Estimating the Social Costs of Friendsourcing

Jeffrey M. Rzeszotarski[1,2], Meredith Ringel Morris[1]
[1]Microsoft Research, Redmond, WA, USA
[2]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA
jeffrz@cs.cmu.edu, merrie@microsoft.com

ABSTRACT
Every day users of social networking services ask their followers and friends millions of questions. These *friendsourced* questions not only provide informational benefits, but also may reinforce social bonds. However, there is a limit to how much a person may want to friendsource. They may be uncomfortable asking questions that are too private, might not want to expend others' time or effort, or may feel as though they have already accrued too many social debts. These perceived social costs limit the potential benefits of friendsourcing. In this paper we explore the perceived social costs of friendsourcing on Twitter via a monetary choice. We develop a model of how users value the attention and effort of their social network while friendsourcing, compare and contrast it with paid question answering in a crowdsourced labor market, and provide future design considerations for better supporting friendsourcing.

Author Keywords
Friendsourcing; crowdsourcing; SNS Q&A; Twitter

ACM Classification Keywords
H.5.m. Information interfaces and presentation: Misc.

INTRODUCTION

One does not have to look far on a social networking service (SNS) like Twitter or Facebook to see people *friendsourcing* information seeking by asking questions of their networks. Prior research by Morris et al. [16] has shown that over 50% of social network users self-report as having asked questions of their network, and commonly seek out recommendations, opinions, or factual information. These questions not only provide informational benefit when answered, but also provide social benefits to both questioner and answerer [16,17]. Further, questions are one of the more valued types of content on social networks [1].

However, people have also reported that they were not inclined to post or answer personal, religious, political, monetary, or health questions because of how they might be perceived [16]. This aligns with studies of self-censorship on Facebook [23,24] that show users actively monitoring the way they present themselves, suggesting that many categories of questions are discouraged by existing friendsourcing practices. Further, most questions that pass the self-censorship bar still remain unanswered [12,22].

Morris et al. also explored the transactional nature of SNS Q&A [16]. Questioners reported feeling like they "owed" anyone who answered the question, and answerers reported a similar transaction, either gaining future favor or repaying an obligation; such norms of reciprocity [28] are particularly important among certain demographics, such as users in Asia [31] or users with disabilities [3]. As people continue to ask questions of their social network, these costs may rise above a person's level of comfort, forcing them to either stop asking questions or delay [21].

In such cases, *perceived social costs* on the part of the questioner influence the benefits they are able to realize from friendsourcing. These costs include the questioner's valuation of friends' expending time, attention, and effort, the cost of future efforts needed for "repaying" answerers, as well as the possible impact on a questioner's curated online persona. As the costs rise, questioners may limit informational requests, diminishing the potential efficiencies of friendsourcing. A richer understanding of the factors influencing the likelihood of friendsourcing can benefit not only for users of social networks, but also designers of tools and systems that support friendsourcing.

In this paper we introduce an experimental methodology to investigate the perceived social costs of SNS question asking by assigning a monetary value to friendsourcing. We chose Twitter as the basis of our experiment; previous literature has demonstrated vibrant Twitter question asking behavior across a wide variety of users [12,18,20,22]. We contribute a quantitative and qualitative findings about the factors that influence users' valuation of the social costs of friendsourcing, including demographic characteristics, the number of questions previously asked, user's interest level in a given question, and audience-appropriateness.

Additionally, we investigate an alternative means of answering social questions that may be better suited to private or anonymous questions. Previous literature has investigated using paid crowdsourcing to answer social questions [12,19], finding not only that crowd labor can provide good answers, but also that the anonymity can make participants feel more comfortable. Using the monetary values from our experiment, we contribute a comparison that suggests that

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada

ACM 978-1-4503-2473-1/14/04…$15.00.
http://dx.doi.org/10.1145/2556288.2557181

for prices less than participants' valuations of social costs, one can get quality informational answers from the crowd.

Combining these two findings, there is an interesting opportunity for future friendsourcing technology based on modeling and adapting to users' perceived costs. For some questions where friends may have extra domain knowledge or necessary context, systems might reduce the barriers to entry and costs for friendsourcing them. Others that anyone may be able to answer could go to a crowd of ready and willing knowledge workers or a web search engine. A system might help a person anonymize or redirect questions that are deemed too costly or personal and are at risk for self-censorship. By combining the power of friendsourcing and crowdsourcing with an awareness of social and personal costs, technology might become even more helpful for answering important questions.

RELATED WORK

Morris et al. [16] conducted a survey about question asking behavior on Facebook and Twitter, finding that over 50% of users ask questions, and that they commonly seek out questions, most commonly seeking recommendations, opinions, or factual information; a more recent survey found that half of Facebook users, one-third of Twitter users, and one-quarter of Google+ users have engaged in friendsourced information-seeking [18]. In general, such questions seem to be well liked by followers and friends [1]. Cross-cultural studies have found similar social network Q&A behavior in Asia, where expectations of reciprocity were more explicit than in the U.S. [31].

Paul et al.'s work [22] complemented those survey studies by using a log-based approach, studying public "?" tweets manually labeled by Mechanical Turkers. They noted a preponderance of rhetorical questions, and discovered that the response rate for questions was surprisingly low [22]. A more recent, larger-scale study of Twitter logs [12] found that only 35% of "?" tweets that used hashtags indicative of information seeking received any replies. Teevan et al. [26] identified factors that influence response rate and time to social network questions, such as punctuation and phrasing.

Friendsourcing on social networks is closely related to the relationships of friendsourcers and their respondents as well as their potential audience as a whole. Users of Facebook who sought social support were appreciative of the network's affordances for broadcast, but needed to balance the social costs of privacy with their anticipated audience [29]. Indeed, some friendsourced questions may not have any informational utility at all, but purely exist for the purpose of satisfying emotional support [10]. While weak tie responses were perceived as more useful, askers' level of satisfaction did not align with answerers' tie strength [10]. However, on Facebook self-reported information-seeking behavior aligns with perceived bridging social capital [14].

