Social Preferences? Google Answers![1]

Tobias Regner
Friedrich-Schiller-Universität Jena
tobias.regner@uni-jena.de

December 2006
JEL classifications: C24, C70, C93, D82, L86
keywords: social preferences, reciprocity, moral hazard, reputation, Internet, psychological game theory

Abstract:

We study the pricing and tipping behaviour in the online service 'Google Answers'. While users set a price for the answer to their question ex ante, they can additionally give a tip to the researcher ex post. The obtained data set is analysed and compared to the results of similar games conducted in the laboratory, namely Fehr, Gächter and Kirchsteiger (1997) and Gächter and Falk (2002). Reciprocal theories of social preferences pioneered by Rabin (1993) and extended by Dufwenberg and Kirchsteiger (2004) are useful to explain the observed pattern of behaviour.

In line with the related experimental literature we conclude that an open contracts design encourages people to tip. We find evidence that this is motivated by reciprocity, but also by reputation concerns among frequent users. Moreover, researchers seem to adjust their effort based on the user's previous tipping behaviour. An efficient sorting takes place when a sufficient tip history is available. Users known for tipping in the past receive higher effort answers, while users with an established "bad" reputation for tipping tend to get low effort answers.

[1] I am grateful to Maija Halonen-Akatwijuka and Sebastien Mitraille for valuable discussions and to seminar participants at the University of Bristol, the Royal Economic Society Annual Congress 2005, the World Congress of the Econometric Society 2005, the Max-Planck Institute Jena summer school 2005 - in particular to David Winter, Jürgen Bracht and Osiris Parcero - for their comments.

1 Introduction

While other-regarding behaviour of individuals has been found in numerous lab experiments, it is not too clear yet what the precise drivers of socially-minded behaviour are and whether they also pertain in real-life environments.

The experimental evidence of individuals who consistently make voluntary payments has been explained by theories that take the psychological underpinnings of economic behaviour better into account, namely social preferences.[2] However, the external validity of the lab results is far less studied and merits more attention. Can we observe the behaviour found in the lab as well in real-life contexts and what are the underlying motivations of the occurring voluntary payments?

We collected field data about the pricing and tipping behaviour of "Google Answers" users in order to shed more light on these aspects. In this online service (a sub-service of Google) users can post questions and set a fixed price for the answer. They can also give a tip to the researcher who answered the question. Our new data set contains 6,853 observations (collected between July 2003 and January 2004) with a number of additional explanatory variables thanks to the online availability of past questions of the service. The service started in April 2002 and the average price for one answer is about $20. This rich data set puts us in a position to test the relevance of social preferences in a real-life environment instead of observing behaviour through the lab microscope.

The goal of the paper is to analyse the pricing and tipping behaviour in this non-laboratory test-bed in order to validate the results of related lab experiments. In particular we focus on (i) the underlying cause for the voluntary payments and (ii) the effects of such a design on effort levels and efficiency. We present three possible motivations for the tipping of users and we will test empirically to what extent they drive the behaviour of Google Answers users. First, the tipping could be to conform to a social norm as it is the case in restaurants, for instance. Social preferences could motivate users to leave a tip. Finally, users may decide to tip out of strategic considerations, e.g. reputation.

Social dilemma situations have been analysed in numerous experiments. The Google Answers environment is very similar to the particular setting of Fehr, Gächter and Kirchsteiger (1997). They study labour relations between firms and workers. When mutual opportunities to reciprocate are given (firms can reward or punish the worker ex post), higher effort levels than under stricter contract options are reached. They also find a significant positive correlation between workers' effort and the firms' reaction (reward or punishment). Based on Rabin (1993) they explain the observed behaviour with reciprocity concerns. We follow this approach also taking into account the theory of sequential reciprocity of Dufwenberg and Kirchsteiger (2004). Besides reciprocity frequent users may also be motivated by reputation to leave a tip in Google Answers. Experiments about interaction effects between reciprocity and reputation have been conducted in Gächter and Falk (2002) and we refer to them in our analysis.

[2] See Camerer (2003) and Fehr and Schmidt (2003).

Our real-life findings largely confirm the results of these experiments. About 25% of all answers have been tipped and the three main results of our empirical analysis follow. The more questions users ask over time the more likely they are to tip. Moreover, even single users tip. Around 18% of these observations are tipped. This is a lower rate than for frequent users, but it is still significant. Finally, tipping seems to pay off. Our data confirms that researchers take the past tipping behaviour of users into account and put more effort into the answer, if the user has frequently tipped before. The higher effort increases the benefit of the user and the researcher gets fairly compensated for the extra effort.

In the following section we describe the pitch of our field study - the online service Google Answers. Section 3 presents the related experimental and theoretical literature. Section 4 describes our data set, while section 5 analyses it. The conclusions are in section 6.

2 The Online Service Google Answers

Google is the most popular search engine and an essential tool to find information online. However, Google offers more than its standard search tool as sometimes even experienced Internet users need help finding exactly the answer they want to a question.[3] The service Google Answers (www.answers.google.com) provides help to these Internet users. It offers assistance from researchers with expertise in online searching.

