Genkii - The Effects of Varying Sequential Rewards for User's Motivation to Complete a Set of Crowdfunded Tasks

Ruben Geraldes
rubengeraldes@gmail.com
Instituto Superior Técnico,
Universidade de Lisboa,
Av. Prof. Cavaco Silva,
Taguspark Porto Salvo, Portugal

Daniela Fontes
daniela.fontes@ist.utl.pt
Instituto Superior Técnico,
Universidade de Lisboa,
Av. Prof. Cavaco Silva,
Taguspark Porto Salvo, Portugal

Helmut Prendinger
helmut@nii.ac.jp
National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo, 101-8430, Japan

Rui Prada
rui.prada@tecnico.ulisboa.pt
INESC-ID ,
Av. Prof. Cavaco Silva,
Taguspark Porto Salvo, Portugal

Abstract

A good abstract will leave people satisfied that they know
what you did,
why you did it,
and what you found out.
Keep it at 120 words.

1 Introduction

1.1 Motivation

The ubiquity of hand-held devices opened up new opportunities for research. Nowadays it is accessible to perform off-site studies and collect data in real-time and across great geographical spans.

One of the ways to deploy these off-site studies is by using micro-tasks markets. We are specially interested in this type of crowdsourcing.

This new paradigm also gives rise to the search for new methodologies, and best practices, which may help the scientific community improve and stage new experiments, in a way that optimizes its resources.

Genkii was conceived after realizing some problems in attracting users for crowdsourced studies in Japan. The initial hypothesis for this problem in attracting users for social studies had to do with privacy issues.

We want to study the effects of varying inducements to get people to perform a determined task repeatedly.

1.2 Outline

2 Related Work

2.1 Crowdsourcing

1. Clearly identify the research goals;
2. Select a study method;
3. Devise an incentive mechanism;
4. Choose the target platform(s);
5. Design and develop the mobile app;
6. Prepare data collection;
7. Implement a scheme to obtain informed consent from users;
8. Distribute and promote the app;
9. Continuously monitor data collection for a designated time period;
10. Filter and analyze data to answer the research question

(Henze et al. 2013).

2.2 Rewards and Motivation

> When explicit incentives seek to change behavior in areas like education, contributions to public goods, and forming habits, a potential conflict arises between the direct extrinsic effect of the incentives and how these incentives can crowd out intrinsic motivations in the run short and the long run.

(Gneezy, Meier, and Rey-Biel 2011)

The use of rewards to induce a desired behavior has been thoroughly studied. There has been a clear separation between intrinsic rewards and extrinsic rewards, and research points out that a high focus on incentives can ultimately lead to the alienation of creative thinkers. According to (Gneezy, Meier, and Rey-Biel 2011), there are instances where monetary rewards (extrinsic rewards), work well. Mechanical based, self-contained and well-specified tasks seem to be the primary candidates for the usage of extrinsic rewards to motivate, induce or boost the agent's performance(Ariely et al. 2009).

Crowdsourcing is often associated with Gamification, the use of game-design elements to achieve a more compelling user-experience driven by fun.

However there is criticism when using incentives in areas like education and forming habits, because they can hinder natural intrinsic motivation, and the dependence on rewards may lead to reward inflation, and lower the effort the agent is willing to put towards the task (Irlenbusch and Sliwka 2005).

2.3 Crowdsourcing Platforms

Amazons Mechanical Turk Amazon's Mechanical Turk is a platform which serves as an interface for the deployment of crowdsourced tasks that are considered easier for humans than machines.

It creates a labor market in which individuals or corporations can list tasks (also known as HITs or human intelligence tasks), and a specified compensation. The workers can then elect to complete a determined task, against a deadline, and be compensated upon timely completion. The compensation for the worker is either being paid a determined amount, or a free volunteered work (Mason and Watts 2010). This model has been referred to in the literature as a microtasks market(Kittur, Chi, and Suh 2008).

