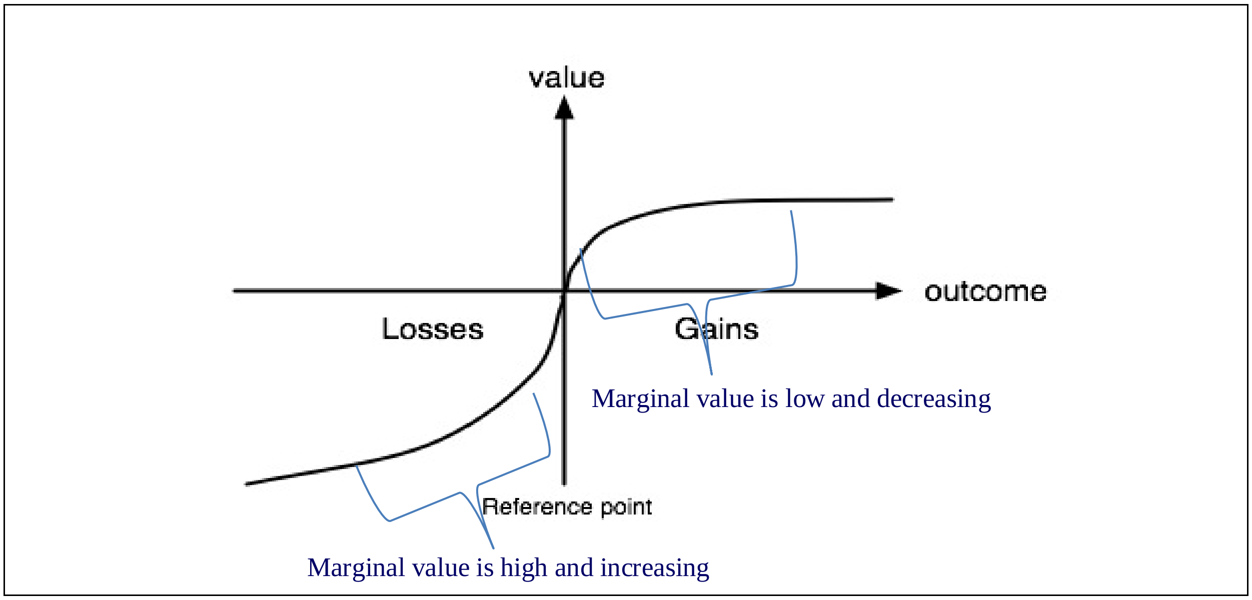
Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5(4): 297-323.

Wu G, Heath C, Larrick RP (2008) A prospect theory model of goal behavior. Working Paper, Graduate School of Business, University of Chicago (Version April 22, 2008).

Zhang Y, Huang SC (2010) How endowed versus earned progress affects consumer goal commitment and motivation. *Journal of Consumer Research* 37(4): 641-654.

Tables and Figures

Figure 1 Value Function of Prospect Theory*



* The value function curve is adapted from Wikipedia (http://en.wikipedia.org/wiki/File:Valuefun.jpg); accessed December 6th, 2013

