mimir: A Market-Based Real-Time Question and Answer Service

Gary Hsieh
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
garyh@cs.cmu.edu

Scott Counts
Microsoft Research
One Microsoft Way
Redmond, Washington 98052
counts@microsoft.com

ABSTRACT

Community-based question and answer (Q&A) systems facilitate information exchange and enable the creation of reusable knowledge repositories. While these systems are growing in usage and are changing how people find and share information, current designs are inefficient, wasting the time and attention of their users. Furthermore, existing systems do not support signaling and screening of joking and non-serious questions. Coupling Q&A services with instant and text messaging for faster questions and answers may exacerbate these issues, causing Q&A services to incur high interruption costs on their users.

In this paper we present the design and evaluation of a market-based real-time Q&A system. We compared its use to a similar Q&A system without a market. We found that while markets can reduce wasted resources by reducing the number of less important questions and low quality answers, it may also reduce the socially conducive questions and usages that are vital to sustaining a Q&A community.

Author Keywords

Question and answer, Q&A, market, synchronous.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Question and answer (Q&A) services are designed to facilitate the transfer of information and expertise. Recently, such services have manifested themselves in the form of online community Q&A sites, where community members can post and respond to one another's questions. Examples include Yahoo! Answers (answers.yahoo.com) and Live QnA (qna.live.com), where users can post

questions and others can help answer them. These sites act as valuable alternatives to online searches and also generate knowledge repositories that can later be revisited.

Unfortunately, existing community Q&A services are not as efficient as they could be. One problem is the high percentage of non-serious and spam questions. When browsing through Q&A sites, visitors will notice questions that are not serious questions or do not make sense: "have you actually had ants in your pants," or "enter your question!" Potential answerers may spend valuable time and attention on these non-serious questions, missing out on more serious questions that really need answers. Another problem is that existing Q&A sites are not designed to balance the need of the asker with the availability of the answerer. This is especially problematic for real-time Q&A services such as Twitter Answers (www.mosio.com/twitter/), where question askers may want answers urgently but broadcasting the questions can result in costly interruptions for potential answerers.

Economic markets may provide solutions to these problems by facilitating the matching of resources based on the needs of the askers and the availability of the answerer. In an information exchange market, askers can offer to pay for the solution to their questions similar to the way they would purchase other goods and services. Answerers can then be compensated for their expertise, time and attention. This type of market offers three characteristics that are desirable in this domain. First, paying for help in economic markets can reduce spam and non-serious questions by forcing askers to be more selective in the questions they ask, as indicated by prior work on anti-spam [11,17]. Second, questions that are more important to askers, as signaled by a higher price, should receive more attention from potential helpers. Finally, monetary compensation can act as an additional incentive to motivate more participation from knowledgeable experts.

The idea of market based Q&A services is not new. In fact Google, with Google Answers, tried to create a Q&A market that ultimately was shut down in Dec. 2006 (answers.google.com). Possible reasons for shutting down the service include a decline in number of users, and a lack of real-time notification to the question asker when the question has been answered [4]. Despite this interest by a

major corporation in creating a Q&A market, how to successfully leverage economic markets in the online Q&A domain remains nebulous. In order for these markets to become successful, a number of research questions have to be answered. From an application standpoint, how do we design a usable market for real-time question and answer exchange? From a theoretical standpoint, how does having the market impact question and answering? Recent work has compared Google Answers to other Q&A sites [7], but the comparisons were across different Q&A sites with different characteristics. A more controlled setting is required to generate a better understanding of the effect of market mechanisms in this domain.

In our work, we designed and deployed two real-time Q&A services to 108 users, over the span of three weeks. One was a market-based real-time Q&A service. The second was similar, but had a simple reputation system instead of the market. First we compared how the services were used, then conducted a controlled study on question answering. Our results suggest that the market mechanism does impact question asking and answering. Askers and answerers are more selective in what they ask and answer in market-based Q&A systems, reducing the unnecessary attention and time costs. However, this selectivity reduced the amount of content in the system overall and consequently may have also reduced its users' sense of community.

Multiple contributions stem from these results. First, our study provides insight into how to design for a real-time, market-based Q&A service. To our knowledge, we have designed the first Q&A service that has combined payment markets with real-time Q&A. Second, our results provide theoretical insight into how incorporating a market mechanism impacts question and answering.

BACKGROUND AND RELATED WORK

In the past few years community-based Q&A sites have become increasing popular. Yahoo! Answers, one of the most popular sites of this kind, has more than 21 million users in the US and 120 million worldwide [12]. These sites offer an alternative to traditional search engines and act as communities where users can contribute and share expertise. Users can post a question to the site and other users in the community can respond and answer the posted questions. The generated knowledge is valuable as a knowledge repository. These community-based Q&A sites differ from digital reference services [14] or ask-an-expert services [9] in that community-based Q&A sites leverage the resources and expertise of its users, as opposed to a special group of experts or librarians.

