Increasing the Reach of Snowball Sampling: The Impact of Fixed versus Lottery Incentives

Aditya Vashistha
University of Washington
adityav@cs.washington.edu

Edward Cutrell
Microsoft Research India
cutrell@microsoft.com

William Thies
Microsoft Research India
thies@microsoft.com

ABSTRACT
Though many researchers have studied how to incentivize people to respond to surveys, little is known about how these incentives impact respondents' willingness to recruit others to participate as well. In this paper, we show that the incentives offered for individual survey responses can have a dramatic impact on the overall reach of a survey through a network of peers. In a field experiment in India, we made a survey accessible via mobile phones and offered respondents either a fixed incentive (guaranteed payment of about $0.17) or a lottery incentive (1% chance of winning $17). When asked to choose, a significant fraction of respondents preferred the lottery incentive. However, when encouraged to spread the survey, the fixed incentive spread over 100 times further, reaching about 800 people in a day. We interpret this surprising result and discuss the implications for HCI.

Author Keywords
Snowball sampling; survey response; crowdsourcing

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
To reach out to inaccessible, excluded, or hidden populations, researchers often recruit participants based on chains of personal referrals, a process known as snowball sampling [1,5,11,20]. Snowball sampling suffers from the obvious limitation that participants are not selected randomly from the population, thereby limiting the statistical validity and generalizability of study findings. Nonetheless, snowball sampling is often embraced as the only way to approach hidden or marginalized populations, such as drug users [17], prostitutes [19], pickpockets [15], and low-income respondents in the developing world [13,18,22,23,26]. Researchers have also championed the use of snowball sampling in social computing research, where a global directory of all users is usually unavailable and snowball sampling can be viewed as a form of convenience sample [4].

CSCW '15, March 14 - 18, 2015, Vancouver, BC, Canada.

http://dx.doi.org/10.1145/2675133.2675148

For snowball sampling to be successful, each participant needs to recruit a sufficient number of other people to join the sample. While many researchers have studied incentives for increasing participation in surveys [10], prior work focuses almost exclusively on improving the *response rate*. In the context of snowball sampling, the response rate is only half of the equation; it is equally or more important to know whether someone will *spread* the survey opportunity to others. One way to promote spread is to use referral bonuses that explicitly reward participants for recommending others. However, this approach requires a means of tracking the chain of referrers and can be difficult to implement in practice, especially in low-income communities that have limited access to technology. As an alternative approach, can one design the incentives for survey response so as to make the survey as appealing as possible to spread to others? We are not aware of prior work that considers this question.

In this paper, we demonstrate that the incentives offered for survey response can have an enormous influence on the eventual spread of a survey. We consider two of the most common incentives used in HCI research: a fixed incentive, in which each respondent receives a small reward, and a lottery incentive, in which respondents have a chance of winning a larger prize [27]. There is a large literature that debates the relative merits of these incentives (e.g., [7,12]). In the context of survey response, some studies favor lottery incentives (e.g., [6,14]) while others favor fixed incentives (e.g., [24,25]). However, nothing is known about which of these incentives promotes spreading the survey to others, increasing the overall reach of snowball sampling.

Via a field experiment in India, we show that a mobile survey with fixed payment to respondents spread to about 100 times more people (in a single day) than surveys using a lottery. At the same time, individuals who were given a choice between fixed and lottery incentives preferred the lottery incentives by a significant margin. In other words, what people choose for themselves is different than what they spread to others.

We attribute this surprising result to the immediate gratification afforded by a fixed incentive, as well as the guaranteed social capital that participants accrue by passing this earning opportunity to their friends. While our primary contribution is to the methodology of peer recruitment and snowball sampling, we also discuss the broader implications for HCI, including incentives for crowdsourcing and behavior change.