Figure 2 Time Line of Question Answering

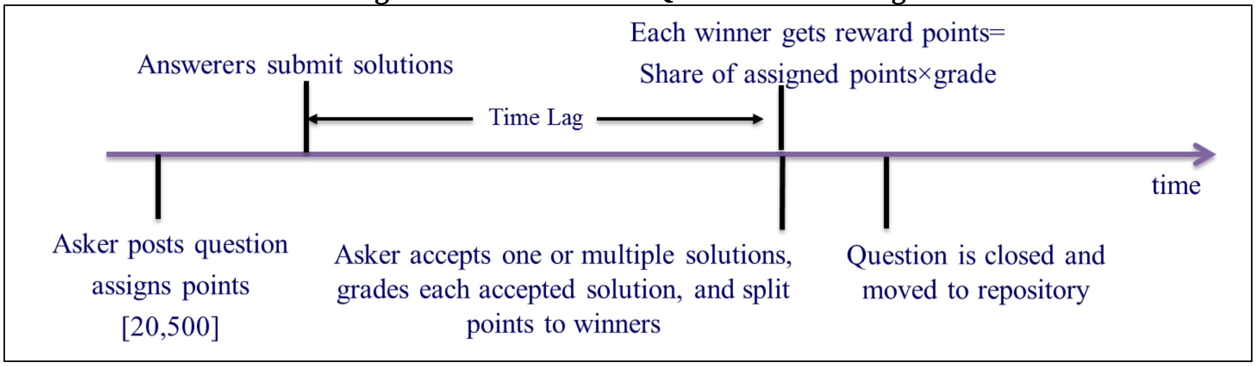


Figure 3 Illustration of Total-Points and Incoming Points

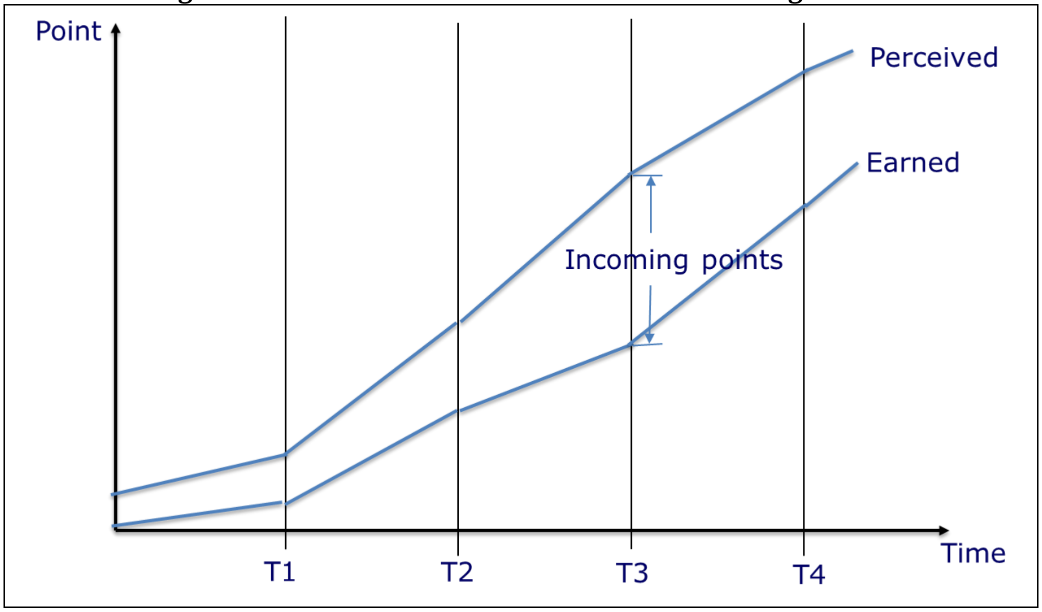


Figure 4 Mean Values of Number of Questions Attempted Given Total-Point

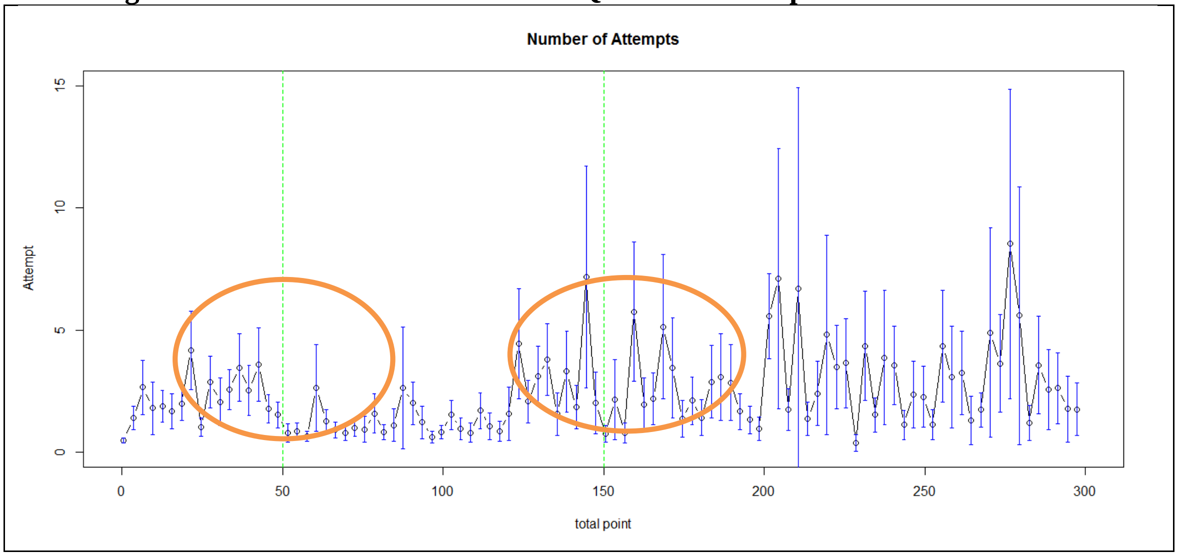


Figure 5 Effect of First Level Goal with Modified Distance

Conditional Expectation	Marginal Effect

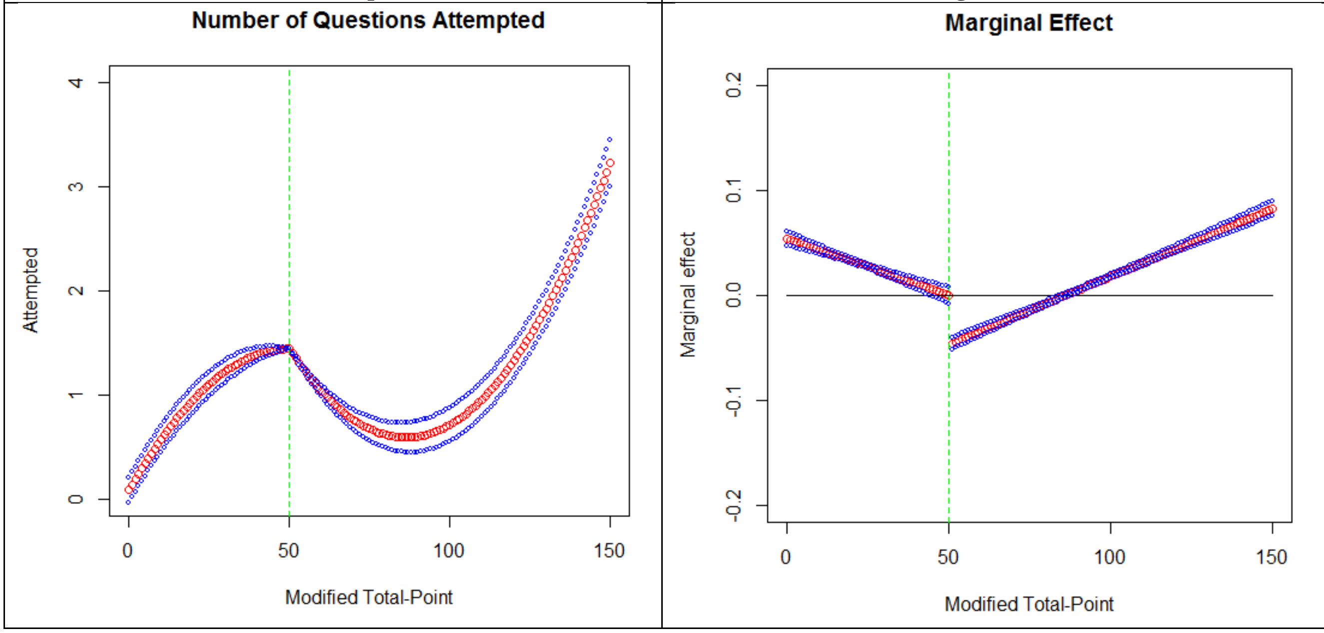