Though community-based Q&A sites are similar in overall design, they differ in many subtle yet important ways. Most of these Q&A sites are asynchronous, using a website as the primary mode of interaction between members. However, others like Twitter Answers have expanded to include synchronous communication media, enabling real-time asking and answering of questions through instant messaging (IM) and short messaging service (SMS). In terms of incentivizing users, most sites have incorporated some form of a reputation-based system to promote top contributors in the community, while others use virtual tokens or even actual money to motivate users to contribute.

The specific focus of this work is at the intersection of real-time Q&A services and the use of economic markets. Real-time Q&A services such as Twitter Answers, Zephyr [1] or IBM Community Tools (ICT) leverage synchronous communication channels for question asking and answering. The synchronicity of these systems allows questions and answers to reach their users directly, leading to faster answers and faster updates on answers. However, as suggested by the study on the use of ICT, there are two potential problems that stem from using real-time Q&A systems: interruptions and inappropriate messages [18].

Economic markets can mitigate these problems with real-time Q&A services. Prior work on spam has shown that charging postages and taxes on the message senders can make the senders more selective about the messages they send out [11,17]. Applied to the Q&A domain, this could reduce spam and non-serious questions. Economic markets can also reduce potentially costly interruptions from receiving incoming questions. By creating a Q&A market, question askers can pay answerers for not just their knowledge, but also time and attention. Conceptually, this is no different than hiring tutors by the hour for their expertise and time. Costly interruptions can be avoided because the potential answerers can prioritize incoming requests based on the offered price. When answerers are busy, they can set a higher filter value, but when they are available, they can lower or set the filter value to 0.

While there has not been prior work examining the effects of markets on questions asked and interruption in synchronous Q&A services, there is a growing body of research studying the effects of economic incentives on question answering in asynchronous Q&A sites [3,7,15]. This research examined Google Answers, a fee-based Q&A site that combined a digital reference service with some community features, and showed that economic motivators are not the only factors influencing question answering behavior [15]. Non-monetary incentives such as ratings also partially account for answering behavior. More recent research comparing answer quality across Q&A sites has shown that fee-based Google Answers led to answer quality and responsiveness metrics superior to those of other Q&A sites such as Yahoo! Answers, leading to the conclusion of "you get what you pay for" [7]. However, due to the fact that the sites studied differed in many dimensions, including user population, size, and even the type of service offered, it is hard to make any specific claims regarding the effects of market mechanism on Q&A services.

Furthermore, research in psychology and experimental economics has suggested that disutility may result from monetary payments [e.g. 2]. Mainly, monetary incentives

tend to crowd-out intrinsic incentives (such as altruism or expected reciprocity) to provide help. Prior research has demonstrated the importance of social motivators in online contribution [13]. Introducing monetary incentives may undermine those motivators and reduce the overall amount and speed of contribution. Therefore, it is imperative and timely to explore the impact of market mechanisms on real-time information exchange. Can markets be used to support and improve real-time Q&A systems? How does the market affect the Q&A community as a whole?

ANALYSIS OF A QUESTION AND ANSWER SITE
We began by looking for potential areas for improvement in existing Q&A sites. We did so by analyzing an existing community-based Q&A site—Live QnA. Live QnA started in August of 2006 and to date has more than 290,000 total registered users, 600,000 questions and 1.8 million answers. For our analysis, we randomly sampled 200 questions from Live QnA. We analyzed the questions and answers on various measures to understand the efficiency of this Q&A system (Table 1). Note that in Live QnA, community voters determine the best answer to a given question, and that additional answers cannot be added to the question after the best answer has been selected. Question askers can vote as well, with their votes carrying the weight of three votes while other users' votes count only once.

	Live QnA
# Questions Analyzed	200
Average # of Answers	2.8 (SD=3.27)
% with answers	80%
% with best answers	56%
Average time until first answer (hh:mm:ss)	2:52:30
Average time until "best" answer (hh:mm:ss)	4:18:17
Average amount of waste	1.42

Table 1. Live QnA statistics

Prior work examining Q&A quality has analyzed answer quality based on the combined quality of all the answers, for a given question [7]. However, that level of analysis fails to consider the resources that are lost when answerers post answers that are not adequate answers. If answer number three is sufficient in answering the question, then the other answer posts prior to it may be considered wasteful. Hence, we created a measurement called waste (Table 1, bottom row), which measures the number of answer posts that were provided prior to the selected best answer. We are mindful, however, that there are cases where Q&A "wastes" are useful and valuable. In fact, the answers prior to a best answer may contribute to the process of reaching the best answer. Nonetheless, we posit that the

waste metric provides insight into the efficiency of existing Q&A systems.

Using the question type breakdown employed in prior work [7], we analyzed the composition of the questions being asked on Live QnA. For this, we recruited 9 coders to categorize each of the 200 questions as not a question, factual, advice, or opinion (inter-rater reliability α=0.89). Note, unless otherwise indicated, all coders used in this work were interns at a major technology corporation. As suspected, there is a fairly high percentage of submitted inquiries that are not real questions (13%, see Figure 1). We also recruited 8 coders to rate the seriousness of the questions asked, using a 5-point Likert scale where 1 is not serious, 3 is moderately serious and 5 is very serious (inter-rater reliability α=0.74). A serious question is defined as a question that you believe the question asker really wanted an answer. We used this to get a sense of the percentage of "real" questions, as opposed to joking or non-serious questions. We took the median across participant ratings to get the seriousness rating per question. The mean of these ratings was 2.63 (SD=0.76).