RELATED WORK

Increasing the spread of a survey is related to viral marketing. However, "viral" spread is characterized by a long chain of referral events, which is sufficient but not necessary for obtaining a large response to a snowball sample. In a recent book, Berger summarizes years of evidence-based research and points to six factors influencing virality: social currency, triggers, emotion, public visibility, practicality, and narrative [3] . Of these, social currency, emotion, and practicality are linked to the spread that we observe in our experiment.

There is also a wealth of research in behavioral economics that examines decision making under risk. For example, Prospect Theory predicts that people give too much weight to small probabilities [16], consistent with individuals' preference for lottery incentives over fixed incentives. However, we are unaware of any prior work that links probabilistic incentives to the eventual spread of a survey.

Crowd mobilization challenges such as the DARPA Network Challenge [21] have demonstrated the power of hierarchical incentive schemes to recruit participants. Our study is different because we do not offer financial rewards for making referrals, which can be difficult to track in practice. Instead we show that differences in task payment alone are sufficient to spur large variations in the reach of a survey.

HYPOTHESES

In prior work on mobile crowdsourcing [13] and phone-based surveys [8], we observed that a small but guaranteed earning opportunity led to rapid referral of new participants via word of mouth. We hypothesized that this rapid spread was caused by the gratification and trust engendered by the guaranteed payout: as soon as participants received payment, they knew the benefit was real and eagerly enrolled their friends. Our intuition was that participants did not have any *a priori* preference for a fixed payment; if asked to choose in advance of performing the task, participants might find it more enticing to have a chance of winning a large prize.

Based on this experience, we formulated more precise hypotheses. Consider a *fixed incentive*[1] that guarantees an immediate micropayment P to each respondent, and a *lottery incentive* that offers each respondent a small chance (1/100) of immediately winning a large payment ($100P$). In expectation, these alternatives offer the same amount of compensation, and a risk-neutral actor may not prefer one over the other. However, our hypotheses are as follows:

H1: Individuals who have not yet completed the task prefer the lottery incentive over the fixed incentive.

H2: In a group where some people have completed the task, the fixed incentive survey spreads to a greater number of total respondents than the lottery incentive survey.

Taken together, these hypotheses imply that participants' personal preference in advance of the task is *different* than what actually spreads in a real-world environment.

EXPERIMENTAL METHODOLOGY

To test our hypotheses, we designed a study that also served as a useful survey for a social enterprise in Bangalore, India. Our partner is Babajob, a mobile and web-based jobs portal for workers in the informal labor sector, such as cooks, maids, drivers, etc. Babajob sought to know whether current and potential job-seekers on its platform own a government-issued ID, such as a passport, driver's license, ration card, etc. For the benefit of the organization, we gathered this information using a phone survey, seeded with registered users and spread to potential users. At the same time, we tested our hypotheses by offering different incentives for users to respond to and spread the survey.

The seed group of participants were all drivers, who were registered as potential job seekers on Babajob's platform. The company supplied us with a list of about 300 drivers, half from Mumbai and half from Delhi. While we did not collect demographic information on these participants, they were likely very similar to those described in a prior paper [8], which surveyed drivers registered with Babajob in a different Indian city (Kolkata). Those drivers were found to be exclusively male and about 31 years old. The majority had received 10 years or fewer of education, and their average individual income was Rs. 8,300 (USD 140) per month. The majority owned a feature phone, or otherwise a basic phone; only one person in 20 owned a smart phone.

Our protocol used a between-subjects design with three conditions (see Table 1). In all of the conditions, we cold-called participants who were already enrolled as drivers in the company database. We introduced ourselves as an associate of the company, and we explained the purpose of the call: to assess whether the user has a government-issued ID. The arms of the study differed by the specific information solicited, and the incentive offered to participants.