Figure 6 Effect of Second Level Goal with Modified Distance

Conditional Expectation	Marginal Effect

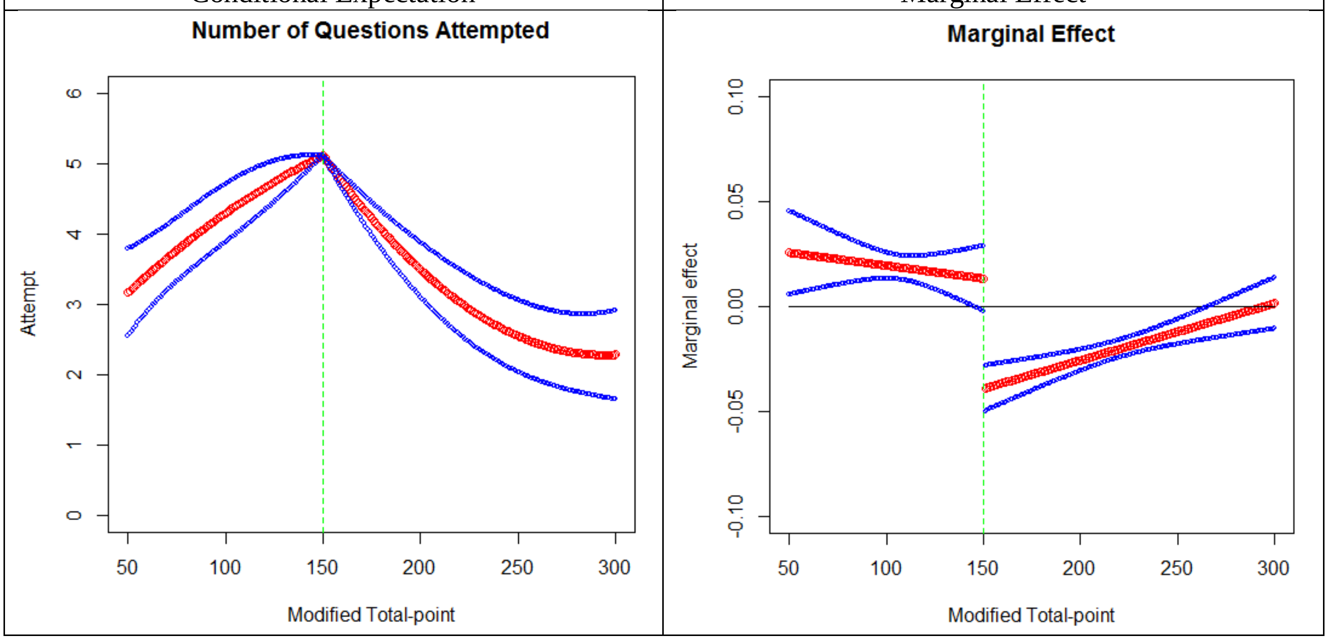
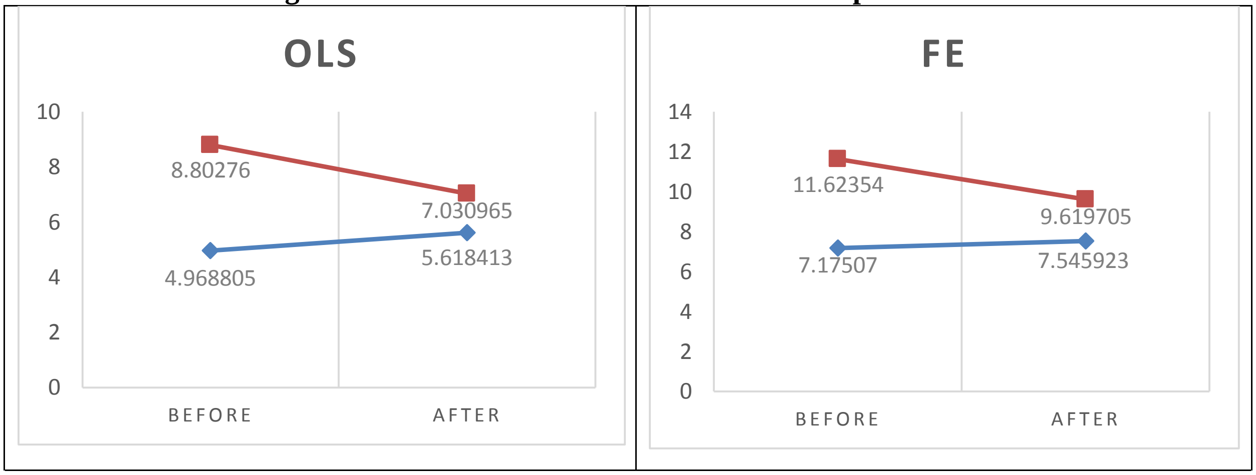


Figure 7 Result of Difference-in-Differences Specification



Note: This is a natural experiment on sub-sample two, # of users=164, N=41256.
The upper lines represent treatment group (The individual-period pairs with more than 50K points).
Effects of control variables are removed.