A one-way ANOVA indicated a significant difference for seriousness across the four question types. Post-hoc analyses showed that advice (M=2.93) and factual (M=2.78) questions were coded as significantly more serious than opinion (M=2.42) and non-questions (M = 2.13); all pairwise p's < .01, except the difference between advice and factual, and opinion and non-questions).
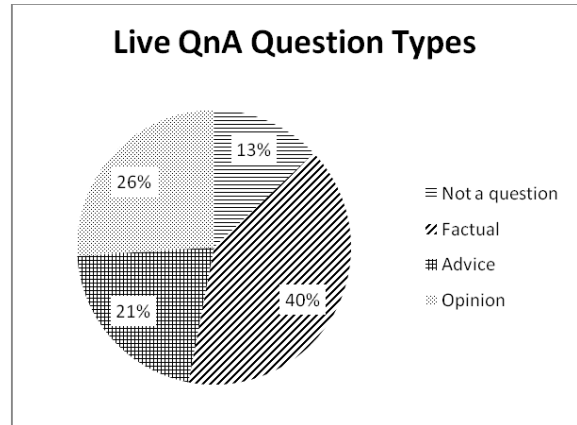

Figure 1. Breakdown of the types of question asked on Live QnA.

These results suggest room for improvement. Incorporating real-time question and answer notification may reduce the average time beween first and best answers. Moreover, utilizing market mechanisms may decrease the number of non and non-serious questions, thereby reducing the amount of answerer's resources wasted.

MIMIR: A REAL-TIME Q&A SERVICE
We implemented two versions of a real-time Q&A service called mimir. These were written as client applications, using C#. We chose to implement the service as client

applications as opposed to web 2.0 applications because doing so offered more control over the question and answer notifications in our comparative study.

The system is functionally similar to a chat broadcast system. In the client window (Figure 2), the user alias and the number of users who are online are displayed. When anyone has a question, they can then click on the "ask!" button, which will trigger a dialog window, within which they will be able to enter in their question. The question is then broadcast to the other users. By default, all incoming question will trigger a "toast" notification in the corner of the user's screen, similar to an incoming email notification. Potential answerers can then glance at the incoming question and decide whether or not to respond.
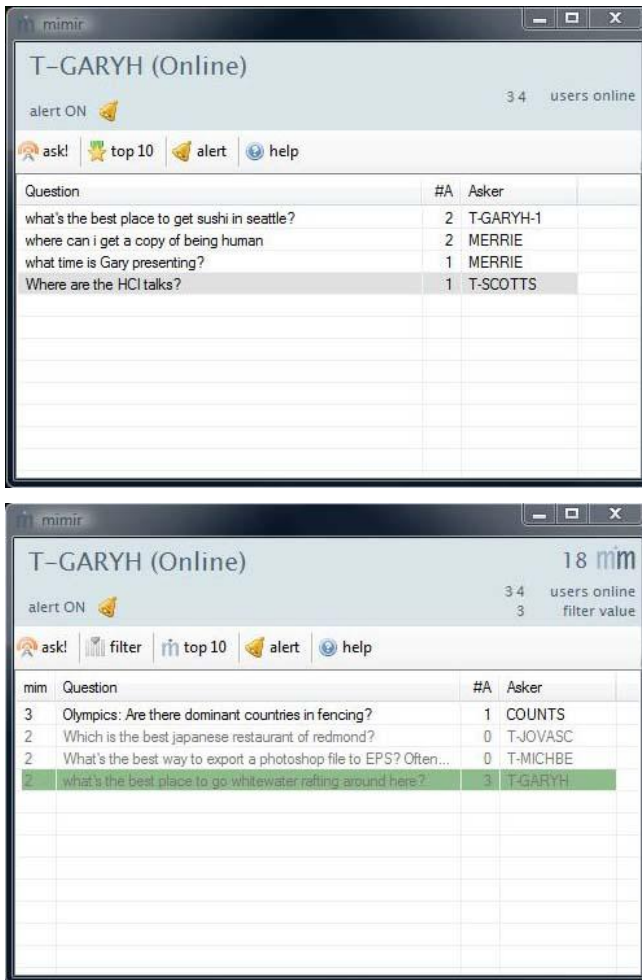

Figure 2. Client windows for mimir: (Top) No market system and (Bottom) Market system.

All questions are presented in a list view in the main window of the application (Figure 2). When a question in the list is clicked on, a separate answer window will appear to the right of the main window. In this answers window, users can read selected question and the list of answers in reverse chronological order (Figure 3). Each answer shows its respective answerer's alias and the time since it was posted. Question askers are allowed to "answer" their own questions. This allows the askers to clarify or to add additional information to the question. However, question askers are not allowed to select their own answers as the "best answer." In the version deployed, only the question asker can determine the best answer, but it is easy to imagine extending a collaborative voting design to this interface. Once a "best answer" is selected the question is marked as answered.

Questions are removed a day after the best answer is selected by the asker. If no best answer has been selected, the question is removed from the client three days after the initial post. Until the question is removed, more answers can still be posted to the question. We envision this client application integrating with a website so that others can search and browse through previous questions and answers.