On social networking services like Facebook and Twitter, activities like sending directed messages or consuming others' content can boost the connectedness between users and improve wellbeing [5,6]. In particular, social network information-seeking activity seems to relate to existing social capital, perhaps due to the potential mix of online and offline interaction [8].

However, interacting online, especially in the case of friendsourcing, may come with additional costs. When Brady et al. expanded a support system for vision-impaired users to social networks, allowing blind participants to ask their friends to help identify labels and read text, users felt resistance towards using the tool due to a desire for privacy, feelings of embarrassment, and a desire to avoid causing extra effort or using attention [3]. In general, psychologists have shown that people are not always willing to ask for help if it may incur costs of effort or make the responder think less of them [7,30]. Norms of reciprocity and fair trade may also make people less likely to incur future debts or make excessive use of friends' time or resources [28].

These costs are also evident when people have a choice between searching the web and friendsourcing. People are wary of spamming a person's social feed or demanding too many resources [17], and prefer search to friendsourcing for very concrete and very vague information needs [21]. However, friendsourced answers often contain personal or contextual information that improves their quality, though they take longer to receive and may not provide more information than a search [9,17].

Recent research has hybridized search and friendsourcing to mitigate costs and maximize benefits. SearchBuddies [11] were automated agents that responded to friendsourcing requests on Facebook with algorithmically generated content, including links to relevant Web resources and suggestions of Facebook friends with relevant expertise. MSR Answers detected questions on Twitter and used a crowd of microtask workers to answer them; answers were largely well written and arrived quickly [12]. Another potential benefit of asking an anonymous crowd over friendsourcing is in their lack of connection to the asker. They may be more frank, or less likely to judge an asker publicly based on their question. In a study that had people friendsource and crowdsource fashion decisions, participants appreciated the blatant honesty of the crowd, though friends knew more about their personality and activities [19].

EXPERIMENTAL DESIGN

To estimate the social costs of friendsourcing, we considered approaching the issue using canonical measures such as social capital. However, previous literature suggests that both bridging and bonding social capital may not relate strongly to favor asking on social networking services [13]. Further, such measures do not capture the broader set of behaviors that may indicate rising social costs. Not only might there be rising transactional changes in capital that make a user feel less inclined to "owe even more," but participants may also self-censor in order to maintain their

online persona [32] if a question's contents are sensitive, or they may empathize with others and actively try to avoid bothering them too much. To accommodate this broader set of possible outcomes, we chose to approach *social costs* in friendsourcing more generally.

While we choose to observe *social costs* in a general sense, there are several likely components of this construct with respect to friendsourcing. For example, as users post questions, friends expend time and attention noticing these questions. Should friends reply they might also expend significant effort to give an answer. A questioner may value these friend expenses as a social cost. Replies may come with costly expectations of future reciprocation (or the perceived presence of them). Participants in social networks also engage in personality maintenance [23]. Posting overly personal or too many questions may impact a users' managed persona, incurring more social costs.

One way to understand the intrinsic motivations associated with rising or falling social costs is to apply an opposing extrinsic motivation. Imagine a situation where a person earns five cents for every question they friendsource during the next hour. This may provide enough motivation to post, overcoming any social costs. However, if they only receive one cent per question, then the extrinsic reward may not sufficiently cover the social costs of friendsourcing. By varying payment across participants, we are able to assign a monetary value to the social costs of friendsourcing information seeking under specific circumstances.

In our experiment, we have participants make a series of 20 decisions with mounting social costs. Participants are asked to choose either to friendsource a question on Twitter or to sacrifice some of their potential study compensation in order to pose the question to a crowd of microtask workers. If a participant chooses to friendsource, then the question is scheduled to appear on their feed with a set delay. As the study continues, a queue of posts and sacrificed money builds. We vary the absolute amount of money that may be sacrificed as a between-subjects experimental condition, though it is always $1/20^{th}$ of the total gratuity. For example, a participant may pay $0.05 from their $1.00 bonus money to avoid friendsourcing a question, while another may be asked to pay $0.25 from their $5.00 bonus. This limits the influence to an absolute value difference rather than a difference relative to the total bonus.

After choosing to friendsource or crowdsource each of their selected questions, participants are debriefed (and paid their base gratuity plus the full bonus amount). Following the end of the experiment, our system automatically posted the questions marked for friendsourcing to the participant's Twitter feed spaced at eight-hour intervals, so that we could measure response rate and the quality of answers received.

Choosing a Social Network

There are numerous candidate social networks for this experiment, particularly Facebook, Twitter, and Google+ which are used for casual friendsourcing by a substantial proportion of their users (one-half, one-third, and one-quarter, respectively [18]). We wanted to select a social network that had as much of a *broadcast* feel as possible, the default behavior to be sharing, and, if possible, the outcomes of our participants' perceived choices to be public so as to maximize the potential social costs.

Facts / Opinions	Polls	A or B
#help	#justcurious	#decisions
#qtna	#curious	#pickone
#replytweet	#twitterpoll	#dilemma
#ineedanswers	#daremequestions	#thatisthequestion
#asktwitter	#ask_answer	#choices
#askingforafriend	#discuss	#wouldyourather
#questions	#openquestion	**Rhetorical**
#twoogle	#youropinion	#confused
#justasking	#whatif	#why
#question	**Recommendation**	#what
#seriousquestion	#needhelp	#thoughts
#randomquestion	#suggestions	#randomthought
#opinions	#helpme	#hmm
#lazyweb		**Requests**
		#please

Table 1: Hashtags of interest for question sampling based on Mechanical Turk ratings.

Facebook and Google+, both because of privacy features and the system's and participants' active selection of who can and cannot see content, are ill suited for this kind of participation. Facebook and Google+ also have a relatively slow pace of post and response that may change the impact of potential costs. In the case of Facebook, additional algorithms operate on Newsfeeds, only allowing subsets of the audience to see a given post. This makes it hard for users to estimate the size/composition of their audience [2].