To test H1, we conducted a single arm, the BASELINE. In this arm, we assessed whether participants owned a government ID by way of a verbal question and response. We described two alternate payments for the participants' response: a guaranteed payment of Rs. 10 (USD 0.17), or a 1-in-100 chance of winning Rs. 1,000 (USD 17), both delivered as mobile airtime credit ("top-up"). To collect users' preferred payment, we sent them an SMS that re-iterated the options and included two phone numbers on which to reply (see Table 1). Participants indicated their choice of payment by giving a missed call on one of the numbers. (A *missed call*, common in India, is one that rings but hangs up before it is answered, thereby incurring no expense to the caller). Upon receiving the missed call, we

[1] This term has two aspects: guaranteed (non-zero) payment, and constant amount of payment. We use the term "fixed" to align with prior literature, but it may be that the guarantee of payment is the more important part. A guaranteed but variable payment (with the same expected value as our fixed payment) would be interesting to explore in future work.

Table 1: SMS's used in the study.

Condition	SMS (Translated from Hindi)
BASELINE	Greetings from Babajob. If you want to get a guaranteed recharge of 10 Rs then give a missed call on <number1>. If you want to try to win 1000 Rs lottery recharge, then give a missed call on <number2>. Please note that only 1 in 100 people will win the lottery. You can choose only one of the two options. This offer is valid for only 1 hour. Thanks!
FIXED	Do you have a government ID? If your answer is Yes, give a missed call on <number1>. If your answer is No, give a missed call on <number2>. As soon as you give a missed call on any one of the two numbers, **we will recharge your mobile with 10 Rs. Everyone who gives a missed call will get a 10 Rs top-up.** This offer is valid only until <date>, 7PM. Please tell your friends as well.
LOTTERY	*Same as fixed incentive, except bold text is as follows:* **… you will be eligible to win a lottery of 1000 Rs mobile recharge. Please note that only 1 person in 100 people will win 1000 Rs mobile recharge …**

issued payment within 15 minutes using an online portal, easymobilerecharge.com. The reasons that we used an SMS to solicit the user's response (as opposed to asking for it in the phone call) are two-fold: (i) to minimize the response bias inherent in verbal interviews [9], and (ii) to avoid a selection bias relative to other arms, which require participants to understand SMS. The order of options in the phone conversation and the SMS was randomized and balanced across participants. In this baseline condition, we told participants not to spread the survey to anyone else.

To test H2, we conducted two additional arms: FIXED and LOTTERY. In both arms, we called the participants as before. However, instead of soliciting their ownership of a government ID over the phone, we asked them to submit this information by sending a missed call on one of the two numbers (one for `yes' and another for `no'). In the fixed incentive arm, we offered Rs. 10 (USD 0.17) for their reply, while in the lottery incentive arm, we offered a 1-in-100 chance of a Rs. 1,000 (USD 17) payment for their reply. After explaining this offer over the phone, we sent users an SMS that restated the survey question and provided the phone numbers on which to reply (see Table 1). Unlike the baseline condition, we encouraged users to spread this survey question (and earning opportunity) to as many people as possible, up until 7:00pm on the day of the experiment. Because the survey was summarized in a single SMS, users could easily forward the message to others.

Whenever we conducted a lottery, we determined the payment to each participant using a random number generator that was evaluated only at the time of payment. This lent credibility to the experimenter, who could offer each participant an honest 1-in-100 chance of winning. However, it also implied that very few participants (in fact, zero) actually won the Rs. 1,000 payment. In order to study the effect of winning on the spread of the survey, we also conducted another variation of the experiment ("lucky incentive") where the first 5 participants won Rs. 1,000 each.

Table 2: Details of the experiments.

Condition	Date	City	Received our call	Received our call & replied
Baseline	May 10	Mumbai	34	20
	May 17	Delhi	28	20
Fixed Incentive	May 16	Mumbai	15	12
Lottery Incentive	May 13	Mumbai	15	13
	May 20	Delhi	24	14
	May 21	Delhi	15	14
Lucky Incentive	May 22	Delhi	10	6

In the fixed incentive arm, we contacted 15 drivers via phone; 12 of them replied with a missed call. In the lottery incentive arm, we contacted drivers until at least as many of them (12) replied with a missed call. In the lucky incentive arm, we contacted participants until five of them replied; each won Rs. 1,000 (we would have made more awards if not for budget constraints). We ran each study from about 10:30am to 7:00pm. While each arm restricted attention to a specific location (Mumbai or Delhi), drivers within a given location were selected for the experiments at random. We checked that the phone numbers of participants were not repeated across any two of the experiments.