Figure 3. mimir's answer window.

Baseline, No Market Version (*No Market System*)
The baseline comparison condition is a no market system with a simple top 10 list. The top 10 list displays the top 10 user aliases who provided the highest number of best answers. The list was viewable by clicking the "top 10" button in the main application window (see Figure 2, Top). We decided to include a simple top 10 list in our no market baseline system because most existing Q&A systems, such as Live QnA, have some sort of reputation system built-in.

Market-Based Version (*Market System*)
The experimental version was the market-based system, which used a currency called mims. When asking a question, the question asker attaches an escrow payment (in mims) to the question. By default, the payment amount is set to 0 mim (Figure 4).

Mimir users can filter incoming questions based on the number of mims offered. For example, users can set their mim filter value to 5, which will suppress notifications for questions less than 5 mims. They can reach the filter control by clicking on the "filter" button on the client window (see Figure 2, Bottom).

Figure 4. Market system's question asking window.

Prior work has shown that the overhead cognitive costs in using the market (e.g., having to think about what price to ask) may impact the efficiency of the market in a question and answer scenario [8]. To minimize such costs and to facilitate the decision process in mimir, the average filter price of the potential answerers is shown to the users in the question-asking dialog view (Figure 4). This way, question askers are provided a general sense of the "going rate" for answers in the Q&A market.

As with all economic markets, the Q&A market mechanism needs to be carefully designed to prevent misuse. Question askers may try to have their questions answered for free, while answers may try to get paid even though they did not provide an adequate answer. To mitigate misuse, we used the following design for our system:

1. An escrow payment is deducted from the asker's account as soon as the question is asked.

2. If after a day the question receives no answer, then the question is removed and the payment is fully refunded to the question asker.

3. If there are answers, the escrow payment is rewarded to the user who posted the best answer the instant the best answer was selected by the question asker.

4. If after three days none of the answers were selected as the best answer, then the question asker will receive a partial refund. 1-10 mim questions will incur a 1 mim "service" fee, 11-20 mim questions will incur a 2 mim fee, and so on.

This design allows question askers to decide if their questions have been answered, while also increasing the cost to questions askers who want to abuse the system.

HYPOTHESES

We make the following hypotheses regarding the comparison between the two types of Q&A systems.

First, we expect the escrow payment mechanism in the market system to reduce the number of non-serious questions asked, even though we allow users to select 0 mim payments. There are both social and monetary factors at play here. Social norms and potential social costs limit the number of questions asked with no payments offered. Monetary costs make the users more selective when asking questions. If it costs $1 to ask a question, then the askers should only ask questions whose answers are worth at least $1 to them. Otherwise, they will be "wasting" their money.

H1: The market system will lead to higher average seriousness in questions asked, but will result in fewer total number of questions asked compared to the no market system.

Due to the reduction in questions and the added filtering mechanism in the market system, we would also expect lower interruption costs in the market system.

H2: The market system will incur lower interruption costs on its users.

We also expect a combination of forces to influence question answering behavior. While economic incentives may be useful to motivate help, prior research has shown that it may also reduce and crowd out intrinsic motivation for help [5]. The gain in extrinsic motivation for answering in the market system may simply cancel out the loss in intrinsic motivation.

Prior work also suggests that the introduction of a market changes the perception of the existing interaction [6]. What was once a social interaction can turn into a market transaction. In our design, the no market version, with its simple reputation system, will be more of a community-based help/chat system like ICT or Zephyr. Meanwhile, the market will create a more serious questions and answers atmosphere. This change in perception can prevent answerers from answering when they do not know the answer and reduce waste. However, the reduction in low-quality answers would also lead to fewer responses per question and consequently slower answers.

H3: The no market system will have faster responses and more answers. In the market system, overall answer quality will be higher, and there will be less waste.

Cutting down on the non-serious questions and answers may also reduce the number of social and informal interactions. Prior work has highlighted the importance of informal communication in organizations, suggesting that it supports social relationships [10]. The self-filtering of silly but sometimes socially valuable questions and answers could weaken the sense of community in the market system.

H4: Users of the no market system will feel a stronger sense of community.

Within the market system we would expect market prices to be used as signals: question askers should vary their mim depending on the seriousness of the question. Recent work on attention markets for email communication showed that senders and receivers were able to use synthetic tokens as signals [16].

H5: More serious questions will have higher escrow payments.

DEPLOYMENT OF MIMIR

The mimir systems were deployed and used by interns at a major technology corporation (Microsoft). Users were recruited via email from the interns' mailing list, which had around one thousand individual interns' email addresses. The list of intern emails was randomly split in two: one group of interns received email invitations to try out the market system, the other the no market system. The invitation emails were largely identical, except the market system participants were told they would be given $5 worth of tokens (20 mims) to begin and that the amount of tokens they have at the end of the study will be converted into lunch coupons. The no market participants were told that they would receive a $5 lunch coupon at the end of the study for installing and running the system.