In (Mason and Watts 2010) it is shown and discussed the potential use of Amazons Mechanical Turk for laboratory experiments for a low fee (between 0.01$ and 0.10$). The main advantages cited for the merit of these studies are the convenience of the platform in reaching many users in a relatively short notice and it is mentioned that the average cost per user is very low, in the order of cents per task. Also it has been shown (Kittur, Chi, and Suh 2008) that, provided that the type of task is well-specified, easily measurable, and do not put a lot of emphasis on creativity, the quality of work performed on Amazons Mechanical Turk is as good, and maybe even better than, work performed by experts paid under traditional contracting arrangements.

Further experiment design recommendations offered in (Kittur, Chi, and Suh 2008) including the concern in devising the tasks such that it discourages random or malicious completion, by making a good performance require the same or less effort than a tampered one.

Yahoo Crowdsourcing Japan The purpose of Yahoo Crowdsourcing Japan is to

3 Genkii

Genkii is an application that enables gps localized satisfaction reports. Users report their "Genkiiness", by performing three different gestures:

- Circle meaning that the person feels happy/genkii;
- Triangle which represents an "OK" state;
- Cross which denotes sadness.

3.1 Implementation

3.2 Crowdsourcing Campaign

Following the methodology suggested by (Choi et al. 2014), we want to perform a study where one group will be considered our baseline by receiving a fixed reward for each task over time. Besides our control group, we devise a second incentive scheme, that using the same amount of points, poses as a progressive reward system. We can compare both reward schemes on 1.

Our goal is to study this and compare these two incentive schemes, specially by studying user enrollment rates and user drop rates.

Table 1: Scheme of rewards used for the crowdsourcing campaign. Using the same amount of reward points, for our first group these rewards are the same for every task over time. On the second group, we devised an increasing reward mechanism.

Reward	Task 1	Task 2	Task 3	Task 4	Task 5	T
Stable Rewards	20	20	20	20	20	
Increasing Rewards	5	10	15	30	35	

The first campaign ran from 19/06/2015 to 26/06/2015. The second campaign featuring increasing rewards ran from 08/07/2015 to 15/07/2015.

3.3 Expectations and Hypotheses

4 Results

The first campaign featuring a fixed reward scheme counted with 436 genkii reports. 115 users installed the Genkii application, and 79 users provided at least one report.

The main goal of this study is to compare the effect of rewards on the user's reporting behaviour. At first glance in order to understand the data collected it's important to study the overall frequency of the user reports. In the Figure 1 we can observe that in fact there seems to be considerable onboarding effect with 39 users making 1 report. But we also have a group of 15 users that made more than the strictly required (10 reports).

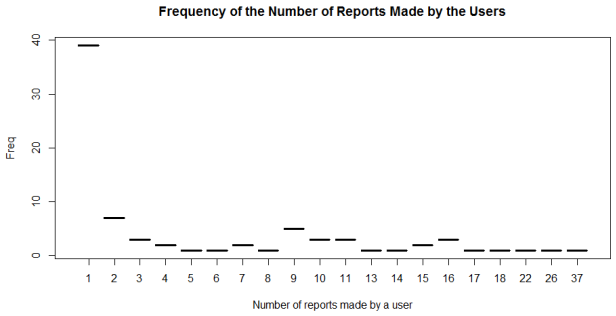

Figure 1: Frequency of reports made by users. As expected, the number of users that made only a single Genkii report (39) stands out. On average each user makes 5.5 reports with a standard deviation of 6.9 reports, the median is 2 reports.

As we can observe in Figure 2 the campaign quickly took off. During the last three days we verified a drop in the number of reports being made. Our hypothesis for this has to do with the architecture of the Yahoo crowdsourcing and a limit

of times the each task could be unlocked, as we were primarily interested on studying the sequential behavior. Also users understood that given the time constraint between reports (4 hours), it was became more difficult to finish the set of 10 tasks.

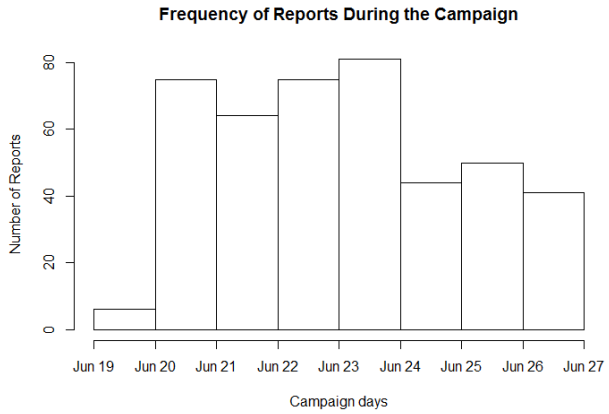

Figure 2: Number of Reports made on each day of the campaign.