Table 1 Description of Main Variables

Variable	Description
$NumAttempted_{i,t}$	Number of questions participated (no matter solution accepted) by user *i* at time *t*
$PointAttempted_{i,t}$	Number of question points (in thousands) attempted by user *i* at time *t*
$NumSolved_{i,t}$	Number of questions solved by user *i* at time *t*
$TotalPoint_{i,t}$	Total-point value (in thousands) user *i* has accumulated at time *t*
$IncomingPoint_{i,t}$	Expected incoming points for user *i* at time *t*
$(Level)\ Goal_{i,t}$	An indicator that equals to 1 if user *i* reaches goal (at some level) at time *t*
$AskCount_{i,t}$	Number of questions asked by user *i* at time *t*
$Tenure_{i,t}$	Number of weeks user *i* has been in market since registration at time *t*

Table 2 Summary Statistics of Main Variables

Variable	Mean	Standard Dev.	Min	Max
$NumAttempted_{i,t}$	0.549	3.558	0	254
$PointAttempted_{i,t}$	0.237	1.523	0	105.068
$NumSolved_{i,t}$	0.314	2.511	0	188
$TotalPoint_{i,t}$	76.489	371.854	0	6765.569
$IncomingPoint_{i,t}$	1.197	6.185	0	289.517
$FirstGoal_{i,t}$	0.148	0.355	0	1
$SecondGoal_{i,t}$	0.080	0.271	0	1
$AskCount_{i,t}$	0.086	0.465	0	13
$Tenure_{i,t}$	218.659	151.368	0	794
Observations	# of individuals = 994, # of individual-period pairs = 104,622			

Table 3 Estimation of Parametric Models with Modified Distance (First Level Goal)

Dept. Var.	**Number of Questions Attempted**					
At Threshold	**Continuity**			**Discontinuity**		
Polynomial Order	**0**	**1**	**2***	**0**	**1**	**2**
$\beta_{1,1}$		0.0195***	5.5408e-5		0.0370***	0.0274***
		(0.0012)	(0.0040)		(0.0013)	(0.0046)
$\beta_{1,2}$			-5.4259e-4***			-1.7404e-4*
			(7.1373e-5)			(7.6933e-5)
$\beta_{2,1}$		0.0066***	-0.0470***		0.0246***	-0.0134***
		(0.0010)	(0.0030)		(0.0012)	(0.0040)
$\beta_{2,2}$			6.4849e-4***			3.8043e-4***
			(3.1994e-5)			(3.8245e-5)
$FirstGoal_{it}$				0.0102	-1.8479***	-1.2359***
				(0.0466)	(0.0723)	(0.0968)
$LogAskCount_{it}$	1.6581***	1.6435***	1.6249***	1.6581***	1.6240***	1.6214***
	(0.0357)	(0.0356)	(0.0355)	(0.0357)	(0.0355)	(0.035)
$LogTenure_{it}$	-0.3117***	-0.3778***	-0.3978***	-0.3120***	-0.3953***	-0.3975***
	(0.0146)	(0.0148)	(0.0149)	(0.0146)	(0.0148)	(0.0149)
N (# of users)	96267 (974)	96267(974)	96267(974)	96267 (974)	96267 (974)	96267 (974)
$Deviance$	124810.3	124290	123691.1	124810.2	123631.8	123526.6
AIC	126772.3	126256	125661.1	126774.2	125599.8	125498.6
BIC	136067.1	135569.8	134993.9	136078.5	134923	134840.8

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1.
Coefficients of individual dummies and year dummies are not reported.
Standard errors are in parentheses.

Table 4 The Effect of First Level Goal Given Different Distances

Distance (loss)	**-50**	**-40**	**-30**	**-20**	**-10**	**-5**	**-1**
Conditional Expectation	0.0830	0.5719	0.9523	1.2241	1.3875	1.4284	1.4417
	(0.0630)	(0.0634)	(0.0632)	(0.0544)	(0.0337)	(0.0185)	(0.0040)
Marginal Effect	0.0543*	0.0435*	0.0326*	0.0218*	0.0109*	0.0055	0.0011
	(0.0035)	(0.0022)	(0.0013)	(0.0016)	(0.0027)	(0.0034)	(0.0039)
Distance (gain)	**1**	**5**	**10**	**20**	**30**	**40**	**50**
Conditional Expectation	1.3959	1.2234	1.0369	0.7613	0.6153	0.5991	0.7125
	(0.0030)	(0.0144)	(0.0274)	(0.0490)	(0.0650)	(0.0758)	(0.0821)
Marginal Effect	-0.0457*	-0.0405*	-0.0341*	-0.0211*	-0.0081*	0.0049*	0.0178*
	(0.0030)	(0.0027)	(0.0025)	(0.0019)	(0.0014)	(0.0011)	(0.0011)

Note: *significant at 0.05.
Standard errors are in parentheses.

Table 5 Estimation of Parametric Models with Modified Distance (Second Level Goal)

Dept. Var.	Number of Questions Attempted					
At Threshold	Continuity			Discontinuity		
Polynomial Order	0	1	2*	0	1	2
$\beta_{1,1}$		0.0140*** (0.0027)	0.0113+ (0.0079)		0.0283*** (0.0032)	0.0418*** (0.0086)
$\beta_{1,2}$			-6.1423e-5 (8.5184e-5)			1.5214e-4+ (8.9058e-5)
$\beta_{2,1}$		-0.0211*** (0.0021)	-0.0393*** (0.0056)		-0.0087*** (0.0025)	6.2532e-4 (0.0075)
$\beta_{2,2}$			1.3585e-4*** (3.6637e-5)			-5.7797e-5 (4.3884e-5)
$SecondGoal_{it}$				-1.5137*** (0.1625)	-2.0848*** (0.2401)	-2.4907*** (0.3129)
$LogAskCount_{it}$	2.7572*** (0.2127)	2.6687*** (0.2120)	2.6780*** (0.2118)	2.7979*** (0.2119)	2.6835*** (0.2112)	2.6771*** (0.2112)
$LogTenure_{it}$	-2.5015*** (0.1667)	-2.0078*** (0.1824)	-2.0809*** (0.1836)		-2.1441*** (0.1825)	-2.1106*** (0.1831)
N (# of users)	10180 (134)	10180 (134)	10180 (134)	10180 (134)	10180 (134)	10180 (134)
$Deviance$	27549.33	27437.76	27420.76	27461.71	27361.58	27356.68
AIC	27831.33	27723.76	27710.76	27745.71	27649.58	27648.68
BIC	28850.5	28757.39	28758.85	28772.11	28690.44	28703.99

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1.
Coefficients of individual dummies and year dummies are not reported.
Standard errors are in parentheses.