The usage period lasted for three weeks. To study overall usage and to analyze the questions and answers, we logged usage and content posted to mimir in our database. To explore users' perception of the system, at the end of the second week of the study, we conducted a questionnaire asking users to rate general Q&A quality along multiple dimensions. The questionnaire used a 5-point Likert scale, with 1 as strongly disagree, 3 as neither disagree nor agree, and 5 as strongly agree.

In addition to the community generated questions and answers, we also submitted a test-bed set of controlled questions to both systems in order to study answering behavior in a more controlled fashion. Out of 60 answered questions randomly sampled from Live QnA we selected 24 questions to ask—eight factual questions, eight advice questions and eight opinion questions. We posted the questions to each version of the system over the course of 5 days at the end of the study period. Five coders rated the questions on question difficulty (α=0.60), which was used to set the mim value in the market system. Unlike other coders used in this study, these coders were distant colleagues of the researchers. While the inter-rater reliability was not as high as we would have liked, it was not used directly to determine the question mim value. We had included a little bit of randomness—low difficulty questions were offered 0 or 1 mims, medium difficulty questions were offered 2 or 3 mims and high difficulty questions were offered 4 or 5 mims. We chose to partially tie mim values to difficulty rather than completely randomizing the mim values attached to the questions because it seemed more natural: if mim values were randomly assigned, some hard questions would have low

value while some easy questions would have high value, which would elicit unrealistic answering behavior.

In order to appear as though the test questions were coming from a variety of askers, we used ten different aliases of real interns who were not participating in the study to post the questions to the mimir systems (we asked for their permissions before posting the questions.) The questions were posted at random times during the day, with the same questions being posted to each system as the same time. As the question asker, we minimized the interaction with the two mimir systems. We did, however, decide that it would be important to occasionally select the best answers as a typical user would. Since it is possible that the time *when* a best answer is selected would impact the users' answering behavior, we selected the best answers in the two systems at the same time.

RESULTS

During the 3 weeks of use, 108 users installed our mimir system- 58 for the market system and 50 for the no market system. During this period, there were typically around 20 users on the market-based system and around 25 users in the no market system. Most of the users were between the ages of 18-22 and male (7 females total). 31 users from each condition completed the second-week questionnaire.

General Usage

There were 68 questions asked through the no market mimr system and 50 questions asked through the market mimir system, a non-significant difference. Questions asked covered a wide range of topics, from questions regarding the company and the internship, to restaurant recommendations and programming help.

H1-User-asked Questions

Our usage results indicated a non-significant trend toward more questions being asked in the no market system, but was there a difference in the types of questions asked? Mimir questions were coded for question type and seriousness by the same coders who rated the 200 Live QnA questions (nine for types and nine for seriousness).

Our results showed that the seriousness of the questions asked were not significantly different between the two conditions (market=3.23 and no market=3.36, t(116)=1.01, p=0.31). However, when we surveyed users, those in the market condition perceiving the questions asked to be of higher importance than the users of the no market system (Likert rating of 2.78 to 2.39, t(60)=1.72, p=0.04, 1-tailed).

Compared to Live QnA, we saw that besides a reduction in non-questions (Figure 5), the questions being asked through mimir in general were of higher seriousness (see Table 2, last row, F(2,315)=31.2, p<0.001). This could be due to a number of things. For example, users may be more careful with what they ask through mimir because intra-organizational deployment makes each individual user more accountable or because knowing the other users are fellow interns allows for better targeting of questions.

	No Market	*Market*	Live QnA
#questions	68	50	200
Avg. #answers	3.18 (SD=2.86)	2.14 (SD=1.6)	2.8 (SD=3.27)
% with answers	90%	86%	80%
% with best answers	52.5%	58%	56%
Avg. time till first answer (hh:mm:ss)	1:14:44	1:17:07	2:52:30
Avg. time till best answer (hh:mm:ss)	3:41:51	4:35:26	4:18:17
Avg. amount of waste	2.39	1.24	1.42
Avg. seriousness	3.36 (SD=0.70)	3.23 (SD=0.70)	2.63 (SD=0.76)

Table 2. Usage comparision between mimir and Live QnA

When comparing question types (Figure 5), there was a trend toward a difference in question types in the market condition versus the other conditions (Likelihood ratio X^2=11.7, p=0.07). Mainly, the market system had significantly more factual questions. The increase in percentage factual questions corresponds to a decrease in percentage for non-questions and opinion questions. Interestingly, these are the two lowest categories in regards to question seriousness. Regarding hypothesis 1, we saw trends in the expected direction for usage, and user perception in line with our prediction. Secondary metrics like question type supported our hypothesis.
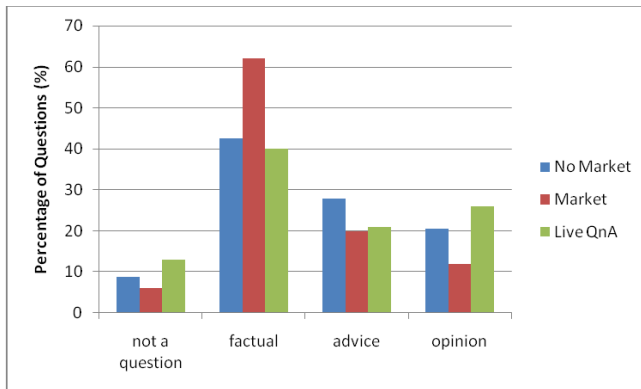