In comparison, Twitter is generally open and fast-paced. As of August 2013, 18% of U.S. residents had accounts, and 30% of those aged 18-29 [4]. Because of the 140-character limit, posts are generally succinct. The barrier for replying is very low, since the length limit constrains the complexity of the answer. All tweets are potentially visible to all followers, as a "Newsfeed" style algorithm does not mediate them. The Twitter API allows for an app to pose as a user (with their permission) and post. For these reasons, Twitter seemed like an ideal candidate for our preliminary work in this area. The social (and monetary) costs of friendsourced information-seeking on other networks likely differs; comparing the impact of social network design is an area for future work; however, this paper focuses primarily on the costs associated with friendsourcing on Twitter.

Identifying and Sampling Canonical Questions

In order to avoid variations in the content or difficulty of the questions participants asked, we employed a set of general questions that did not require a high degree of personal context or domain knowledge. While such factors are

among the benefits of friendsourcing [17], for the purposes of a controlled experiment between many different people with varying social network compositions, leveling these factors is crucial. To avoid potential experimenter biases if we were to create the questions, we sought existing exemplar questions on Twitter using a grounded approach.

We sampled 61,286,532 public tweets that we suspected were questions from the Twitter Firehose during a one-week period between May 9th, 2013 and May 16th, 2013. We only collected tweets that ended with a question mark (ignoring hashtags), did not contain URLs, were from the English language region, and weren't replies or retweets. From that initial set we further restricted our criteria, selecting only tweets that started with question words; contained time zone information (a mark of an active user); and included hashtags. This left us with 5,261,440 tweets. Of these question tweets, 1,175,098 received at least 1 reply (a response rate of 22.3%). This seems to parallel existing log studies of question asking on Twitter [12,22]. We selected the 2,365 hashtags (out of 261,255 total hashtags) that had at least 50 tweets made to them over our week. We manually searched for tags that implied questions, ultimately producing a list of 40 hashtags to use for finding questions.

From these 40 candidate hashtags, we randomly sampled 30 questions each and had 3 different Mechanical Turk workers rate their quality, classify their content as factual/opinion/recommendation, friendship/favors, or rhetorical, and mark whether a stranger who doesn't know the author could conceivably answer the question. Table 1 shows a coding of the hashtags based on the worker ratings. Based on the Turkers' labels, we identified four common question categories common on Twitter: Factual, Personal, Product, and Recommendation. For each category we chose 15 tweets that required minimum context and could be answered with at most a web search. We corrected grammar and unified the hashtag topical markers to correspond to the four identified categories. Table 2 contains several example questions from each category.

Evaluating Friendsourcing Costs

To conduct the study, we recruited participants through general Twitter posts to a Twitter account dedicated to the study and through paid promoted tweets via Twitter Ads.

Condition	N	Total ?s Friend-sourced	Total ?s Crowdsourced
$0.05	31	416 – 70.7%	172 – 29.3%
$0.10	31	358 – 60.6%	233 – 39.4%
$0.15	30	415 – 70.0%	178 – 30.0%
$0.20	30	454 – 76.4%	140 – 23.6%
$0.25	31	540 – 90.0%	62 – 10.0%
$0.50	31	498 – 81.8%	111 – 18.2%

Table 3: Conditions and resulting decisions. Note that some points have been excluded based on poor write-ins.

We also recruited participants through internal organization email lists and Mechanical Turk tasks. Upon later analysis, the source of participants did not have a significant effect on our results. We implemented a prescreen system that used the Twitter API. We limited eligibility to U.S.-based participants aged 18 or older with had public Twitter accounts, at least 50 tweets in total, at least 10 tweets over the last 30 days, at least 20 friends, and at least 20 followers. This prescreening was meant to ensure that participants had enough Twitter activity to make posting several tweets during the study seem natural, and had a sufficient network size to incur potential social costs. Participants who passed were assigned conditions by round-robin (Table 3).

Our prescreen received roughly 4200 hits, of which approximately 750 actually approved our Twitter app. Of those, 397 passed the prescreen. 184 of those approved applicants successfully completed the entire study (1 from the listservs, 27 crowd workers, and 156 from Twitter). These limited response numbers are not surprising since we stated in the prescreen our Twitter app may be making posts to one's account during the study; several potential participants suggested that requirement was too onerous. Once participants passed the prescreen, they were directed to a SurveyGizmo page that contained the study. We implemented a system using Django and the Twitter API to schedule posts if a user chose to friendsource. We included a description of crowdsourcing markets so as to make sure participants knew what asking a crowd of workers entailed.

Experimental Conditions and Variables

We wanted to explore a variety of different payment conditions to help estimate social costs. We chose $0.05 as our lowest payment for crowdsourcing a question. From there

Factual	How many active Facebook users are there in Africa? #twoogle Can people in witness protection use social media? #twoogle What do you call the end piece on a loaf of bread? #twoogle
Personal	What do you guys do before a workout? #seriousquestion Is it acceptable to go to the Cinema on your own? #seriousquestion How do you tell someone you don't wanna talk anymore without being rude? #seriousquestion
Product	Do I need to read the books if I watch Game of Thrones? #asktwitter Do people prefer twitter on mobile or on laptop? #asktwitter Is paleo diet beneficial? Has everyone tried it? #asktwitter
Recommendation	Who makes good laptops these days? #suggestions What country's food should I try next? #suggestions Are there any TV shows like Breaking Bad I can watch while I wait for the new season? #suggestions

Table 2: Samples of the fifteen canonical questions for each category.

we added five cents for each condition, up to $0.25. To explore very high costs, we also added a $0.50 condition. Each condition had between 30 and 31 participants (Table 3). Participants received a base gratuity of $5 for completing the study, with a potential bonus of 20 times their condition's payment level (i.e., a maximum bonus of $1 in the five-cent condition, or $10 in the fifty-cent condition).