Table 6 The Effect of Second Level Goal Given Different Distances

Distance (loss)	-50	-40	-30	-20	-10	-5	-1
Conditional Expectation	4.2972	4.4859	4.6623	4.8264	4.9782	5.0496	5.1044
	(0.2121)	(0.1958)	(0.1680)	(0.1268)	(0.0711)	(0.0375)	(0.0078)
Marginal Effect	0.0195*	0.0183*	0.0170*	0.0158*	0.0146*	0.0140*	0.0135
	(0.0032)	(0.0030)	(0.0037)	(0.0049)	(0.0063)	(0.0071)	(0.0077)
Distance (gain)	**1**	**5**	**10**	**20**	**30**	**40**	**50**
Conditional Expectation	5.0786	4.9248	4.7386	4.3865	4.0617	3.7640	3.4935
	(0.0056)	(0.0271)	(0.0526)	(0.0985)	(0.1379)	(0.1711)	(0.1982)
Marginal Effect	-0.0390*	-0.0379*	-0.0366*	-0.0339*	-0.0311*	-0.0284*	-0.0257*
	(0.0055)	(0.0053)	(0.0049)	(0.0043)	(0.0037)	(0.0031)	(0.0026)

Note: *significant at 0.05.
Standard errors are in parentheses.

Table 7 Estimates of Regression Analysis on Exogenous Introduction of Hierarchy

Variables	OLS			Quasi-Poisson			Fixed Effects		
Sample	**Sub 1**	**Sub 2**	**Full**	**Sub 1**	**Sub 2**	**Full**	**Sub 1**	**Sub 2**	**Full**
$Constant$	5.6283***	4.9688***	1.0190***	1.6066***	1.6370***	0.0420			
	(0.3012)	(0.1707)	(0.0238)	(0.0742)	(0.0563)	(0.0266)			
$AfterShock$	-0.8454***	0.6496***	0.2321***	-0.4177***	0.0550	0.2901***	-0.9819***	1.7530***	0.3236***
	(0.1420)	(0.1197)	(0.0145)	(0.0480)	(0.0591)	(0.0276)	(0.1551)	(0.1525)	(0.0197)
$AboveThreshold$		3.8340***	4.3914***		1.4666***	2.8601***		3.2617***	1.7313***
		(0.1182)	(0.0384)		(0.0449)	(0.0255)		(0.1450)	(0.0536)
$AfterShock \times$		-2.4214***	-2.4381***		-0.5340***	-0.7058***		-2.4352***	-2.0797***
$AboveThreshold$		(0.1581)	(0.0469)		(0.0676)	(0.0362)		(0.1717)	(0.0531)
$LogAskCount$	5.4687***	5.1586***	2.5092***	0.9282***	0.9780***	1.4161***	6.2228***	5.8781***	2.5814***
	(0.2151)	(0.1530)	(0.0293)	(0.0382)	(0.0317)	(0.0162)	(0.2141)	(0.1584)	(0.0305)
$LogBioLength$	0.3491***	0.2918***	0.1124***	0.1157***	0.1280***	0.1692***			
	(0.0254)	(0.0161)	(0.0031)	(0.0080)	(0.0061)	(0.0038)			
$LogWeek$	-0.7739***	-1.0859***	-0.2429***	-0.1902***	-0.4377***	-0.4570***	-1.0092**	-1.4490***	-0.2490***
	(0.0639)	(0.0373)	(0.0053)	(0.0156)	(0.0127)	(0.0064)	(0.0663)	(0.0522)	(0.0072)
N (# of users)	21236	41256	269800	21236	41256	269800	21236	41256	269800
	(62)	(164)	(1999)	(62)	(164)	(1999)	(62)	(164)	(1999)
Dispersion Parameter				19.941	20.677	10.387			
Individual fixed effects	No	No	No	No	No	No	Yes	Yes	Yes

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1
For Quasi-Poisson models, dispersion parameters are presented.
For fixed effect models, coefficients of individual dummies are not reported.
Standard errors are in parentheses.

Table 8 Effect of First Level Goal in Cross-Sectional Data (OLS)

Variables	Two-week			Four-week		
$Constant$	2.9307***	4.5884***	5.5262***	3.8545***	5.5148***	6.9341***
	(0.4729)	(0.6011)	(0.6288)	(0.7091)	(0.8948)	(0.9349)
$FirstGoal$	1.2764**		-3.6865***	2.9058***		-5.6036**
	(0.4195)		(0.8527)	(0.6295)		(1.2699)
$Distance(before)$		0.0377***	0.0596***		0.0392**	0.0724***
		(0.0086)	(0.0099)		(0.0128)	(0.0147)
$Distance(after)$		0.0089	0.0498***		0.0516***	0.1138***
		(0.0093)	(0.0131)		(0.0138)	(0.0196)
$LogAskCount$	1.0644**	0.9768**	0.8686*	1.8653***	1.8388***	1.7265***
	(0.3696)	(0.3591)	(0.3534)	(0.4077)	(0.3933)	(0.3867)
$LogBioLength$	0.2083***	0.1517**	0.1524**	0.3093***	0.2363**	0.2375**
	(0.0523)	(0.0515)	(0.0506)	(0.0785)	(0.0768)	(0.0753)
$LogTenure$	-0.5684***	-0.5531***	-0.5315***	-0.7603***	-0.7345***	-0.7010***
	(0.0954)	(0.0925)	(0.0910)	(0.1430)	(0.1378)	(0.1355)
N	480	480	480	480	480	480

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, $^{+}$significant at 0.1
Standard errors are in parentheses.