Figure 5. Comparison of types of quesitons asked across systems

H2-Interruptions from Real-Time Q&A
In the market condition, usage logs indicated 14 users who adjusted their filter value during the study. However, there was no significant difference in self-reported level of interruption between conditions (market=3.2, no market=3.5, where higher values means higher agreement that popups interrupted work). In post usage interviews, users were not overwhelmed by the interruptions that occurred during the study period, although they did complain about a bug in mimir that sometimes caused the answer window to gain topmost focus at random times. More usage is needed to fully explore the use of filtering mechanism to reduce interruption cost. Our current results do not support hypothesis 2.

H3- Question Answering
Our usage results indicated that the no market system did have significantly more answers per question than the market system (3.18 to 2.14, t(114)=2.28, p=0.02). But the no market system also had significantly more waste per question when compared to the market system (2.4 to 1.2, t(102)=2.47, p=0.02). However, was there a difference in the quality of answers? To systematically test for this, we compare the results of the 24 test-bed questions we asked through mimir.

Only one of the 24 questions in the no market system did not receive any answers, compared to 5 questions in the market system. On average, the no market condition had 3.1 answers, compared to 1.8 in the market condition and 4.1 in Live QnA. In terms of speed, the no market system seemed to be the fastest in getting the first answer. While the measured speed of first response was not significantly different across conditions, users in the no market system perceived the questions to be answered more promptly (3.57 to 3.17, t(59)=1.93, p=0.03, 1-tailed).

To address answer quality, eight individual coders were presented with the 24 test-bed questions and all the corresponding answers that were given on Live QnA, and the no market, and market systems in a random order across conditions. The coders were asked to rate the answer as either not an answer, or to use a 1 to 5 Likert scale, where 1 is extremely poor, 3 is average and 5 is excellent. We removed the answers that were not answers, and then took the median from the coder's ratings and compared the question quality across conditions. Inter-rater reliability was very good (α=0.90).

	NM	*M*	*Live*
Not an answer	7	1	2
1 - Extremely poor	11	2	6
2 - Below average	9	2	18
3 - Average	27	14	35
4 - Above average	18	23	33
5 - Excellent	3	3	5

Table 3. Answer quality distribution across systems

Our results showed that there was a significant difference in answer quality rating (F(2,207)=4.64, p=0.01). Specifically, the answer quality in the market (M=3.52) was significantly

higher than that of the no market system (M=2.93; F(1,207)=11.75, p=0.001).When we look at the raw breakdown of answers by quality level, we see that the market led to higher overall quality not because the market system led to more above average and excellent quality answers, but rather because it minimized the low-quality answers (table 3). Similar to what we observed in question asking, answerers in the market system seem to be more selective when answering.

This point can be further seen by examining the questions that had an answer in the no market system, but none in the market system. One example, taken from our 24 test bed questions, is "Why didn't Daniel Negreanu play very well in 2005?" Daniel Negreanu is a professional poker player. To answer this question, one would either (1) spend some time researching or (2) actually follow the professional poker tours and specifically, Daniel Negreanu's game. In our market condition, no answers were given. However, in the no market condition, within two minutes an answer, the only answer to the question, was received. The answer, "lol donkaments," was clearly not helpful and was rated by our coders as not an answer. This goes to show that while questions may be responded to more quickly in the no market condition, it did not necessarily help with question and answering. Another example "what is your favorite drink at starbucks?" While that question received 5 different answers in the no market condition, it received no answers in the market condition. The question asker may be genuinely interested in the answers to the question, but for various reasons, it was ignored in the market condition.

H4-Community and General Perception

Besides analyzing question and answer efficiency, we also explored how the two conditions differed in terms of the sense of community.

When asked about sense of belonging in the mimir community, no market system users felt that they belonged to their mimir community more than the market system users (3.04 to 2.49, t(60)=2.26, p=0.03). Furthermore, no market system users are more willing to keep using their Q&A system (3.52 to 3.04, t(60)=2.16, p=0.03) and they enjoyed using mimir more (3.49 to 3.17, t(60)=1.80, p=0.08).

H5-Usage of Market Price

Within the market system, there was no significant correlation between the number of mims offered for the user-asked questions and the coder-rated question type or seriousness. While our hypothesis was not confirmed, our results do not mean market prices cannot be used as signals in this domain. One reason why we did not observe any significant effects in our study may be because more extended use is needed for users to fully leverage the benefits of signaling and screening offered. It could also mean that the seriousness of a question is not the only factor in determining how valuable getting an answer is to the question asker. More data and longer use is needed to understand how price is used in a Q&A market.

DISCUSSION

Our work aims to understand the effects of incorporating a market in a Q&A system and to provide design insights on how to build a real-time Q&A market. Below, we will discuss findings for these goals separately.