Choosing to friendsource created a queue of posts to be sent after concluding the study. We told participants that our app would space tweets out with 8 hour gaps, as that time seems short enough to evoke feelings of "bothering" someone while at the same time is not so often it could be conceived as pure spamming. Future studies may be able to explore a range of delay conditions.

For each question category we identified earlier (Factual, Personal, Product, Recommendation), we asked participants to choose 5 questions out of the 15 canonical questions per category (Table 2) to build up their set of 20 friendsource/crowdsource decisions. We asked participants to select questions that matched their voice and were interesting. They had the option to write their own if they did not find adequate questions. This sort of approach is also common in information retrieval studies of search queries, as a way to strike a balance between realistic, personally relevant prompts and prompts that can be controlled for and compared across a set of participants [25]. If a person chose to write-in for a particular question choice, we mark that the decision was the result of a write-in. We screened the write-in questions for quality, excluding from analysis any write-in questions that were obviously gibberish or unanswerable. Some example participant write-in questions include: "What song should I buy next on iTunes?" (Recommendation), and "What's your life goal?" (Personal). Participants chose (or wrote) their questions prior to learning that they would need to decide whether to friend- or crowdsource.

We also consider the previous behavior of a participant as a within-subjects variable. For example, if a person has chosen to friendsource twice and pay for the crowd once, then we record that their prior friendsourcing ratio was 0.66. This variable allows us to capture if a person is feeling as though they have "bothered" others enough. We chose to normalize the prior behavior because, depending on the condition, some higher friendsourcing rates are very sparse and we wanted disparate behaviors to remain comparable.

Hypotheses

We predict that payment condition will be positively related to choosing to friendsource. As the costs of turning to the crowd rise, they begin to eclipse the rising social costs.

H1: Higher crowdsourcing costs will result in increased reliance on friendsourcing.

We also predict that a higher friendsourcing ratio will correspond to a reduced likelihood of tweeting another question out of fear of accruing too many social debts.

H2: Higher friendsourcing ratios will reduce the likelihood of selecting friendsourcing for subsequent questions.

Alternatively, if bothering someone is not a serious concern for participants, they typically friendsource lots of questions, they view friendsourcing as relatively low cost, and/or they highly value the answers of their followers, then the relationship may even be positive, as the incremental social cost of tweeting more questions might decrease after a sufficient number have already been posted.

The decision to friendsource is likely to be mediated by a participant's interest in the question (e.g., the value the place upon getting this particular answer from friends). We suggest that as participants are more interested in getting an answer, they will be more likely to pose it to a known, trusted entity (their Twitter followers) rather than a paid crowd of anonymous laborers. Participants also may be likely to be more interested in questions in their own voice or ones that fit their online persona.

H3: Higher interest in a specific question will increase the likelihood of friendsourcing it.

H4: Write-in questions will be more likely to be friendsourced compared to canonical questions

User traits may also impact the valuation of social costs. Participants who tweet more often may be more likely to friendsource as the additional tweets might blend more naturally into their typical level of activity; their past use of the service might also represent a build-up of social capital (such as by responding to others' friendsourcing requests), which could lower the social cost to posting their own questions [5]. We also suspect that younger participants may be more likely to choose friendsourcing over paid crowdsourcing, both because their lower incomes might impact their economic valuation of social costs and because their network composition (likely more casual than professional) might entail less-complex audience-collapse-related [15] persona maintenance, thereby reducing the social costs.

H5: Younger age will relate to an increase in the likelihood of friendsourcing

H6: More pre-study tweets will relate to an increase in the likelihood of friendsourcing

RESULTS

184 participants completed the study. Overall, they generated 3680 different friendsource/crowdsource decisions. The majority of questions that participants chose were from our canonical lists; 416 (11%) were written in by participants. We had two independent coders screen the write-in questions for quality, giving each question a yes/no based on whether it was intelligible and could conceivably be answered by anyone. The coders evaluated the questions with a Cohen's kappa of 0.92. We discarded a decision if either coder thought the write-in question was unusable, causing us to discard 25% of the write-in questions, and leaving us

	Est.	SE	P(>|z|)
(Intercept)	-2.7020	.3050	<.0001
Dollars sacrificed to *crowdsource*	**1.5370**	**.4978**	**0.0020**
% past questions *friendsourced*	**4.6080**	**.1655**	**<0.001**
Interest in getting an answer	**0.2336**	**.0362**	**<0.001**
Question was a write-in : True	**0.5157**	**.2268**	**0.0230**
Self-report past question asking : T	0.3095	.1916	0.1061
Total tweets on Twitter	0.00001	.00001	0.0694
Number of friends on Twitter	-0.0001	.0001	0.3058
Number of followers on Twitter	0.00003	.0002	0.7727
Age in years	**-0.0179**	**.0072**	**0.0133**
Gender identity : M (F as baseline)	0.1993	.1338	0.1363

Table 4: Logistic regression coefficients and p-values (bold indicates p ≤ .01). Positive coefficients correspond to increased log-probabilities of *friendsourcing* rather than *paying*. We employed a mixed effects model to account for the repeated measures of participants.

with 313 valid write-in questions (9% of total choices) and 3577 total decision data points. 75.0% of these decisions were to friendsource rather than crowdsource. A significantly higher proportion of write-in questions were tweeted compared to canonical questions (84.0% vs. 74.1%, c^2=(1, N=3577)=14.52, p=0.0001), supporting *H4*.

The age (M: 26.16) and gender identity (M: 103, F: 81) distributions of our 184 participants are roughly in line with the natural population of Twitter [4]. Participants had made an average of 8441 total tweets (Med: 3290) at the time of our prescreen, following a roughly exponential distribution. They averaged 511 friends and 414 followers.