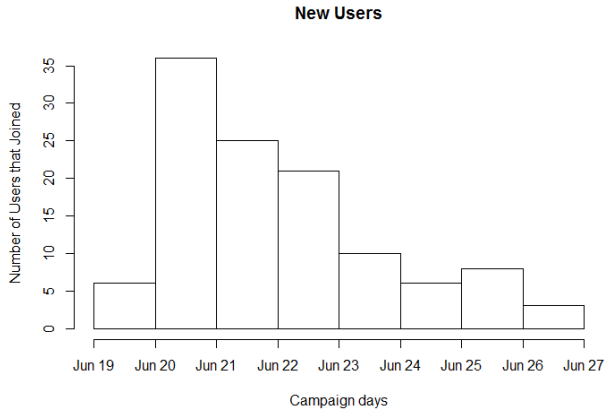

Figure 3: User acquisition throughout the campaign.

The increasing reward campaign run from 08/07/2015 to 15/07/2015. 123 users installed the genkii app during the campaign, with 94 users providing at least one report. We captured 623 genkii reports during this campaign.

This makes the second campaign

4.1 Reward Schemes

In order to compare the two incentive schemes devised we use the user drop rate as a metric for how well the users stick with our 10 task program.

The user participation decreases sharply during the first two quests, on both campaigns. The most notorious difference between the two campaigns is the drop experiences between the first and second quest. The fixed reward scheme has an initial drop rate of 50%. After the first reward, 35% of the users did not claim any more rewards on the increasing reward incentive scheme. In the increasing reward scheme the users have a smaller drop rate between the first and second quest, with 15% more users remaining in comparison with the fixed reward scheme.

Besides the transition from the first to second quest, there seems not to be a major difference in the way the two populations decrease throughout the experiment. However the increasing incentive scheme has a large drop rate (18.2%) between the second reward and the third reward, compared with the drop experienced in the fixed reward scheme (7.3%). The mean difference between the two sets is -3.1% (with a standard deviation of 5%), the fixed reward scheme having a larger drop rate predominantly. However if we do not consider the drop rate that happens on the second task, there is an average -1.8% (with a standard deviation of 2.9%), difference between the increasing rewards and the fixed reward. This points us to the fact that the increasing reward scheme provides a slightly higher incentive to keep users on track.

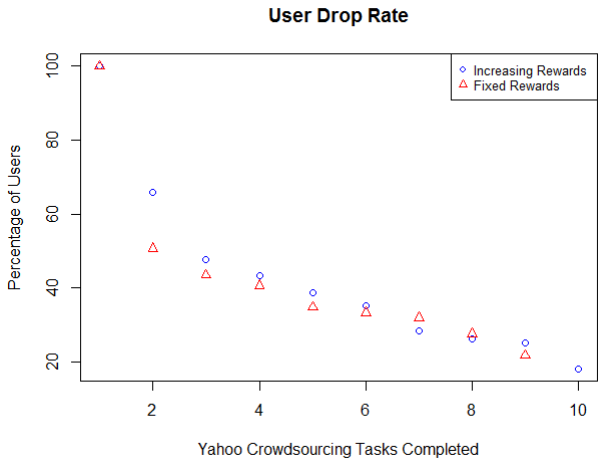

Figure 4: Comparision of the user drop rate for the fixed reward scheme and the increasing reward incentive scheme.

The percentage of rewarded reports during the fixed reward campaign was 64.8%. In the increasing reward campaign the proportion of awarded reports is 62.3%. Although the difference between the two campaigns is 2.5%, users on the increasing reward campaign appear to be more predisposed to report without receiving a reward.

Considering all the reports obtained using two different reward schemes, we verify that the reward does not seem to have an impact on the overall emotional reports captured. This effect can be observed in Figure 5.