Effect of a Market on Q&A

Our results suggest that the market and the no market systems provided different types of value to Q&A. The market system, as we had hypothesized, was better able to screen out non-serious and non-important questions. It also was able to reduce the number of low quality answers. Overall, it minimized waste. The no market system, on the other hand, had more usage: more questions and more answers per question. Furthermore, users were more connected with their mimir community, enjoyed their system more and were more willing to keep using mimir.

Given that our users did not appear to use the market to signal value and filter questions, the main difference between the two systems is that having the market made the users more selective in the questions they ask and the answers they post. When there is no currency involved, the Q&A system feels like a community-based help and chat program. Socially conducive questions are asked along with more serious, work related questions. Following is an example of such asked in the no market system:

What group are you working on & what's your project? I'm curious to see what everyone is upto, and also wondering if I might like doing what you're doing!

One user in the no market condition also compared their mimir system to the "internz" mailing list, which was used often by some interns for Q&A purposes:

I'm not much of an internz poster, and I guess this [mimir, no market] was a bit easier and a little more anonymous even though the alias is there. For some reason, it didn't feel more uncomfortable—just more casual. They can be silly questions too, and I would ask questions for the sake of asking question sometimes just to see what people would say.

This type of usage, while potentially wasting users' resources by overloading them with questions, can be important for fostering a sense of community.

By introducing monetary or token-based payments, Q&A becomes a more serious question and answering service. It becomes less like asking a friend, and more like paying for professional help and this changes how the system is used. Even though question askers could ask 0 mim questions, most of them did not do so. Instead they focused on more serious work related questions and on factual questions that are more likely to be answered. As one user said, "it [the market] makes me not want to ask stupid questions." This difference in perception of the system also impacted how users answered questions. Users in both groups asked the question of "Who is online right now?" The market condition question received a response letting the question

asker know that there is a number of users online count visible in the main mimir window. The no market system, on the other hand, resulted in a roll call, with 11 users reporting they are currently online.

Research in experimental economics has suggested this notion of changing the perception of the interaction due to the introduction of the market mechanism [6]. Specifically, when introducing a fine on parents for showing up late to pick up their kids at the daycare, the number of late-coming parents actually increased, as opposed to decreased. The explanation was that parents started weighing the late pick-up cost against the value of other things they could be doing. Thus a market may change what was once a social situation into a market exchange.

Picking which system to use for a Q&A service will then depend on the goal of the service. The market-based system is useful for balancing question and answer quality with attention and time costs and can lead to a more useful knowledge repository. A no market system, on the other hand, may better foster community.

Building a Real-Time Q&A Market
Just how can a designer build a successful real-time market then? Using users' feedback, we will discuss how to improve our current design focusing on the two main components of the real-time Q&A market— synchronicity and the design of the market.

Synchronicity
One dimension in which mimir differs from most existing Q&A sites is the use of the synchronous communication channel. It is interesting to note that even with far fewer users, mimir was able to have answer speeds that are comparable to an existing Q&A site. While there may be many factors at play here, such as our specific user pool, there is little doubt that synchronicity enables faster exchanges between questions askers and answerers.

One of our users in the no market version of mimir also happens to be an active user of Yahoo! Answers. He uses Yahoo! Answers to search through existing questions, but does not ask questions because "I don't want to wait a day or a week and then keep checking back." Comparing that to the synchronicity of mimir, he said:

I like that a lot…the fact that it pings you as soon as it gets an answer that's really helpful...it's a lot faster if it just pops up on your screen as a notification as opposed to sending it to your email and then you have to check the website.

Unfortunately, we were not able to stress test the use of market to explore potential interruption problems when the amount of synchronous notification increases and the pop-up notifications potentially become distracting.

In addition to using the market for filtering, users suggested the use of categorization to help improve the real-time targeting of questions. Most existing Q&A sites already have some sort of keyword tagging or categorization in place that could be leveraged. Coupling it with the market would allow users to utilize both topical domain information and question value when posting questions and deciding which questions to answer. Questions can then be targeted to a subset of "domain" experts by asking a high price question within a given topic area.

Market Design
Users in the market system thought the market was an interesting addition to a Q&A system and that it could provide added incentives. It was not clear if market offered more incentives than a good reputation system, but perhaps it provided incentives to a different group of users who are extrinsically motivated.

A number of questions arose regarding the payment mechanism. For example, several participants asked whether it was possible to split the offered payment to award multiple answerers. This is a natural tendency given that many questions elicit more than one worthy answer. In fact, during the study, one user actually asked an additional identical question after receiving answers just so he could pay two answerers who both provided good answers. Another user suggested allowing other users to add additional payments to an existing unanswered question if they also have the same question. Still others have suggested allowing users to place time-based payments. For example, if the question is answered within a certain time, the user will pay more. We did not have time to explore every payment design, but these are promising variations to explore in future work.

Given our findings that markets negatively impact the Q&A community, what can we do? One possibility is to allow two different classes of questions in a Q&A system. One type can be opinion questions, while another can be factual and advice questions that will require an escrow payment. This differentiation can allow the socially conducive questions to co-exist with more serious questions that are important and urgent. Reputation can be created for people in both groups: the best opinion posters and the best question answerers. Users who are only interested in one kinds of usage can do so easily.