To improve our confidence in the results, we replicated the baseline and lottery arms of the study multiple times (with different participants each time). The lucky and fixed incentive arms were too costly for us to replicate. A summary of the dates and size of each experiment appears in Table 2.

Phone calls were conducted in Hindi by a native Hindi speaker (male, age 27). For participants who won Rs. 1000, we also conducted follow-up interviews the day after the experiment. While we did not record the conversations, we took detailed notes that form the basis for our reporting.

QUANTITATIVE RESULTS

The baseline experiment encompassed 20 respondents from Mumbai and 20 respondents from Delhi. Results were identical in both locations: 13 drivers preferred the lottery incentive, while 7 preferred the fixed incentive. We tested H1 using a one-tailed binomial test (if drivers had an equal 50/50 preference for the lottery and fixed incentive, what is the chance that we would observe at least 26 of 40 responses in favor of the lottery?) The result ($p=0.04$) supports H1.

Results of the lottery experiments appear in Figure 1. The lottery survey did not spread very far at all. Across three trials, it spread to 6 people, 6 people, and 1 person. Similarly, the experiment with the `lucky' incentive, where the first five people won Rs. 1000, spread to only 6 people (see Figure 2).

Results of the fixed-incentive experiment appear in Figure 3. In contrast to the survey with lottery incentives, the fixed-incentive survey spread rapidly, logging 800 respondents by the announced closure time (7:00pm). Even after payments stopped, it spread to an additional 532 people overnight. We received missed calls from 11 states in India. 62% of calls were from Mumbai (where we seeded the experiment), 8% from other cities in the same state (Maharashtra), 13% from

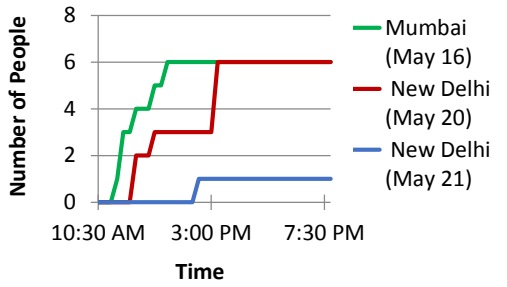

Figure 1: Spread of lottery incentive.

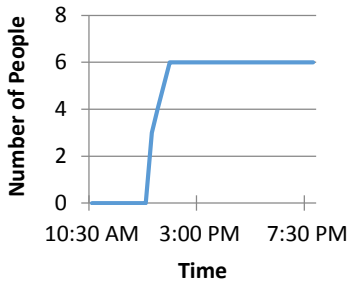

Figure 2: Spread of lucky incentive.

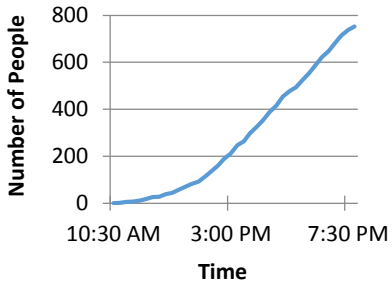

Figure 3: Spread of fixed incentive.

the state of Bihar and rest were spread across 9 other states. The rate of spread was approximately constant during the day. The spread might have been faster if the mobile recharge was automatic; we processed all payments manually, which added a delay of up to 15 minutes. In combination with the results of the lottery surveys, these results confirm H2.

QUALITATIVE RESULTS

To further understand the lack of spread in the lottery case, we conducted interviews with the five participants who were awarded Rs. 1000. As described previously, these awards happened only in the last (`lucky') trial, as an experiment to understand how winners would spread the survey to others. Our interviews confirmed that all of these participants received the Rs. 1000 top-up, and each of them told some of their friends. However, they did not tell many friends; usually they told 2-3 people, and at most 7 people.