Google Answers users ask questions and Google Answers researchers try to answer them in return for a fixed price and a possible tip. After registering with the service users can post a question to Google Answers and specify how much they are willing to pay for an answer. Users can price their question anywhere between $2 and $200. In addition a non-refundable listing fee of $0.50 applies for each question. There is a pool of roughly 400 Google Answers researchers who have the possibility to answer. Once one of them decides to search for an answer, a question will get 'locked' (for 4 hours if the price is below $100, for 8 hours if above). This means a question is actively worked on by a researcher and no other researcher can answer it in that time. The researcher will try to obtain the requested information and will post his answer back to the service. Users are only charged for their question when an answer is given. If the answer received is not satisfying, the user can first ask for additional research through an "answer clarification" request. If still unsatisfied, users can request to have the question reposted for a new answer or apply for a refund.[4] When the answer is completed, they can also rate the quality of the answer. The average rating of a researcher is easily accessible and has an effect on the standing of the researcher towards users and their employer Google. Finally, users can give a tip to the researcher who answered. The tip goes fully to the researcher in contrast to the

[3] Users might also have low Internet skills or simply no time to look for a thorough answer themselves.

[4] However, this is very rare. Our sample of 6,853 observations contains only 8 cases where a refund was granted and the price was returned.

price of a question where Google takes a 25% cut. If answering the question is not attractive to any researcher out of the pool, it will expire after 30 days.

According to Google all researchers are tested to make sure that they are expert searchers with excellent communication skills. Some of them also have expertise in a particular field. Additionally, answers are edited by Google to ensure quality. Researchers are independent contractors and for only a few of them Google Answers is the main job.

Any question that can be answered with words or numbers can get posted. Many users are looking for a specific piece of information like "How much tea was sold in China last year?", "In which San Francisco club did I see the Chemical Brothers play in 1995/96?" or "Race results from Belmont Park 5/24/1990. Who won the 8th & 9th race? And the daily double?". If the answer to the request is online, chances are pretty good that it will be found by the researchers. Moreover, complex questions are posted where background information is demanded and further links are expected. Examples are "How to get information about life in London during the late 1970's: films, television, plays, home decor, music, restaurants, political events, etc." or "Mutual perceptions of Europe and Asia via portraits". Also a number of questions are about marketing or business strategies. Questions are grouped into several categories as explained later.

Naturally, detailed questions regarding financial, medical or legal advice are excluded from Google Answers as is anything related to illegal activities.

3 Related Literature

A great number of experiments studies behaviour in social dilemma games. We particularly refer to Fehr, Gächter and Kirchsteiger (1997), henceforth FGK, and Gächter and Falk (2002).

FGK analyse a simple labour market with firms, workers and excess supply of workers. Three different contracts are simulated in experiments. While contract terms were exogenously enforced in the first treatment, workers were able to reciprocate in the second and both firms and workers were able to reciprocate in the third treatment. Effort levels of workers were significantly higher in the last (strong reciprocity) treatment and a contract that gives the opportunity for mutual reciprocity was found to improve efficiency.

Gächter and Falk (2002) study the interaction effects of reciprocity and repeated game incentives. A gift-exchange game between firms and workers was played in a one-shot and a repeated game treatment. Correlation between wage and effort in both treatments confirms reciprocal motivations. Higher effort levels in the repeated game treatment confirm the positive impact of reciprocal concerns.

3.1 Reciprocity

The set up in FGK consists of two stages - a third one is added in their strong reciprocity treatment. First, firms announce the details of their contract (wage, desired effort, the possible fine for shirking). Then, workers choose an offer they like and their effort level. Shirking, e.g. low effort levels, is verifiable only by chance. Firms' profits depend on the effort. In the final stage firms can reward or punish their workers. Equilibrium effort levels are determined by the offered wage and the amount and likelihood of the fine. If firms and workers are purely selfish, the third stage will not have any impact on equilibrium behaviour as it is costly for firms to reward or punish. Still, FGK found that firms often reciprocated. There was also a significant correlation between workers' effort and the firms' reaction (reward or punishment). Effort levels and profits for workers and firms were higher when firms had the opportunity to reward or punish.

The strategic structure of the Google Answers environment is very similar. Users post a question and set a price. Researchers "compete" for the right to answer. One researcher answers the question and posts it back. The quality of the answer depends on the effort of the researcher, which is not verifiable. The user's value of the labour relation depends on the researcher's effort and is therefore subject to moral hazard. Users can reject answers based on their quality. A rejection and a subsequent refund can be seen as a fine for the researcher, because such an incident affects the researcher's standing within Google Answers.

FGK explain the observed behaviour in their experiments by taking reciprocity motives into account. They relate to the seminal work of Rabin (1993). In addition we consider Dufwenberg and Kirchsteiger (2004) as their theory of sequential reciprocity is better suited for the sequential character of Google Answers. It is important to stress that this approach does not neglect the assumption that individuals maximise their utility. It merely allows their utility to reflect social concerns as well. Besides their own payoff it matters to them as well what the payoffs and intentions of other individuals are. For the relationship between users and researchers in the context of Google Answers concerns for reciprocity seem to play a significant role and we adopt this approach. The following section outlines how the sequential reciprocity equilibrium is determined.