Another possibility is to remove the monetary incentives altogether and instead use tokens without money or extrinsic goods attached to them. This could work well within an organization where the users can be held accountable for their actions. In that case, the token market can still be used for signals and filtering as social and corporate norms will prevent individuals from exploiting the token system. Earning tokens can be viewed as a fun intra-corporate activity, as opposed to a professional one.

Limitations and Generalizability
In the results section, we have presented Live QnA data alongside the mimir data mainly as a point of reference. It is important to note that due to the use of a fairly limited subject population (male-dominated, tech-savvy interns

who are mostly in their twenties) and study duration (3 weeks), there may be subtle nuances that limit the comparisons we can draw between our mimir systems and existing commercial systems, such as Live QnA. In particular, the larger scale of commercial systems may influence the impact of a market. For example, the sense of community may not decrease given the sheer number of users. However, given that typical Q&A sessions are fairly short, we were able to see many instances of completed Q&A sessions. Hence we do expect our findings comparing the experimental conditions to hold with a longer deployment.

CONCLUSION AND FUTURE WORK

In this work, we designed a real-time Q&A system that utilized an economic market. Through our deployment study, we were able to identify ways in which a market Q&A system differs from a no-market Q&A system. The main take-away is that the market reduced overall Q&A waste, which consequently may have reduced the users' sense of community and enjoyment form using the system. Results suggest that this was due to use perception that the market altered the once social and community-based interactions to a more serious Q&A service.

Much research is left in order to fully leverage the potential benefits of the economic market in this Q&A domain. How can we improve the usage of the pricing and screening mechanism? Here we look to the aforementioned variations on the pricing scheme (e.g., time-based sliding scales) suggested by participants. Given the benefits of both the market (efficiency, content quality) and no market (community) Q&A systems, perhaps the most important issue going forward is to determine how to better design a Q&A service so that social and monetary motivators can complement one another?

ACKNOWLEDGMENTS

REFERENCES

1. Ackerman, M. S. and Palen, L. 1996. The Zephyr Help Instance: promoting ongoing activity in a CSCW system. In *Proceedings of the SIGCHI*, pp. 268-275.

2. Deci, E. L., Koestner, R., & Ryan, R. M. 1999. A meta-analytic review of experiments examining the effects of *extrinsic rewards on intr*insic motivation. *Psychological Bulletin*, 125, pp. 627-668.

3. Edelman, B. Earnings and Ratings at Google Answers. 2004. *Unpublished Manuscript*.

4. Google questions *need for Google Answe*rs http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9005481&intsrc=hm_list

5. Gneezy, U. and Rustichini, A., 2000. Pay Enough Or Don'T Pay At All. The *Quarterly Journal of Economics*, MIT Press, vol. 115(3), pp. 791-810.

6. Gneezy, U., and Rustichini, A. A Fine is a Price. 2000 *Journal of Legal Studies*, vol. XXIX, 1, part 1, pp. 1-18.

7. Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. 2008. Predictors of answer quality in online Q&A sites. In *Proceeding of the SIGCHI*, pp. 865-874.

8. Hsieh, G., Kraut, R., Hudson, S.E., Weber, R. 2008. Can Markets Help?: Applying Market Mechanisms to Improve Synchronous Communication. In *Proceeding of CSCW*. pp. 535-544.

9. Janes, J., Hill, C., and Rolfe, A. Ask-an-Expert Services Analysis. 2001. *Journal of the American Society for Information Science and Technology* 52, 13 pp. 1106-1121.

10. Kraut, R. E., Fish, R., Root, R., & Chalfonte, B. 1990. Informal communication in organizations: Form, function, and technology. In *Human reactions to technology: Claremont symposium on applied social psychology*, pp. 145-199.

11. Kraut, R. E., Sunder, S., Telang, R., & Morris, J. 2005. Pricing electronic mail to solve the problem of spam. *Human Computer Interaction, 20, pp. 195-223.*

12. Leibenluft, J. 2007. A Librarian's Worst Nightmare: Yahoo! Answers, where 120 million users can be wrong. *Slate Magazine*, Dec. 7, 2007.

13. Ling, K., et al. 2005. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication, 10*(4), article 10.

14. Pomerantz, J., Nicholson, S., Belanger, Y., Lankes, R. D. 2004. The *Current State of Digital Reference. Information Processing and Management* 40, 2, pp. 347-363.

15. Rafaeli, S., Raban, D., Ravid, G. 2007. How Social Motivation Enhances Economic Activity and Incentives in the Google Answers Knowledge Sharing Market. *International Journal of Knowledge and Learning* 3, 1, pp. 1-11.

16. Reeves, B., Roy, S., Gorman, B., Morley, T. 2008. A Marketplace for Attention: Responses to Synthetic Currency Used to Signal Information Importance in E-mail. *Peer-Reviewed Journal On the Internet*. First Monday, Volume 13 Number 5 - 5 May 2008.

17. van Zandt, T. (2004). Information Overload in a Network of Targeted Communication. *RAND Journal of Economics*, 35 (3), 542-560.

18. Weisz, J. D., Erickson, T., Kellogg, W. A. 2006. Synchronous broadcast messaging: the use of ICT. In *Proceedings of SIGCHI*, pp. 1293-1302.