Table 9 Effect of First Level Goal in Cross-Sectional Data (Quasi-Poisson)

Variables	Two-week			Four-week		
$Constant$	1.1581**	3.6154***	3.8917***	1.2922**	3.3468***	3.6679***
	(0.4233)	(0.5735)	(0.5569)	(0.3934)	(0.5193)	(0.5045)
$FirstGoal$	1.4452**		-2.2157*	1.8150***		-1.9627*
	(0.4409)		(0.9308)	(0.3630)		(0.7638)
$Distance(before)$		0.0536***	0.0624***		0.0462***	0.0554***
		(0.0091)	(0.0090)		(0.0082)	(0.0081)
$Distance(after)$		-0.0005	0.0231*		0.0086^{+}	0.0275**
		(0.0066)	(0.0115)		(0.0052)	(0.0090)
$LogAskCount$	1.2760***	1.3108***	1.2281***	1.2691***	1.3428***	1.3016***
	(0.3487)	(0.3378)	(0.3465)	(0.2042)	(0.1893)	(0.1906)
$LogBioLength$	0.2901***	0.2048**	0.2175**	0.2905***	0.2228***	0.2398***
	(0.0727)	(0.0687)	(0.0690)	(0.0648)	(0.0606)	(0.0605)
$LogTenure$	-0.5939***	-0.6673***	-0.6376***	-0.5554***	-0.6016***	-0.5800***
	(0.0923)	(0.0921)	(0.0913)	(0.0846)	(0.0812)	(0.0804)
N	480	480	480	480	480	480
Dispersion Parameter	5.7452	4.7208	4.5135	6.8392	5.3940	5.1832
Residual Deviance	837.1	675.67	642.46	1161.5	952.96	911.07
Null Deviance	1184.61	1184.61	1184.61	1784.98	1784.98	1784.98

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, $^{+}$significant at 0.1
Dispersion parameters are presented.
Standard errors are in parentheses.

Table 10 A Summary of Results

Hypothesis	Distance Measure	Supported? First	Supported? Second
H1: Contributors of user-generated-content (UGC) websites will reduce levels of effort once they reach goals in a hierarchy. In other words, users' effort levels should be higher before they reach goals than after.	Modified points *	Yes	Yes
	Time	Yes	Yes
H2A: Contributors of user-generated-content (UGC) websites accelerate their effort before they reach the goals (before goal attainment). In other words, for a given length of time, they contribute increasingly more contents as their distance toward the goal decreases.	Modified points	Yes	Yes
	Time	Yes	Yes
H2B: Contributors of user-generated-content (UGC) websites decelerate their effort after they reach the goals (after goal attainment). In other words, for a given length of time, they contribute increasingly fewer contents as their distance past the goal increases.	Modified points	Yes	Yes
	Time	Yes	Yes

Note: H1 is supported by the exogenous shock model and all the alternative specifications.
* "Modified points" refers to the modified point-based distance, which takes into account the number of actual points that the answer had received, as well as the "incoming" points for questions already answered but the askers had not chosen a solution.

Appendix A[1]. What happens when we do not consider "Incoming Points"?

When calculating the distance of goals in our main analysis, we argue that users take into account not only the actual number of points that they had earned, but also the "expected" number of points (incoming points) for questions that they had answered in the previous period.

We now consider an alternative measurement of the distance from goals. If users are motivated only by the *actual* number of points that they have been rewarded, but not the incoming points, then we should observe similar results as reported in our main analysis, i.e. users should increase their efforts all the way up until they reach their goals. This did not turn out to be the case. Specifically, we apply a modified distance metric (based on actual points only, excluding incoming points) to the first level goal, and report the results in Table A1 and Figure A1.[2]

[Insert Table A1 about here]
[Insert Figure A1 about here]

We can see from Figure A1 that under this "naïve" distance metric, users increase their efforts only up until a certain distance *below* the threshold of the first goal. They start reducing their efforts even before the goal is reached. The most plausible explanation for this observation is that users actually can "learn" from their past performance and reach an "expectation" of the number of points that they could earn from questions that they answered previously, even though those questions are still open and the askers had not accepted a solution. In other words, these results lend support to our argument in the main analysis that incoming points—points associated with questions that users had exerted effort on, but had not been rewarded by askers—should indeed be included in the consideration of distance from goals.

[1] All appendices are to be made available online and also directly from the authors.
[2] Results are qualitatively similar for the second-level goal, though the curvature is less prominent due to fewer observations.