The utility function of socially-minded individuals increases not only in their material payoffs but also in the psychological payoffs which depend on the individuals' kindness to others and beliefs about that. The resulting games are solved using the psychological games framework of Geanakoplos, Pearce and Stacchetti (1989). While the action set a_i describes the choices of player i (e.g. the effort of the researcher or the chosen price and tip of the user), b_{ij} defines the belief of i about the choices of player j, whereas $\widetilde{b}_{iji}$ is i's belief about what j believes are i's choices. This framework of beliefs allows us to express the kindness and beliefs about the kindness of individuals towards another individual. This is done by comparing an actual payoff Π to the equitable or fair payoff of a player, Π^e.

The equitable payoff of an individual is the average of his best and worst outcome based on the choices of the other individual.[5] For agent j it is given by:

$$\Pi_j^e(b_{ij}) = \frac{1}{2}(\max\{\Pi_j(a_i, b_{ij})\} + \min\{\Pi_j(a_i, b_{ij})\}) \tag{1}$$

It can be seen as a reference point for how kind i is to j as this kindness κ_{ij} is expressed by relating the actual payoff j is given by i to the equitable payoff of j:

$$\kappa_{ij}(a_i, b_{ij}) = \Pi_j(a_i, b_{ij}) - \Pi_j^e(b_{ij}) \tag{2}$$

Similarly i's belief about the kindness of j to i is:

$$\widetilde{\kappa}_{iji}(b_{ij}, \widetilde{b}_{iji}) = \Pi_i(b_{ij}, \widetilde{b}_{iji}) - \Pi_i^e(\widetilde{b}_{iji}) \tag{3}$$

Incorporating kindness and the beliefs about it gives the following utility function with a material payoff as the first term and the reciprocity payoff in the second term that is weighted by the reciprocity sensitivity α ($\alpha = 0$ is the special case of pure self-interest).

$$U_i = \Pi_i(a_i, b_{ij}) + \alpha_i \cdot \kappa_{ij}(a_i, b_{ij}) \cdot \widetilde{\kappa}_{iji}(b_{ij}, \widetilde{b}_{iji}) \tag{4}$$

The condition to solve the game is that in equilibrium all beliefs and second order beliefs are correct. It is also important to mention that beliefs of players are updated over the course of the game. The individuals apply Bayesian updating.

A positive reciprocity equilibrium exists. The user will give a tip, if his sensitivity to reciprocity is large enough: $\alpha_u > \overline{\alpha}_u$. The possibility of $\alpha_u < \overline{\alpha}_u$ corresponds to the nasty equilibrium.

After establishing conditions for the user to give a tip once the researcher has put in high effort, it has to be analysed whether the researcher will ever work at a high effort level in the first place. He knows that the user will never give a tip when $\alpha_u < \overline{\alpha}_u$ and therefore he will never give high effort. This constitutes the sequential reciprocity equilibrium of (low effort, (no tip, no tip)).

The researcher also knows that the user will act reciprocally once her sensitivity to reciprocity α_u is large enough. That means he assumes the user will reward the choice of high effort with a tip and will reply to low effort by not giving a tip. It can be shown that the condition for the researcher to make the high effort decision is always fulfilled and the sequential reciprocity equilibrium of (high effort, (tip, no tip)) results.[6]

By applying sequential reciprocity theory we can explain when users give a tip. Social preferences are necessary which are incorporated into the utility function with a reciprocity payoff. Once reciprocity gains (from returning kind behaviour) outweigh the material loss of paying a tip, users will prefer to tip. However, users and researchers have to be sufficiently motivated by reciprocity,

[5] The average is used here, because it is straightforward. Using another intermediate value is also possible and it does not affect the qualitative results. See also Dufwenberg and Kirchsteiger (2004) footnote 7.

[6] See Regner (2005) for more details.

e.g. α – their sensitivity to reciprocity – has to be large enough. Moreover, the researcher has to believe that the user's α is large enough in order to provide high effort in the first place.

3.2 Repeated Interaction

Google Answers users have a unique ID which makes them recognisable to researchers. The previous tipping behaviour of users can be observed by researchers and they may also be able to evaluate whether the effort of the respective researcher justified giving a tip. The relationship between reciprocity and reputation concerns in such a repeated games environment has been the topic of Gächter and Falk (2002). They aim to separate between non-strategic (reciprocity) and strategic (reputation) motives in their experimental set up of a gift-exchange game. In a one-shot treatment firms and workers were anonymously matched for 10 periods knowing that they couldn't face the same partner twice, in the repeated game treatment 10 periods were played with a known partner. While the authors do observe reciprocal behaviour in both treatments, the wage-effort relationship is steeper in the repeated game treatment and effort levels are significantly higher in the repeated game treatment (until the last period) than in the one-shot treatment. Moreover, they identify reciprocal, selfish and imitating types among workers.

A possible explanation for the multiple equilibria in repeated games is described by the folk theorem. Since future interactions are taken into account, even purely self-interested individuals would tip, if they discount the future not too much. In that sense tipping is profitable if and only if the increase in future payoffs outweighs the cost of tipping now. Essentially, leaving a tip has to generate enough good reputation that sufficiently many questions in the future are answered with high effort. When the user expects to ask questions frequently in the future, she will benefit often from high effort answers. Then tipping makes sense as long as she does not discount these future profits too much.