Most participants (86%) self-reported as having asked questions using Twitter previously (Y: 158, N: 26), though there may be some self-selection towards question askers in our study population. Participants reported typically tweeting questions daily (11.9%), weekly (34.7%), monthly (26.1%), even less often (13.0%), or not at all (14.1%). We asked participants to estimate the percentage of their Twitter followers that were friends and family, colleagues, or strangers. Participants reported a roughly even split between friends/family (M: 43.5) and strangers (M: 47.4), with a smaller percentage of followers being colleagues (M: 17.2). For each of the 3577 decisions, we had participants rate how interested or disinterested they were in receiving an answer on a seven-point Likert scale. Overall participants were mildly interested (M: 4.36).

Modeling Friendsource vs. Crowdsource

To understand how and why participants choose to friendsource versus crowdsource based on their internal estimated cost/benefit calculation, we modeled their choice. We developed a logistic regression model predicting for either friendsourcing a question to Twitter (positive log odds) or paying to send to the crowd (negative log odds). We used a mixed effects model to account for the repeated measure of a participant over multiple continuing decisions during the

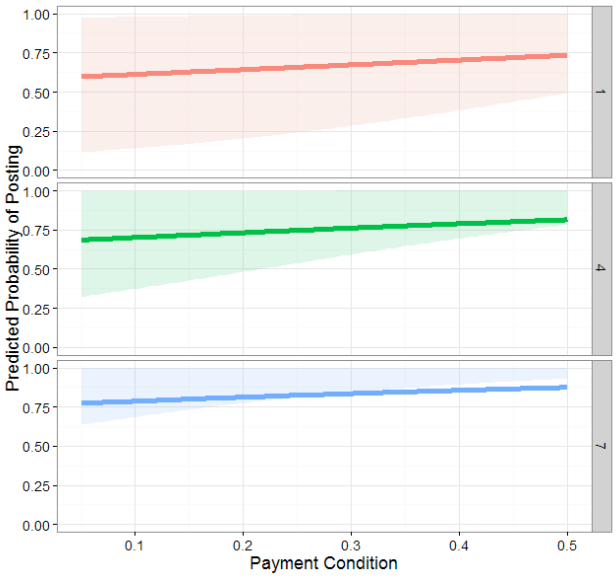

Figure 1: Mean predicted probability of choosing to *friendsource* over *crowdsource* based on payment conditions and level of interest (top = lowest interest, middle = neutral, bottom = highest interest). The bands correspond to quartiles.

study. Table 4 depicts the coefficients, errors, and significance estimates for the model.

As *H1* predicted, the amount of money a person had to sacrifice to choose the crowd influences their choice of friendsourcing or crowdsourcing. Even as social costs rise over several successive friendsourced questions, more money at risk makes a person resistant to sacrificing it. Yet, a portion of people still choose to sacrifice as, for them, the social costs are still too high no matter the bonus. A person's interest in getting an answer to the question also makes them more likely to friendsource, supporting *H3* (Figure 1). Interestingly, the prior percentage of questions a person friendsourced seems to increase the odds of future friendsourcing. This disproves *H2*, and may suggest that once people have friendsourced a few questions, they may consider it easier to keep going. Participants might not be as worried as previously thought about bothering others, or there may just be a general preference towards friendsourcing (or self-selection toward this preference in the group that completed our study). This is an area for future study.

Participants who had made more tweets prior to the study were marginally more likely to keep friendsourcing (p = .07), providing weak support for *H6*. Older participants were less likely to friendsource questions, supporting *H5*. We hypothesize that this may be due to different compositions of their expected audience. Participant age is significantly negatively correlated with the percentage of friends and family followers (r=-0.39, p<0.0001), positively correlated with the percentage of colleagues (r=0.20, p=0.0067), and positively correlated with the percentage of strangers (r=0.28, p<0.0001). Having more strangers or colleagues in a feed might imply higher social costs because of social presentation maintenance; the casual question categories

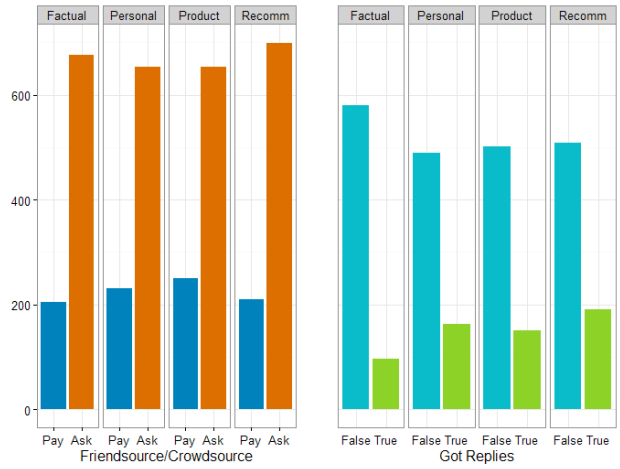

Figure 2: Count of Friendsource/Crowdsource decisions and of questions that Got Replies separated by question category

used in the experiment may also have been less suited for a professional network (or professionals may choose to friendsource questions on different platforms such as enterprise-specific networks [27]). Repeating the experiment with professionally oriented questions and/or higher payment conditions might illuminate which audience and economic factors impact older users' friendsourcing choices.

Question Characteristics and Replies

After all of the friendsourced questions had been tweeted, we polled the Twitter API for all replies to the questions. We located 1010 replies for 602 friendsourced questions, giving a response rate of 16.8%. 135 participants, 73.3%, received at least one reply (when asked at the end of the study, 25.9% did not expect to get any responses, which turned out to be an apt estimate). The median latency for receiving the first reply to a question was 695 seconds, or roughly 10 minutes (M: 7,230). Median reply length was 48 characters (M 54). Generally, participants received answers for more than one question (M: 3.29). As one might expect, the $\log_{10}$-transformed number of followers of participants who got replies (M: 2.336) were significantly higher ($t(88)=-2.459$, $p=0.016$) than those who did not get replies (M: 2.164). Participants who got replies (M: 25.12) were also significantly younger ($t(67)=2.13$, $p=0.037$) than those who did not (M: 29.01). There was a marginal relationship between total tweets posted prior to the study and receiving a reply, though log(number of followers) and log(total tweets) are significantly correlated ($r=0.62$, $p<0.001$). Write-in questions were also more likely to receive a response (15.9% vs. 26.5%, $c^2(1, N=3577)=22.24$, $p<0.001$), perhaps because they were tailored to a certain audience.