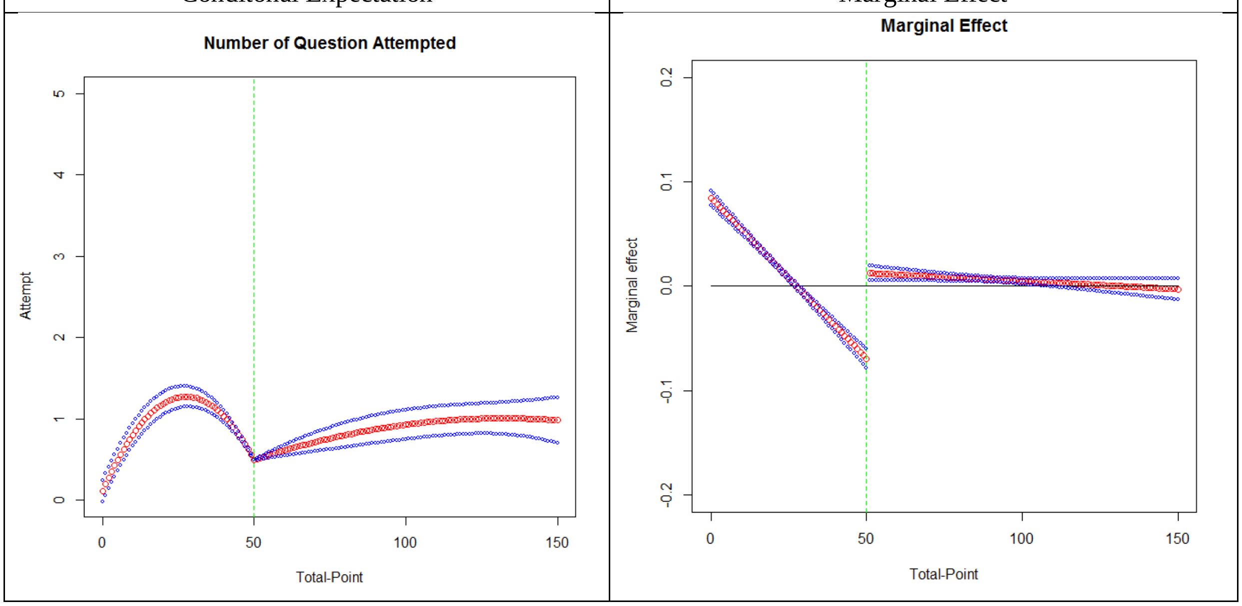
Table A1 Estimation of Parametric Models with Original Distance (First Level Goal)

Dept. Var.	Number of Questions Attempted					
At Threshold	Continuity			Discontinuity		
Polynomial Order	0	1	2*	0	1	2
$\beta_{1,1}$		0.0136^{***}	-0.0692^{***}		0.0235^{***}	-0.0747^{***}
		(0.0012)	(0.0045)		(0.0015)	(0.0059)
$\beta_{1,2}$			-0.0015^{***}			-0.0016^{***}
			(0.0001)			(0.0001)
$\beta_{2,1}$		-0.0026^{*}	0.0125^{***}		0.0049^{***}	0.0092^{*}
		(0.0011)	(0.0037)		(0.0013)	(0.0043)
$\beta_{2,2}$			$-7.6313e\text{-}5^{+}$			-4.6066e-5
			(4.2536e-5)			(4.7267e-5)
$FirstGoal_{it}$				0.0102	-0.8431^{***}	-0.1564
				(0.0466)	(0.0745)	(0.1066)
$LogAskCount_{it}$	1.6581^{***}	1.6552^{***}	1.6389^{***}	1.6581^{***}	1.6486^{***}	1.6390^{***}
	(0.0357)	(0.0357)	(0.0356)	(0.0357)	(0.0357)	(0.0356)
$LogTenure_{it}$	-0.3117^{***}	-0.3476^{***}	-0.3888^{***}	-0.3120^{***}	-0.3603^{***}	-0.3890^{***}
	(0.0146)	(0.0149)	(0.0150)	(0.0146)	(0.0149)	(0.0150)
N (# of users)	96267 (974)	96267 (974)	96267 (974)	96267 (974)	96267 (974)	96267 (974)
$Deviance$	124810.3	124665.8	124237.9	124810.2	124536.4	124235.7
AIC	126772.3	126631.8	126207.9	126774.2	126504.4	126207.7
BIC	136067.1	135945.6	135540.6	136078.5	125827.7	135549.9

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, $^{+}$significant at 0.1
Coefficients of individual dummies and year dummies are not reported.
Standard errors are in parentheses.

Figure A1 Effect of First Level Goal (with 95% Confidence Interval)

Conditonal Expectation	Marginal Effect



Appendix B. Time-Based Distance Specifications

Table A3. Estimation of Parametric Models with Time Distance (First Level Goal)

Dept. Var.	Number of Questions Attempted					
At Threshold	Continuity			Discontinuity		
Polynomial Order	0	1	2^{*}	0	1	2
$\beta_{1,1}$		0.0210^{***}	0.0523^{***}		0.0257^{***}	0.0727^{***}
		(0.0029)	(0.0041)		(0.0029)	(0.0045)
$\beta_{1,2}$			1.7435e-4***			2.5785e-4***
			(1.7432e-5)			(1.9155e-5)
$\beta_{2,1}$		0.0027	-0.0242^{***}		0.0064^{*}	-0.0070
		(0.0029)	(0.0040)		(0.0029)	(0.0043)
$\beta_{2,2}$			1.4214e-4***			7.2027e-5***
			(1.5704e-5)			(1.7053e-5)
$FirstGoal_{it}$				-0.7309^{***}	-1.0052^{***}	-1.7455^{***}
				(0.1323)	(0.1348)	(0.1690)
$LogAskCount_{it}$	3.7689^{***}	3.7647^{***}	3.6951^{***}	3.7656^{***}	3.7671^{***}	3.6442^{***}
	(0.2278)	(0.2269)	(0.2259)	(0.2276)	(0.2264)	(0.2251)
N (# of users)	13895 (127)	13895 (127)	13895 (127)	13895 (127)	13895 (127)	13895 (127)
$Deviance$	42375.59	42246.84	42107.66	42344.82	42190.81	42000.29
AIC	42641.59	42516.84	42381.66	42612.82	42462.81	42276.29
BIC	43644.31	43534.65	43414.54	43623.08	43488.16	43316.71

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, $^{+}$significant at 0.1.
Coefficients of individual dummies and year dummies are not reported.
Standard errors are in parentheses.