Alternatively, repeated interaction can be interpreted as a reputation mechanism where an updating process about a players' "type" takes place. When the decision to cooperate depends on the type of a player, e.g. good or bad, Kreps, Milgrom, Roberts and Wilson (1982) for instance show that cooperative equilibria can be reached. This kind of reputation model is based on Bayesian updating of beliefs and in the context of Google Answers it means that researchers update their beliefs about the tipping behaviour of the user they face. We can distinguish two different types, socially-minded and selfish users.

When there is no tip history available, researchers have an uninformed belief μ_0 regarding the chance a specific user will leave a tip. Researchers who know the past tipping behaviour of their user will take this into account when they make their effort decision. Their updated belief μ_n replaces the uninformed belief μ_0. When researchers observe that a user has a positive tip history, they will update their belief of the probability this user is of the "tipping" type accordingly. Given sufficient observations (past answers) the updated belief tends to converge to 0 for selfish users who do not tip and to 1 for users who do tip high effort answers.

Hence, the Bayesian updating of users' past tipping behaviour reduces the uncertainty the researchers face. The more they are able to inform themselves about the user's past behaviour, the better they are able to identify the user's type. They will have a better idea whether or not to expect a tip and will put in high effort when it is likely to be rewarded. They will put in standard effort, if chances for receiving a tip from the user they face are too low. Socially-minded users reciprocate and tip high effort answers, while selfish users do not tip. However, selfish frequent users may also take the researchers' updating into account and they might decide to imitate the "socially-minded" type. By tipping high quality answers they build up a good reputation and encourage high effort answers in the future.

Social preferences among researchers would reinforce these strategic considerations. Researchers are able to observe the previous tipping behaviour of users and they may also be able to evaluate, whether a tip was not given due to low effort. As explained before that means researchers will update their beliefs about the tipping behaviour of the user they face. They would take the kindness of 'their' user towards other researchers into account, if they are also motivated by indirect reciprocity.[7] Then the researcher's belief about the kindness of the user is updated based on the user's previous actions and the researcher will put in high effort, if the user has a good enough track record of tipping and rewarding high effort answers.

3.3 Summary

The section presented the results of two experimental studies and showed their similaritites to the Google Answers environment. In line with FGK we relate our analysis to sequential reciprocity theory and study whether reciprocity explains the voluntary payments. Google Answers users may use the service repeatedly. These users may anticipate the benefits from establishing a good reputation and the resulting high quality answers in the future. Therefore, reputation concerns may motivate kind behaviour (e.g. tipping) besides reciprocity. Similar to Gächter and Falk (2002) we analyse the impact of such repeated interaction on the voluntary payments.

Like FGK we test whether an open contracts design - providing mutual opportunities to reciprocate - produces voluntary payments, e.g. tips, by single users and whether these tips are motivated by reciprocity, i.e. a higher effort level of the researcher.

Hypothesis 1: The average tip of single users is significantly greater than 0. There is also a causal relationship between effort and tip.

Turning to repeated interaction two rationales (folk theorem, type updating) can explain the tipping of selfish frequent users. Hence, tipping out of

[7] See Seinen and Schram (2005) for an experimental study of indirect reciprocity where observed records of cooperativeness of a player induce others to cooperate with him.

strategic considerations hinges on the frequency of use and the belief updating of researchers.

Hypothesis 2: Frequent users tend to tip at least as often as single users.

Like Gächter and Falk (2002) we distinguish between socially-minded and selfish types of users. Since there is no end-period effect in the field data, we cannot separate imitators from truly reciprocal users like they did. In order to confirm this classification, users who tip must have had a tendency to tip in the past, likewise users who do not tip must have had a tendency not to tip in the past. Hence, users are of a certain type and stick to their strategy or preference, respectively.

Hypothesis 3: Tips are not distributed equally across users. The previous behavioural pattern is followed and the types "tipping" and "non-tipping" can be distinguished.

When different tipping types exist, researchers may inform themselves about a user's tip history and update their belief about the probability with which a user might tip. That affects their effort decision.

Hypothesis 4: After enough observations to establish a reputation the questions of users with a good tip history are answered with higher effort, questions of users with a "bad" reputation for tipping are answered with lower effort. Tip history and effort are positively correlated.

Finally, we test, whether an open contracts design has a similarly positive effect on efficiency in Google Answers as in FGK.

Hypothesis 5: After sufficient observations to reveal the behaviour of users, efficiency increases as researchers put in more effort (which means more value to users) and they get rewarded with a tip.

4 Description of the Data Set

Answers and comments to questions on Google Answers are not kept private to the user who posted the question. Instead, they are explicitly intended for the public by Google. Thus, everybody interested can benefit from the answers found. Past questions with the entire thread of comments, answers and answer clarifications plus information about their price, tip, rating and category are archived online. This gave us the opportunity to collect a large amount of detailed data. While we believe our data set contains all questions asked within the observation interval (question ID 230,000 to 300,000), we cannot guarantee the completeness of the set. The failure of an observation getting extracted would however be random and would therefore not affect our sample.

Our data set starts in July 2003 and ends in January 2004. Within this period we collected 13,948 questions and 6,853 of these were answered. The rest expired 30 days after the question was posted. Thus, total observations for our analysis amount to 6,853.[8] The number of answered questions over time is very stable. The range for the five full months of observation (August to December) is between 1,027 and 1,182. Overall, 1,745 answers have been tipped, which is a ratio of slightly more than a quarter.