Each participant chose (or wrote in) five questions from each of the four categories (Factual, Personal, Product, Recommendation) to create their 20 total questions. There are slight differences in the number of decisions in each of the categories because of excluded write-in questions (Figure 2). In general, Factual and Recommendation questions were friendsourced marginally more often (76.7% and 76.9% rates) than Personal and Product questions (73.9% and 72.3%, $c^2(3,3577)=7.21$, $p=0.065$). However, Factual questions were the least likely to receive replies (14.2%), Personal and Product had roughly the same proportion of questions that got replies (25.0%, 23.1%), and Recommendation questions got slightly more replies (27.3%). These differences are significant ($c^2(3,2681)=38.5$, $p<0.0001$), suggesting that participants' followers were more interested in providing recommendations than answering factual questions that a participant could answer using a search engine.

Qualitative Response

Participants' qualitative responses reinforce some key aspects of the model. When people chose to friendsource, 57.6% professed a desire to save bonus money. 29.3% also cited wanting to know what followers had to say as their reason for friendsourcing. Of those that chose to sacrifice money, 50% cited concerns about bothering friends.

In our retrospective questions during the mandatory debrief, 60.3% of participants cited wanting to save bonus money as a goal of choosing friendsourcing, while 37% were more interested in what friends had to say than the crowd, and 26.1% didn't want to bother friends and followers. A majority of participants (70.2%) thought that it was acceptable to friendsource questions to Twitter at least once a day. Some cited Twitter as a "tool to utilize as often as necessary," while others explained "it gets annoying to some." They cited concerns about flooding and spamming, stating, "too many questions will annoy people" and "too much looks like an annoying spambot." One participant mentioned, "because twitter is a rapid conversation sort of thing… it's more acceptable to just shoot questions into blank space and see if you get a response." On the other hand, another participant stated "only rarely do I ask advice on Twitter, because it really isn't part of my Twitter 'persona.'" Another cited using search engines instead, saying, "I usually tweet about specific things. I don't ask global questions. I go to search engines to get answers…" These opposing perspectives illustrate that improvements in modeling social costs and the "economics" of friendsourcing might be made by personalizing models to account for individuals. Such costs (and personal preferences) may also change over time as norms of social network use evolve.

Evaluating Crowdsourcing Costs

While we have investigated potential costs of friendsourcing, we have not yet examined the alternative: the costs of crowdsourcing the same questions. Would crowd workers demand an amount of money comparable to what the participants were willing to sacrifice to avoid social costs? We sent the same batch of 60 canonical questions to crowd workers on Amazon's Mechanical Turk task market, seeking 5 different responses for each question. We replicated this process at the same time for each day of the workweek, increasing the payment per question each day in $0.01 in-

	1c	2c	3c	4c	5c
Log Duration	M:1.48 SD:0.39	M:1.81 SD:0.46	M:1.65 SD:0.58	M:1.77 SD:0.61	M:1.47 SD:0.46
Length of answer	M:56.0 SD:38.8	M:38.8 SD:35.2	M:53.5 SD:42.9	M:57.0 SD:45.7	M:57.9 SD:44.3
Rating (-2 to 2)	M:0.90 SD:0.84	M:0.80 SD:0.80	M:0.91 SD:0.83	M:0.92 SD:0.75	M:1.21 SD:0.68

Table 5: Means and standard deviations for crowd questions/answers by payment condition. Note that durations tend to be long because workers accept many tasks and then wait before resuming and completing them

crements from $0.01 to $0.05 per response. Workers could answer up to all 60 questions (of which 1 worker did, in general M: 9.6). Overall 156 workers provided 1500 responses (5 conditions * 5 replications * 60 questions). Examining the duration between posting the question-answering task and getting responses, the workers took a mean 8,765 seconds (2.4 hours) and median 6,046 seconds (1.7 hours), and a standard deviation of 9044 seconds. Some workers submitted responses in as little as 17 seconds from the time the job was posted. The median length of responses was 40 characters (M: 52).

We then had two new crowd workers rate each response on a scale of quality that ran from -2 to 2. -2 corresponded to blank, gibberish, or non-answers. -1 were responses that were intelligible but not answering the question. 0 were ones that were incorrect or incomplete. 1 were satisfactory answers, and 2 were excellent answers. We also had them decide whether or not they would be satisfied getting that response. We averaged the two raters' responses.

Across all questions the raters generally considered the crowd answers to be satisfactory (-2 to 2, M: 0.95). 1010 answers were deemed satisfactory, 354 split the raters' opinions, and 136 were deemed unsatisfactory. This aligns with the quality ratings, of which only 121 responses were rated below 0 quality. The answers that were at least satisfactory related slightly to higher payment amounts (M: 2.87c vs. M: 3.06c; $t(1049)=-2.53$, $p=0.011$). There were generally few differences between payment conditions (Table 5). The 5c condition received slightly higher ratings, and 2c responses had on average responses with 12 fewer characters. If we consider the payment condition as a continuous variable, there are no significant differences. In general, whether we paid the Turkers 1 cent per answer or 5 cents, we received decent, useful responses in a timely fashion. It is plausible that the differences we saw between payment conditions were just artifacts of worker self-selection and the population that was online that particular day of the week.