Between reward incentives, the fixed reward scheme also verifies the trend displayed in Figure 5, with very slight differences in the overall emotional reports obtained. On the increasing reward scheme, however we found that these dif-

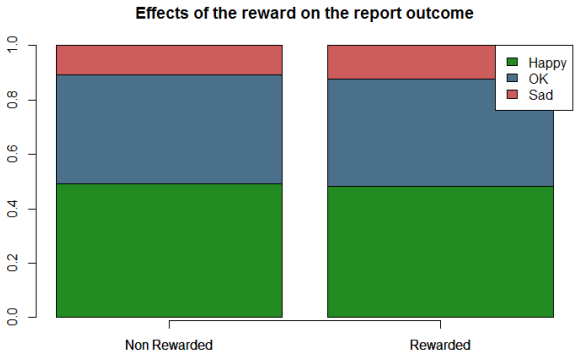

Figure 5: The rewards given do not affect the proportion of emotions reported.

ferences in the emotional reports were more accentuated the rewarded reports displaying a higher "Feeling Sad" percentage 6.

Figure 6: Difference between the proportion of the different emotional states when the report is rewarded or not.

4.2 Genkii Territory

As location enabled emotional report application, Genkii provides insights on the location of the users, and overall mobility.

To capture mobility information we created the concept of "Genkii Territory", which we define as the area that spans between the maximum and minimum latitude and the maximum and minimum longitude reached by a player, adjusted for the earth shape, by assuming a sinusoidal projection. This area is an estimate of overall mobility. It is important to keep in mind that we limited the ability to report consecutively on the same location.

We calculated the "Genkii Territory" for all players that report at least 6 times during both campaigns. There were 65 users that fit the requirements set.

Given the fact that the distribution of the areas that we obtained (in km^2), has a great variance, the standard deviation is 19105.4km^2, the mean being 4698.00, and the median 81.72. We categorize the players, regarding their mobility according to the bins in Figure 7.

Most of the Genkii users who report more often seem to travel daily distances associated with commuting. Either those who stay practically on the same spot ("House Dwellers"), or those who travel long distances seem to be a minority.

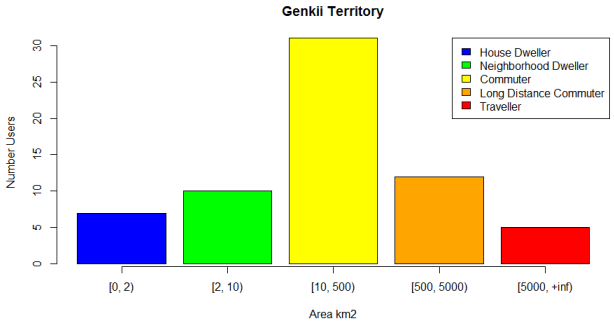

Figure 7: Genkii territory. Given approximate sizes of cities, we figured 5 category ranges to classify our users with regards to their mobility. Those who have a "Genkii Territory" that spans for less than 2km^2 are called "House Dwellers". If they have a territory no larger than 10km^2 they are classified as "Neighborhood Dwellers". Most of our users have a territory comprised between 10km^2 and 500km^2 we figured these values could be common for commuters. We also define a "'Long Distance Commuter" category. And for users which have a "Genkii Territory" larger than 5000km^2 we supply a "Traveller" label.

4.3 Genkii Distribution

In order to capture trends over time we user all the reports collected during both campaigns. With these results we seek to understand more about the users who signed with Yahoo Crowdsourcing to provide Genkii reports.

The mean proportion of registered emotions is 49% of "Happy reports", 39% of "OK", and 12% of "Sad" Reports.

In the Figure 8 capture the hourly distribution of the reports. We can observe that there seems to be a cyclic pattern during the day, with the lowest amount of reports made at 1, 9, and 17. Culturally it is interesting to verify that these are common commuter times.

The fact that we used a total of 1059 reports to construct this visualization seems to help us capture a overall stable group behavior. The daily activity is clearly separated in 8 hour periods.