The observations of our data set are generated by 4,840 different users. The highest number of observations posted by the same user is 77. Still, the majority of users just asked a single question. The median of the distribution is therefore 1 and the average number of questions per user is 1.42.

We collected the following data for each answer: The user ID of the person who posted the question, the price he set, the tip he possibly gave, the ID of the researcher who answered, date and time of posting the question, date and time of posting the answer, the rating of the researcher that was possibly left, the category of the question, the word count of the answer and the word count of the possible answer clarification.

Out of this data we computed additional variables. We calculated the time it took to answer a question (the difference between when the question was answered and when it was posted) and the frequency of use of the service (the number of questions posted (answered or not) by each user during the observation period). We expressed the effort of the researcher in two ways. We relate (i) the word count of the answer as well as (ii) the time it took to answer to the price. Moreover, we created a dummy if there was an answer clarification and various category dummies as explained later.

TABLE 1 : DESCRIPTIVE STATISTICS

variable	obs	mean	median	mode	st. dev.	min	max
price	6853	21.59	10	10	33.77	2	200
tip	1745	8.94	5	5	15.39	1	100
ratio tip to price	6853	0.204	0	0	1.271	0	50
rating	4359	4.70	5	5	0.63	1	5
time difference	6853	1.80	0.20	0.03	4.66	0.001	29.99
word count	6853	581.90	360	94	719.02	1	11482
answer clarification	6853	0.2934	0	0	0.4554	0	1
effort1 (word count)	6853	59.32	35.4	25	80.72	0.067	1657
effort2 (time)	6853	146.75	52.25	90	463.39	0.067	19200

where obs = number of observations, st. dev. = standard deviation

The range of prices is pre-determined by Google Answers. The lowest price users can set is $2 and the highest price possible is $200. These are also minimum and maximum price of the sample. The average price conditional on the question being answered (6,853 observations) is $21.59, while the average price of the 7,095 questions that expired without an answer is lower. It is only $19.23. Median and mode of the price distribution are $10.

[8] Since the focus of our analysis is the tipping aspect we decided to deliberately truncate the data set considering only answered questions as observations. We are aware of the fact that a more general model would analyse all questions and why some are not answered. We only touch this issue in our paper.

Minimum and maximum values for the tip are also pre-set by the service. There is an upper limit of \$100 for the tip. The mean of the distribution is \$8.94 and its median and mode equal \$5.

The time difference between question and answer is expressed in days, so 1 equals one day and for instance the value 0.25 means it took 6 hours to answer. The quickest answer came after only two minutes, the slowest was given just before the 30 day expiration deadline. The median of the distribution is 0.2. That means half of all answers were posted within 5 hours.

The word count is the number of words of an answer. The shortest answer was a single word (”No” to be precise) and the longest contained 11,482 words (a \$190 question with \$65 tip).

A rating has been given for 4,359 answers, roughly two thirds of the total. The possible range is from 1 to 5, with 5 being the top rating. If users decided to give a rating, they did not mind giving the highest possible as median and mode are 5 and the average rating is 4.7.

The dummy variable answer clarification equals one if a clarification was given and zero otherwise. An answer clarification was requested and given in 29.34% of all times.

The variable ‘effort1’ describes the word count-based effort the researcher has put into an answer. It is the ratio of ‘word count’ over ‘price’. The more words researchers have included in answers of equally priced questions, the higher their effort has been. The average words per dollar are 59.32 and again median and mode are below the mean. The most words per dollar received have been 1,657.

The variable ‘effort2’ describes the effort of the researcher in terms of the time it took to answer. It is the ratio of ‘price’ over ’time difference’. The quicker researchers have delivered answers of equally priced questions, the higher their effort has been. Since our measure for time is the difference between the posting of the question and the posting of the answer this variable has to be taken with some caution. The variable ’time difference’ might not always be the time a researcher has worked on a question. It is exactly that, if the researcher started to work right after the question has been posted and this is arguably the case for many questions. However, questions might remain in the pool of unanswered questions for a while before a researcher decides to work on the answer. This can be up to 30 days after the posting of the question. The ’time difference’ is then the time worked on the answer plus the time that passed until the researcher started working.

Table 2 lists the ten different categories in which users can post their questions. We created dummies for all of them except the last one: ‘Miscellaneous’. Their popularity is quite different. While only 216 observations are in category ‘Sports and Recreation’, the most popular category after ‘Miscellaneous’ was ‘Computers’ with 1,209 entries. About 31% of all observations in ‘Arts & Entertainment’ or ‘Sports and Recreation’ have been tipped. Users in the ‘Business & Money’ category appear to be the least generous as only 21.68% of these questions have been tipped. The tip rate of the other categories is fairly close to the overall average of 25.46%. The ‘Business & Money’ category also features the highest average price (\$34.32).