Comparing Costs

We also had another set of Mechanical Turk raters rate each one of the 1010 replies that friendsourcers received using the same methodology as for the crowdsourced questions. In general the quality ratings of friendsourced replies were slightly lower than crowdsourced replies (M: 0.64 for friendsourced replies vs. M: 0.95 for crowds, $t(1728)=7.72$, $p<0.0001$). 604 of the replies were deemed satisfactory, 250 split raters, and 156 were unsatisfactory (these rates are roughly concordant with the crowdsourced condition). The Twitter replies overall arrived faster than the crowdsourced replies, with a mean latency of 7,136 seconds later (2.0 hours), a median of 677 seconds (0.2 hours), and a standard deviation of 29,282 seconds. The time difference between our crowd workers and real Twitter responses is significant after log-transforming response latency to account for exponential distributions ($t(1652)=21.2$, $p<0.0001$). Some Twitter responses came in as soon as 3 seconds after posting. There was no significant length difference between Twitter and worker responses.

One possible reason for the difference in rated quality is that friendsourced answers may require information about social context to interpret properly. A remark that seemingly does not answer anything may in fact be more valuable to the friendsourcer than an informational response. Interestingly, no crowd worker submitted responses that contained jokes, while the raters noted that several of the tweeted responses were off-color or humorous. For example, one participant chose to friendsource the question "Is stretching more important before or after a workout?" The received the response "After. Stretch the truth to say you've done more than you actually have." This points to one specific advantage that friendsourced answers possess: they can make use of shared social context and previous rapport. While crowd workers are most certainly capable of social banter [12], their answers are usually intended for a general audience, though this can also be a benefit if the question asker purely needs information and is not also using friendsourcing as a means of social interaction [19].

Regardless, crowd workers provided slightly superior informational value compared to friendsourcing. Because they were paid, they delivered a 100% response rate.

DISCUSSION

Through our controlled study and regression analyses, we have explored several potential factors that affect users' decision to friendsource information seeking on Twitter. By varying payment, we were able to gain an understanding of how participants estimated the value of their friends' attention and effort, their own future reciprocal efforts, and/or their persona maintenance (teasing apart the valuation of these and other possible components of social costs is a rich area for further study). Participants were more likely to friendsource when provided higher monetary incentives. In the high payment conditions this amounted to a 10-20% increase in likelihood of friendsourcing. This is strongly mediated by several factors, including question content and interest in receiving an answer to a particular question.

However, several participants suggested that friendsourcing anything at all was too onerous. They opted to sacrifice all of their bonus money rather than post. This is a potential

limitation. While we controlled for past question asking behavior by incorporating a self-report into our model, we may not have explored the full spectrum of how people valuate the costs of friendsourcing. There may be other border cases such as those who never friendsource, or those whose social cost valuation is higher than the incentives provided in our study, and there may be some self-selection based on who chose to participant in the study. Some people who estimate friendsourcing to be very costly may not participate in a friendsourcing study. This may help explain why friendsourcing was the default behavior.

The fact that we opted to conduct a controlled study with monetary choices also presents limitations. Because we had participants make choices within an experimental environment, their expectations and social cost calculations may have been skewed. While we gained some specificity in terms of identifying the way people estimate social costs, it would be reinforced by future in-situ examinations of behavior. Observation bias might compel participants to ask more questions if they feel that is the more desirable behavior. The use of sacrifice rather than pure bonus in the experimental design is also a potential limitation; using a positive paradigm (earning rather than sacrificing money) may have shifted users' price points.

Similarly, because we used a synthetic situation, the timing and content of questions may not perfectly reflect actual friendsourcing behavior online. However, because we gave participants the ability to write their own questions and we selected participants who posted at least twice a day already, these factors were hopefully mitigated. While we controlled the rate at which questions would be posted as it is likely to influence the social cost calculation, in practice people regularly change their posting rate as a means to reduce social costs. Further, because the study ran over the course of two weeks, different social conditions across the Twitter network may have affected participants' likelihood of posting. In the future one might vary the rate of posting much like we varied price, looking for the way participants estimate the costs of "bothering" over different time spans. One could also introduce more strict control of question content, for example evaluating the costs of very general questions versus private health questions versus questions that may affect online personality maintenance.

Twitter also provides an inherent limitation to this work. Posting behavior on a particular social network is not fully generalizable to other social networking services. One significant area for future research would be to expand the scope of this methodology into different social networking services such as Facebook or Google+, which may exhibit different effects as a result of more reciprocal behavior, closer ties, different conventions regarding posting and replying, and potentially different audience compositions. Studying social network/forum hybrids such as Quora and StackOverflow may also provide interesting findings in this space. Regardless of platform, culture and differing social relationships (exacerbated by context collapse) might also affect social cost estimates.

Despite these limitations, our research points to several areas for future design and research. We already see in our analysis and in past work that people do choose not to friendsource and instead self-censor. This may because they don't want friends to expend too much effort, or may also be related to other factors such as the contents of questions. For example, a person may not want to friendsource a question about an embarrassing health problem, even if they have several friends who have medical training and could give advice. This points to a rich area for design of friendsourcing technology that reduces the perceived costs of asking. In the case of private information, we might be able to reduce costs by granting more anonymity or providing a use case in-between directed messaging and broadcasting to get at domain or trustworthy knowledge sources. We might be able to design networks that better expose the actual effort people go to in answering questions so as to make sure people are not off in their social cost estimations. Interestingly, social proof might actually exacerbate the problem. We already see participants feeling a push for reciprocity, and if we surface that someone is answering everyone's questions, friendsourcers may feel more in debt.

Our work with Mechanical Turk answerers and past work such as MSR Answers [12] suggest another direction for designing friendsourcing marketplaces – for some questions, it may be better to have an anonymous crowd ask them. We have shown that people estimate their friends' attention and effort at drastically higher values than it costs to get a comparable (or higher!) quality answer from crowd workers. This suggests benefits in developing approaches that hybridize friendsourcing and crowdsourcing. When a question might be too bothersome, we could provide a person the ability to ask a crowd with as little effort as posting. Participants even may be more comfortable asking private questions to strangers [19]. However, the crowd may not be able to answer all questions. The crowd is ill suited to answer those that require personal context or ones that might better be fulfilled with social rather than informational content. Intelligent agents could dynamically identify questions that are suited for different types of responders and assign them accordingly. This may even function within a social network, perhaps directing friendsourced questions to domain experts or those especially skilled at providing social support. As more and more people friendsource against finite attention and effort resources, this sort of intelligent question assignment will become increasingly important.