Why didn't the Rs. 1000 winners tell more of their friends about the survey? Three themes emerged out of our interviews. First, their friends usually thought that the award was a joke or scam; they had to see the SMS payment confirmation in order to believe that Rs. 1000 was paid out. As one participant said (quotes translated from Hindi):

> When I first told them, they said it is fake, it is a lie and it is a trick to lure me. After that, I showed them the SMS. I also showed them congratulations SMS and after seeing that they somewhat believed me. Otherwise, it is hard to believe.

In a similar vein, the winners did not want to risk their own credibility and social capital by convincing friends to enter a lottery that they would not win. As one respondent said:

> When three friends gave a missed call and received the message 'Sorry you did not win', then I didn't tell anyone else. Because I thought that people I tell will think I am playing a prank on them.

The third theme emerging from the interviews is that winning participants did not want to share the survey too broadly out of a desire to maintain exclusivity. By winning the prize, they had become `special' and desired to maintain that distinction in the eyes of their friends. One person said:

> The profit I have earned, why should others get it? That is why I will tell it only to some special people.

DISCUSSION

Why did the fixed incentive spread faster than the lottery incentive? We offer three explanations. First, the fixed incentive offers immediate gratification, leading to a high state of arousal that can increase sharing behavior [2]. Second, immediate payment proves to participants that the offer is legitimate, boosting their confidence that they will gain social capital by spreading the news to friends [3]. Third, even winners of the lottery are unlikely to spread the survey widely, because the large prize gives them a sense of exclusivity. A small, guaranteed incentive may be "just right" to share with a large number of people.

Why, then, might people prefer the lottery incentive to the fixed incentive when given a choice? It could be that the lure of Rs. 1000 (almost four times their daily wage) is very enticing, or that people overestimate their personal odds of winning. It's possible that Rs. 10 is simply too small to interest some respondents. Finally, the chance of a large incentive may feel more like a game, making the process exciting even if the endpoint is unlikely to be profitable.

Our study has some limitations that will be important to address in future work. Our users were limited to low-income drivers in India, and generalizability to other contexts remains untested. Evaluating other incentives, such as a 50/50 chance of payment, a guaranteed base payment with variable bonus, or longer delays between response and payment, will be important to understand the full space of parameters and their influence on sharing behavior.

CONCLUSIONS

We conducted a field experiment that compares the effects of fixed incentives and lottery incentives on the total response volume of a snowball sample. While participants prefer the lottery incentive for themselves, they are much more likely to spread the fixed incentive to others.

While this study contributes most directly to the domain of snowball sampling, there could also be broader implications for HCI. Completing a survey is only one of many tasks that might be beneficial to spread via social networks. While crowd mobilization challenges such as the DARPA Network Challenge [21] have explored incentives for the spread of tasks, there was only a single winner. Thus, the strategies explored are variants of a lottery rather than a fixed payment. Our results suggest that a fixed payment to each participant may be able to mobilize an even larger crowd in a disaster scenario, something that is worthy of follow-up study.

Our results may also have implications for approaches to behavioral change. If an intervention offers small benefits with high probability, it might spread further (and thus impact more people) than an intervention which offers large benefits with low probability. By the same reasoning, the jobs portal in our study is considering changing its marketing strategy: rather than promote the chance of getting a new job (a big but unlikely prize), they could advertise a small but assured benefit (e.g., direct payment for creating a profile). To control costs, payment can be done on a first-come first-served basis, up to some maximum number of respondents.

ACKNOWLEDGMENTS

We are grateful to Sean Blagsvedt and Maya Chandrasekaran of Babajob for their assistance with the study. We also thank the anonymous reviewers for their helpful comments.