Table 2 : Question Categories						
Category Name	total	with tip	% tip	avg. p	avg. tip	avg.% tip
Arts & Entertainment	696	220	0.3161	14.75	2.18	0.1476
Business & Money	1107	240	0.2168	34.32	2.68	0.0781
Computers	1209	322	0.2663	20.42	3.02	0.1479
Family and Home	287	68	0.2369	13.26	1.84	0.1386
Health	488	124	0.2541	25.15	2.71	0.1079
Reference, Education, News	795	206	0.2591	21.29	2.66	0.1249
Relationships and Society	304	80	0.2632	19.83	1.87	0.0944
Science	453	112	0.2472	20.31	2.39	0.1175
Sports and Recreation	216	67	0.3102	13.62	1.66	0.1218
Miscellaneous	1298	306	0.2357	18.35	2.15	0.1174
all	6853	1745	0.2546	21.59	2.47	0.1143

5 Analysis of the Data

The results of a panel regression of the data set are presented first, then we analyse some specific aspects in more detail. We focus on the possible motivations for tipping that have been identified: conforming to a social norm, social preferences or strategic considerations due to reputation concerns. Finally, we analyse the relationship between updating, effort decision and efficiency in the data set.

5.1 Estimations

Three different arguments should explain the tipping behaviour in the data and we use a set of proxy variables to test them. Firstly, reputation may matter. Frequent users of the service have an incentive to build up a good reputation and may regard tipping as a strategic device. Secondly, social preferences would make people tip. Users who are socially-minded should leave a tip as long as there is a reason to reciprocate. Thirdly, the tip should simply be affected by the price of the question. Users may tend to tip proportionally to the price, giving a high tip for a highly priced question and vice versa.

Our proxy for reputation concerns is the frequency with which a user asked questions during the observation period. A high frequency of use means the user should put much weight on her future income and this is positively affected by tipping now. The more questions posted the more generous users should be with the tip – simply out of strategic considerations. We use the logarithmic value of the frequency of use in our regression.

To take account of behaviour that elicits social preferences we use the following set of proxies. The effort involved in a given answer shows how hard a researcher worked for the answer. Users sufficiently motivated by social preferences would then reciprocate and tip. Has a rating been left, a user seems to care about the benefit of the researcher, although only non-monetarily. Moreover, if an answer clarification has been given, the researcher put in an extra effort and this could once again trigger positive reciprocity. While we are aware that these variables can only be rather crude surrogates for what motivates voluntary giving, we believe that this quantification can nevertheless contribute to a better understanding of social preferences.

Effort is metered in terms of time and word count relative to the price of the question. An answer of average quality that comes very fast might have a higher value for the user as will an answer that comes within the usual time but is very comprehensive with a lot more background information than expected. Therefore, a question that has been answered with relatively high effort is more likely to generate value for the user. Users with social preferences would tend to return the kind behaviour of the researcher and give a tip, when their question has been answered with high effort.

Once a user leaves a rating, it seems fair to assume that she is not entirely self-interested. It only costs time and a positive impact on a user's reputation seems hard to imagine. It shows on the other hand that the user cares about the researcher since researchers' ratings are fairly important to them. There is no monetary sharing of course, but leaving a rating can be seen as a sign for a minimum of social preferences, necessary for giving a tip.

An answer clarification is given only on request, after the answer itself has been posted. It is likely that the clarification adds more value to the answer, but it could also be argued that the researcher puts in some extra effort and this should trigger reciprocal behaviour of the user. Hence, we use the answer clarification dummy as another proxy for social preferences.

The tip is the dependent variable in our regression and the equation we estimate is

$$t = k + bX + \epsilon$$

where k is the constant, b is the vector of the coefficients, X is the vector of our variables and ϵ is the error term. The explanatory variables are the price of the question, the frequency of use (logarithmic), the rating, effort1 (word count-based), effort2 (time-based), the answer clarification dummy and the dummies for the categories. Table 11 lists the variables, their coefficients and respective standard errors for our estimations.

Since no negative tip can be given the distribution of the tip is left-censored at zero. It is also right-censored at 100 by design of the service. Therefore, a censored regression model appears appropriate for our data. The Tobit model takes limits of the range of the dependent variable into account to ensure unbiased and consistent estimates. The Tobit maximum likelihood estimates are shown in column I. These are the results of the standard Tobit model which assumes a single distribution function for the dependent variable. However, there is reason to believe that the decision on whether to tip or not and the decision how much to tip (given one has chosen to tip) are separated. Different distributions could be underlying. A two-equation model of Cragg (1971) will take this into account. Using a two-step approach we can split up the two decisions whether to tip and if so then how much to tip by using different probability functions in the Tobit model. (Amemiya 1984) A Probit model estimates the binary decision of whether to tip or not and a truncated regression is used to estimate the size of the tip. The results are given in columns II and III, respectively. We compare the fit of the restricted Tobit model and the unrestricted composite model of Probit and truncated regression to see if the two-equation approach should be considered. A likelihood ratio test of restricted against unrestricted model rejects the null hypothesis clearly. Separating the decisions and estimating a Probit model combined with a truncated regression is advisable.