Table A4 Estimation of Parametric Models with Time Distance (Second Level Goal)

Dept. Var.	Number of Questions Attempted					
At Threshold	Continuity			Discontinuity		
Polynomial Order	0	1	2^{*}	0	1	2
$\beta_{1,1}$		0.0007	0.0444^{**}		0.0459^{***}	0.2004^{***}
		(0.0073)	(0.0154)		(0.0080)	(0.0187)
$\beta_{1,2}$			3.9072e-4*			0.0017^{***}
			(1.6997e-4)			(1.9201e-4)
$\beta_{2,1}$		-0.0252^{***}	-0.0597^{***}		-0.0168^{**}	-0.0261^{***}
		(0.0059)	(0.0076)		(0.0059)	(0.0079)
$\beta_{2,2}$			1.9741e-4***			6.2915e-5*
			(2.7819e-5)			(2.8851e-5)
$SecondGoal_{it}$				-2.4602^{***}	-3.5905^{***}	-4.9677^{***}
				(0.2487)	(0.2807)	(0.3518)
$LogAskCount_{it}$	3.3185^{***}	3.1899^{***}	3.0722^{***}	3.3151^{***}	3.0379^{***}	2.9780^{***}
	(0.3556)	(0.3555)	(0.3540)	(0.3519)	(0.3495)	(0.3466)
N (# of users)	4527 (54)	4527 (54)	4527 (54)	4527 (54)	4527 (54)	4527 (54)
$Deviance$	15621.88	15591.59	15540.59	15523.76	15428.59	15342.68
AIC	15741.88	15715.59	15668.59	15645.76	15554.59	15472.68
BIC	16128.26	16114.84	16080.73	16038.58	15960.29	15891.26

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, $^{+}$significant at 0.1
Coefficients of individual dummies and year dummies are not reported.
Standard errors are in parentheses.

Appendix C. Alternative Dependent Variables

Table A5 Estimation of Parametric Models with Modified Distance (First Level Goal)

Dept. Var.	Number of Points Attempted					
At Threshold	Continuity			Discontinuity		
Polynomial Order	0	1	2*	0	1	2
$\beta_{1,1}$		0.0090***	-0.0014		0.0169***	0.0132***
		(0.0005)	(0.0018)		(0.0006)	(0.0020)
$\beta_{1,2}$			-2.2586e-4***			-6.5599e-5*
			(3.0963e-5)			(3.3375e-5)
$\beta_{2,1}$		0.0039***	-0.0211***		0.0120***	-0.0065***
		(0.0004)	(0.0013)		(0.0005)	(0.0017)
$\beta_{2,2}$			3.0140e-4***			1.8484e-4***
			(1.3880e-5)			(1.6591e-5)
$FirstGoal_{it}$				-0.0456*	-0.8258***	-0.5375***
				(0.0202)	(0.0314)	(0.0420)
$LogAskCount_{it}$	0.7317***	0.7244***	0.7161***	0.7317***	0.7156***	0.7146***
	(0.0155)	(0.0155)	(0.0154)	(0.0155)	(0.0154)	(0.0154)
$LogTenure_{it}$	-0.1261***	-0.1584***	-0.1670***	-0.1278***	-0.1662***	-0.1669***
	(0.0063)	(0.0064)	(0.0065)	(0.0064)	(0.0064)	(0.0064)
N (# of users)	96267(974)	96267(974)	96267(974)	96267(974)	96267(974)	96267(974)
$Deviance$	-35768.63	-36434.39	-37096.52	-35773.77	-37132.53	-37261.86
AIC	-33806.63	-34468.39	-35126.52	-33809.77	-35164.53	-35289.86
BIC	-24511.77	-25154.58	-25793.76	-24505.44	-25841.25	-25947.62

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1
Coefficients of individual dummies and year dummies are not reported.
Standard errors are in parentheses.

Table A6 Estimation of Parametric Models with Original Distance (First Level Goal)

Dept. Var.	Number of Questions Solved					
At Threshold	Continuity			Discontinuity		
Polynomial Order	0	1	2*	0	1	2
$\beta_{1,1}$		0.0087***	-0.0361***		0.0137***	-0.0398**
		(0.0007)	(0.0028)		(0.0009)	(0.0036)
$\beta_{1,2}$			-8.2801e-4***			-8.7950e-4***
			(4.7460e-5)			(5.7335e-5)
$\beta_{2,1}$		-0.0021***	0.0070**		0.0018*	0.0048+
		(0.0007)	(0.0023)		(0.0008)	(0.0027)
$\beta_{2,2}$			-5.2924e-5*			-3.2693e-5
			(2.6083e-5)			(2.8984e-5)
$FirstGoal_{it}$				0.0344	-0.4327***	0.1046
				(0.0286)	(0.0456)	(0.0654)
$LogAskCount_{it}$	0.2966***	0.2950**	0.2863***	0.2966***	0.2916***	0.2864***
	(0.0219)	(0.0007)	(0.0218)	(0.0219)	(0.0219)	(0.0218)
$LogTenure_{it}$	-0.1692***	-0.1913***	-0.2134***	-0.1704***	-0.1979***	-0.2135***
	(0.0089)	(0.0091)	(0.0092)	(0.0090)	(0.0091)	(0.0092)
N (# of users)	96267(974)	96267(974)	96267(974)	96267(974)	96267(974)	96267(974)
$Deviance$	30554.39	30403.06	30076.67	30552.93	30312.29	30074.08
AIC	32516.39	32369.06	32046.67	32516.93	32280.29	32046.08
BIC	41811.25	41682.87	41379.43	41821.26	41603.57	41388.32

Note: ***significant at 0.001, **significant at 0.01, *significant at 0.05, +significant at 0.1
Coefficients of individual dummies and year dummies are not reported.
Standard errors are in parentheses.

Appendix D. Natural Trend of Effort Level

Figure A3 shows the natural trend of effort before the introduction of a hierarchy system. We fit the data with the same polynomial model (excluding yearly dummies), and then remove the effects of fixed effects and control variables from the outcome variable. Clearly, we observe the trend of effort levels near the thresholds are linear and slightly downward sloped.

Figure A3 Natural Trend of Effort Level Before Introduction of Hierarchy