In contrast the higher amount of reports seems to be made around 4, 12 and 20. Once again we can hypothesize that around 4 we have a lot of reports from people going out at night. The highest peak of our registered reports happens

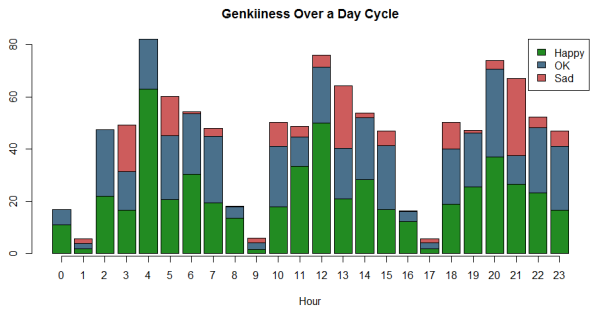

Figure 8: Captured feelings overa day cycle.

just before the public transport resumes in the city of Tokyo (around 5). Although the relative proportion of emotions reported seems not to vary much throughout the day, the 3, 4, and 5 seem to be privileged time for expressing emotions. In the 4 a.m. slot we captured the highest peak of "Feeling Happy" report. The 3 and 5 a.m. slots see some of the highest proportions of "Feeling Sad" reports.

Meal times or meal breaks seem also to be privileged by users to make reports. We always verify during the hour after the slot with a report peak there is a rise on general dissatisfaction, which is interesting because the peak report times (4,12, and 20) also represent peaks of the "Feeling Happy" reports.

These results reinforce that users tend to use the Genkii app as a pastime.

4.4 Gestures

The accuracy (the gesture predicted by the system being the same as the gesture confirmed), is 85.3% for the first campaign, and 83.6% for the second campaign. The prevalence of left-handed people is similar in both campaigns with 7% of self reported left-handed people in the first campaign, and 10% on the second campaign. Overall the percentage of left-handed people was 9.7%

5 Discussion

Despite the inherent advantages of time and reach cost, crowdsourced studies end up forgoing the control over the experimental setting. It is difficult to assure conditions for representative demographics. This fine grained control can be replaced by large numbers in statistical sampling.

During the first campaign, we noticed that 19% of the users that provide reports ended up providing more reports than what the rewarded ones. This points us to the merit of the application in capturing the users' interest. Perhaps the effect of the stable reward experiment made users shift their focus from the reward acquisition to the application. Besides clearly displaying the instructions, and the presence of the timer, Genkii allows the user to explore and play with the application, making reports (as long as successive reports are not made in the same place). The design decision to use gestures started as a way to promote a mechanical task but ended up providing a playful way to make a Genkii report.

6 Conclusions

Future developments for Genkii include the application of the knowledge acquired to In this study, we considered only the usage of monetary inducements. There's an extensive literature on gamification, and the usage of other methods to keep the users engaged. Future steps include offering a version with gamification elements, and after a similar user acquisition compare the engagement and drop rates of the study.

The emotional report data collected seems to plausibly explain trends and cultural aspects. One future work could be trying to match and validate this emotional report method with sentiment analysis from Twitter, for example.

7 Acknowledgments

We thank Daniel Morais for his contribution on the visualization of the Genkii Map website.

References

Ariely, D.; Gneezy, U.; Loewenstein, G.; and Mazar, N. 2009. Large stakes and big mistakes. *Review of Economic Studies* 76(2):451–469.

Choi, J.; Choi, H.; So, W.; Lee, J.; and You, J. 2014. A Study about Designing Reward for Gamified Crowdsourcing System. 678–687.

Gneezy, U.; Meier, S.; and Rey-Biel, P. 2011. When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives* 25(4):191–210.

Henze, N.; Shirazi, A. S.; Schmidt, A.; Pielot, M.; and Michahelles, F. 2013. Empirical research through ubiquitous data collection. *IEEE Computer* 74–76.

Irlenbusch, B., and Sliwka, D. 2005. Incentives, decision frames, and motivation crowding out an experimental investigation. IZA Discussion Papers 1758, Institute for the Study of Labor (IZA).

Kittur, A.; Chi, E. H.; and Suh, B. 2008. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, 453–456. New York, NY, USA: ACM.

Mason, W., and Watts, D. J. 2010. Financial incentives and the "performance of crowds". *SIGKDD Explor. Newsl.* 11(2):100–108.