Table 3: Estimation Results

	I: TOBIT		II: PROBIT		III: truncated	
Explanatory variable	coeff.	st. error	coeff.	st. error	coeff.	st. error
Price	.2099 ***	.0097	.0034 **	.0011	1.874 ***	.3367
Frequency of use	1.0083 ***	.2427	.0923 *	.0477	13.361 ***	4.825
Rating	12.306 ***	.6005	1.100 ***	.0710	88.840 **	26.142
Effort1	.0225 ***	.0038	.0016 ***	.0003	.1256 **	.0590
Effort2	.5315	.4367	-.0045	.0417	-3.187	12.50
D_ANSCL	4.147 ***	.6887	.2795 ***	.0670	60.24 ***	15.89
D_ART	1.2923	1.212	.2710 **	.1250	-51.72 *	26.81
D_BIZ	-1.7325	1.139	.0522	.1273	15.51	18.59
D_COM	-.41105	1.066	.0885	.1149	-4.772	17.49
D_FAM	.0851	1.798	.1300	.1744	-6.632	36.97
D_HEA	.4860	1.424	.1300	.1462	56.88 **	26.80
D_REF	.4030	1.200	.0804	.1227	-7.735	20.01
D_REL	-2.3915	1.674	.0063	.1714	-49.77	38.34
D_SCI	-1.741	1.442	.0464	.1535	-9.345	24.78
D_SPO	3.0465	1.930	.4770 **	.1955	-119.16 *	58.81
Constant	-73.58 ***	3.129	-6.121 ***	.3716	-783.9 ***	189.1
Sample size	6853		6853		1745	
Log likelihood	-8730.1987		-2443.3426		-5399.3065	
Statistical significance	*=10% / **=5% / ***=1%					

where D_* = dummy variable for #

ANSCL = answer clarification, ART = arts & entertainment, BIZ = business & money

COM = computers, FAM = family & home, HEA = health

REF = reference, education & news, REL = relationship & society

SCI = science, SPO = sports and recreation

if all category dummies = 0, we have observation in 'miscellaneous'

The price is highly significant in all regressions. One argument for separating the tipping decision and the decision of how much to tip was that the price of the question might not affect the first, but even more the second decision. In fact, it turns out that the price does affect both decisions, but the decision whether to tip to a lesser extent. The data also confirms the significance of reputation concerns. The estimators for the coefficient of the frequency of use are statistically significant, yet again at a lower level (10%) in the Probit regression.

The effect of the word count-based effort is clearly positive as well. While the Probit estimator for 'effort1' is very significant, it is less significant in the truncated regression. It seems high effort only affects the decision to leave a tip or not, it has a lesser impact on the size of the tip. The regression results for the time-based effort give no indication that this variable is significant. Either the described problems of computing the variable are too big or users do not actually care how fast they receive answers. The rating is one of the best variables to explain the tip. Its coefficients are highly significant throughout the regressions. It also clearly matters whether an answer clarification has been given. Again, the coefficients are very significant throughout.

Our dummy variables show, if there are different tipping patterns in the various categories. While the standard Tobit specification of the model does not find significance of a category dummy, we get some meaningful results in the two-equation approach. Dummies for the categories 'Arts & Entertainment' and 'Sports and Recreation' are significant in Probit at the 5%-level. These are the categories with tip rates (31%) clearly above the average of 25%. However, their coefficients are significant and negative in the truncated regression. Users in this category tend to tip often, but their tips are very small. On the other hand, users in the category 'Health' seem to tip a lot, if they decide to at all.

It would be intuitive to include the tip history into the regression as well. A high tip history (e.g. five previous answers out of seven tipped) should indicate a higher chance of a tip given at the present answer than when the tip history is low (e.g. no previous answers tipped). In fact, the tip history is highly significant, but it causes multicollinearity with the frequency of use and biases the regression. Therefore, it was not included. Regressions with the tip ratio as the left-hand side variable instead of the price on the right-hand side deliver largely similar results.

Our censored regression models are based on maximum likelihood and they assume a normal distribution of the error term and homoscedasticity. A Bera-Jarque test rejected the normality assumption. A censored least absolute deviations estimator offers an alternative as it is robust to changes in the error distribution. Its estimators are consistent, but inefficient. However, this could not be performed due to software limitations.

Nevertheless, all our estimation results confirm the positive effects of social preferences and reputation on the tip that theory and experiments suggest. The coefficients of frequency of use, effort1, rating and answer clarification all have the expected sign and they are significant at least at the 5%-level. Moreover, the fit of the regressions are sufficient at 15%.

5.2 Social Norm

Tipping is widespread in many service professions. Azar (2004) and Lynn (2005) survey tipping behaviour in common service situations like a restaurant visit, for instance. While originally (16th and 17th century in Europe) people tipped out of gratitude for extra service, out of compassion or to encourage better service, it soon became a social norm. Nowadays people rather feel obliged to tip. They tip mostly in order to conform to the social norm or to avoid embarrassment. In many occasions tipping is very institutionalised and a quite precise fraction of the bill is to be tipped. In restaurants people would tip roundabout the same percentage of their respective bill. (Azar 2004)

If a social norm similar to restaurant tipping is the motivation for the tipping in Google Answers, we should observe a similar pattern. Many users should tip and the size of their tip (a percentage of the price paid) should not vary too much. However, the data appears to suggest otherwise. About 25% of all answers have been tipped. In the majority of the cases people do not tip. Users do not seem to be motivated by a widely followed social norm. It could be argued that the online environment and the fact that there is no direct contact makes less people adhere to the norm of tipping, but that it is still the social obligation that drives them to leave a tip. The high variation of tips contradicts that. The mean of the ratio "tip to price" is 0.204, however the standard deviation is 1.271. Many users do not tip at all, but also quite a few users leave very high tips compared to the price of the question. We observe tips fifty times the price of the respective question. This does not fit the tipping pattern we know from restaurants. Hence, we conclude that it seems unlikely that Google Answers users tip due to a social norm.