CONCLUSIONS

In this paper we developed a novel methodology for studying how social network users estimate the social costs of friendsourcing through a system of monetary choices. We employed this methodology to examine social question asking on Twitter, demonstrating that participants assign specific social costs to friendsourcing, and that assigning in-

creased monetary costs to avoiding friendsourcing makes participants more likely to friendsource despite the social costs. We demonstrated a link between question content, participants' desire for an answer, and age with the way participants estimated the social costs of friendsourcing.

We demonstrated that even a five-cent difference in monetary cost changes participants' question asking behavior, and parallel it with the costs of getting comparable answers from a crowd of workers. In general, even one cent can provide comparable or better informational quality as compared with Twitter replies, though with slightly increased latency. With these two findings in mind, there are rich areas for future research, including further examining social costs both within Twitter and in other social networking services, and also in designing new systems that might minimize friendsourcing social costs or hybridize social friendsourcing with microtask markets.

REFERENCES

1. André, P., Bernstein, M., and Luther, K. Who gives a tweet?: evaluating microblog content value. *CSCW '12*.
2. Bernstein, M., Bakshy, E., Burke, M., and Karrer, B. Quantifying the invisible audience in social networks. *CHI '13*.
3. Brady, E.L., Zhong, Y., Morris, M.R., and Bigham, J.P. Investigating the appropriateness of social network question asking as a resource for blind users. *CSCW '13*.
4. Brenner, J. and Smith, A. *Pew Internet Report: 72 % of Online Adults are Social Networking Site Users*. 2013.
5. Burke, M., Kraut, R., and Marlow, C. Social capital on Facebook: Differentiating uses and users. *CHI '11*.
6. Burke, M., Marlow, C., and Lento, T. Social network activity and social well-being. *CHI '10*.
7. DePaulo, B. and Fisher, J. The costs of asking for help. *Basic and Applied Social Psychology*, (1980).
8. Ellison, N.B., Steinfield, C., and Lampe, C. Connection strategies: Social capital implications of Facebook-enabled communication practices. *New Media & Society 13*,6 2011.
9. Evans, B.M., Kairam, S., and Pirolli, P. Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Information Processing & Management 46*, 6 (2010), 679–692.
10. Gray, R., Ellison, N., Vitak, J., and Lampe, C. Who wants to know?: question-asking and answering practices among facebook users. *CSCW '13*.
11. Hecht, B., Teevan, J., Morris, M.R., and Liebling, D. SearchBuddies: Bringing Search Engines into the Conversation. *ICWSM '12*.
12. Jeong, J., Morris, M.R., Teevan, J., and Liebling, D. A Crowd-Powered Socially Embedded Search Engine. *ICWSM '13*.
13. Jung, Y., Gray, R., Lampe, C., and Ellison, N. Favors from facebook friends: unpacking dimensions of social capital. *CHI '13*.
14. Lampe, C., Vitak, J., Gray, R., and Ellison, N. Perceptions of Facebook's value as an information source. *CHI '12*.
15. Marwick, A.E. and boyd, d. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society 13*, 1 (2010).
16. Morris, M.R., Teevan, J., and Panovich, K. What do people ask their social networks, and why?: a survey study of status message q&a behavior. *CHI '10*.
17. Morris, M.R., Teevan, J., and Panovich, K. A Comparison of Information Seeking Using Search Engines and Social Networks. *ICWSM '10*.
18. Morris, M.R. Collaborative search revisited. *CSCW '13*.
19. Morris, M.R., Inkpen, K., and Venolia, G. Remote Shopping Advice: Enhancing In-Store Shopping with Social Technologies. *CSCW '14*.
20. Nichols, J. and Kang, J-H. Asking questions of targeted strangers on social networks. *CSCW '12*.
21. Oeldorf-Hirsch, A., Hecht, B., Morris, M.R., Teevan, J., and Gergle, D. To Search or to Ask: The Routing of Information Needs between Traditional Search Engines and Social Networks. *CSCW '14*.
22. Paul, S., Hong, L., and Chi, E. Is Twitter a Good Place for Asking Questions? A Characterization Study. *ICWSM '11*.
23. Sleeper, M., Balebako, R., and Das, S. The post that wasn't: exploring self-censorship on facebook. *CSCW '13*.
24. Stutzman, F., Vitak, J., Ellison, N., Gray, R., and Lampe, C. Privacy in Interaction: Exploring Disclosure and Social Capital in Facebook. *ICWSM '12*.
25. Teevan, J., Dumais, S.T., and Horvitz, E. Potential for personalization. *ToCHI 17*, 1 (2010), 1–31.
26. Teevan, J., Morris, M.R., and Panovich, K. Factors affecting response quantity, quality, and speed for questions asked via social network status messages. *ICWSM '11*.
27. Thom, J., Helsley, S., and Matthews, T. What are you working on? status message q&a in an enterprise sns. *ECSCW '11*.
28. Uehara, E. Reciprocity reconsidered: Gouldner'smoral norm of reciprocity'and social support. *Journal of Social and Personal Relationships*, (1995).
29. Vitak, J. and Ellison, N.B. 'There's a network out there you might as well tap': Exploring the benefits of and barriers to exchanging informational and support-based resources on Facebook. *New Media & Society 15*, 2 (2012), 243–259.
30. Wills, T. Perceptual Consequences of Helping Another Person. *Annual Convention of the APA*, (1976).
31. Yang, J., Morris, M.R., and Teevan, J. Culture Matters: A Survey Study of Social Q&A Behavior. *ICWSM '11*.
32. Zhao, X., Salehi, N., and Naranjit, S. The many faces of Facebook: Experiencing social media as performance, exhibition, and personal archive. *CHI '13*.