REFERENCES

1. Atkinson, R. and Flint, J. Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social Research Update 33*, (2001).
2. Berger, J. Arousal increases social transmission of information. *Psychological science 22*, 7 (2011).
3. Berger, J. *Contagious: Why Things Catch On*. Simon & Schuster, 2013.
4. Bernstein, M.S., Ackerman, M.S., Chi, E.H., and Miller, R.C. The trouble with social computing systems research. *CHI EA*, (2011).
5. Biernacki, P. and Waldorf, D. Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods Research 10*, 2 (1981), 141–163.
6. Bosnjak, M. Prepaid and Promised Incentives in Web Surveys. *Social Science Computer Review*, (2003).
7. Brown, J., Schonfeld, T., and Gordon, B. "You May Have Already Won...": An Examination of the Use of Lottery Payments in Research. *IRB: Ethics and Human Research*, (2006).
8. Chakraborty, D., Medhi, I., Cutrell, E., and Thies, W. Man versus machine: evaluating IVR versus a live operator for phone surveys in India. *DEV*, (2013).
9. Dell, N., Vaidyanathan, V., Medhi, I., Cutrell, E., and Thies, W. "Yours is better!" Participant Response Bias in HCI. *CHI*, (2012).
10. Edwards, P., Roberts, I., Clarke, M., et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database of Systematic Reviews*, (2009).
11. Faugier, J. and Sargeant, M. Sampling hard to reach populations. *Journal of Advanced Nursing 26*, 4 (1997), 790–797.
12. Gajic, A., Cameron, D., and Hurley, J. The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey. *European Journal of Health Economics,* (2012).
13. Gupta, A., Thies, W., Cutrell, E., and Balakrishnan, R. mClerk: Enabling Mobile Crowdsourcing in Developing Regions. *CHI*, (2012).
14. Haisley, E., Volpp, K., Pellathy, T., and Loewenstein, G. The Impact of Alternative Incentive Schemes on Completion of Health Risk Assessments. *American Journal of Health Promotion*, (2012).
15. Inciardi, J.A. In search of the class canon: a field study of professional pickpockets. In R.S. Weppner, ed., *Street Ethnography: Selected Studies of Crime and Drug Use in Natural Settings*. Sage, 1977.
16. Kahneman, D. and Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica*, (1979).
17. Kaplan, C.D., Korf, D., and Sterk, C. Temporal and social contexts of heroin-using populations. An illustration of the snowball sampling technique. *The Journal of nervous and mental disease 175*, 9 (1987).
18. Kumar, N. and Parikh, T.S. Mobiles, music, and materiality. *CHI*, (2013).
19. McNamara, R.P. *The Times Square Hustler: Male Prostitution in New York City*. Praeger, 1994.
20. Penrod, J., Preston, D.B., Cain, R.E., and Starks, M.T. A Discussion of Chain Referral As a Method of Sampling Hard-to-Reach Populations. *Journal of Transcultural Nursing 14*, 2 (2003), 100–107.
21. Pickard, G., Pan, W., Rahwan, I., et al. Time-Critical Social Mobilization. *Science*, (2011).
22. Sambasivan, N., Cutrell, E., and Toyama, K. ViralVCD: Tracing information-diffusion paths with low cost media in developing communities. *CHI*, (2010).
23. Shroff, G. and Kam, M. Towards a design model for women's empowerment in the developing world. *CHI*, ACM Press (2011).
24. Ulrich, C., Danis, M., Koziol, D., Garrett-Mayer, E., Hubbard, R., and Grady, C. Does It Pay to Pay? *Nursing Research*, (2005).
25. Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., and Mcspurren, K. Charities, no; Lotteries, no; Cash, yes. *Public Opinion Quarterly*, (1996).
26. Wyche, S.P., Smyth, T.N., Chetty, M., Aoki, P.M., and Grinter, R.E. Deliberate interactions: characterizing technology use in Nairobi, Kenya. *CHI*, (2010).
27. Zangeneh, Z., Barmaki, R., Gibson-Wood, H., Levitan, M., Romeo, R., and Bottoms, J. Research Compensation and Lottery: An Online Empirical Pilot Study. *Int J Ment Health Addiction*, (2008).