5.3 Social Preferences

Reputation effects may influence the tipping behaviour of users. We study the behaviour of single users only, in order to control for reputation and focus on social preferences.

Over the entire observation period 18% of all single users did tip.[9] Since our data set ends in January, we would treat users incorrectly as single users who joined shortly before the cut-off line and continued asking questions afterwards. These users might tip out of reputation concerns, while we would consider them as single users. This may in deed be the case as the tip rate of single users over the observation months increases towards the end of the period.

Table 4 shows the tip rate of (single) users and also the ratio of the single user tip rate to the tip rate of all users, since there is a slight upward trend of the tip rate over the months. The ratio is the lowest in October which means that we have the fewest tipping single users in this month. Single users in October are three months distant from having possibly posted a question before the observation period or posting another question in the future, yet more than 17% of them decided to tip their answer.

Table 4: Tip Rate of (single) Users over Time

month	July	August	September	October	November	December	January	all
single users	16.56	15.38	16.7	17.28	17.53	21.13	22.38	17.98
all users	21.75	22.93	25.22	26.42	23.85	28.69	29.45	25.46
ratio	0.7614	0.6707	0.6622	0.6540	0.7350	0.7365	0.7599	0.7062

[9] See Table 5 in the next subsection for the data about single users.

In the entire data set there are 2,942 single users and 529 of them did leave a tip. The mean of the distribution is 1.47 and the standard deviation is 5.79. The frequency of tips by single users is statistically significant from zero at the 1% level. A separate regression only with single users delivers results similar to the main regression. The word count-based effort is statistically significant at the 5% level.

After controlling for the impact of reputation effects we find that tips are still prevalent, albeit at a lower rate than among frequent users. Moreover, tips are explained by effort. This confirms hypothesis 1. Our approach to control for repeated game incentives is naturally limited by the field data set and cannot be regarded as bullet proof. Nevertheless, the results are in line with comparable experimental and field studies. Voluntary payments at a significant level are also observed in another field study where reputation effects cannot play a role. (Regner and Barria 2005)

5.4 Reputation Concerns

Frequent users could have an interest in building up a reputation of appreciating good effort and acknowledging it with a tip. This way they may attract workers who recognise them as generous and will deliver good quality. This motivation may be of particular relevance in online environments, since buyers and sellers do not see each other online.

In order to test the impact of reputation effects on the tipping behaviour we split our sample into subgroups. We computed each user's frequency of use which is the number of questions posted during the observation period. Some users may not have a clear idea of how often they are going to use the service when they start with the first question, but on average they should be aware of that. Therefore, we believe the frequency of use is a good indicator to what extent users should be concerned about their reputation. Since our sample is taken out of an ongoing stream of questions, there is no last period or rather last observation. Hence, we will not be able to observe a last period effect that may distinguish behaviour motivated by reputation from social preferences.

We clustered our observations into four subgroups according to the number of questions users asked over the entire period: single users, occasional users who asked two or three questions, frequent ones with four to nine questions asked and very frequent users who asked more than ten questions during the observation period. The following table shows the pricing and tipping behaviour of users in each subgroup.

Table 5: Subgroups by Frequency of Use

sub group	users	total obs.	no tip	with tip	tip rate	avg. price
single users	2942	2942	2413	529	0.1798	21.80
2 to 3 questions	1105	1395	1025	370	0.2645	23.24
4 to 9 questions	681	1169	778	391	0.3345	25.17
10+ questions	112	1347	892	455	0.3378	16.30
all	4840	6853	5108	1745	0.2546	21.59

The clustering is based on the number of questions asked (13,948), while total observations is the number of questions answered (6,853) as naturally only these can contain information about the answer given. The column 'users' gives the number of users within each group.

The largest group of our sample are single users. Almost 18% of them gave a tip, which is far below the total sample average of 25%. However, with increasing number of questions asked we can observe a steadily increasing tip rate. Already about 26% of the observations by users who asked two or three questions were tipped. The tip rate goes up to more than 33% for the subgroup of users that asked four to nine questions and basically stays at that level for the group of very frequent users.

These results lead us to conclude that occasional users already take reputation concerns into account. For frequent and very frequent users reputation concerns matter even more as the tip rate increases further. However, there seems to be a satiation effect as the rates for the last two subgroups are virtually the same.

Frequent users tip consistently more often than single users, similar to the experimental findings of Gächter and Falk (2002). In fact, the tip rate increases with the frequency of use, which confirms hypothesis 2.

5.5 Updating, effort decision and efficiency

This section tries to shed more light on the decision making of researchers. They may update their beliefs about the likeliness the user they face will tip (if effort is high). In the data set we can specify the tip history of each user at each number of question she posted. It is the amount of answers she tipped divided by the total of answers she received at that point. Recall that this information is not very straightforward to get for the researchers. It is not shown next to the user name as the past average like the rating of researchers is for instance.[10] Table 6 splits the sample into different sub groups with respect to the question number asked. Essentially we see that the tip rate increases for users who keep on asking questions which is not surprising as we know that frequent user tend to tip more often.

Table 6: Question no. and tip rate

question no.	obs	tip rate	avg. price	avg. tip (if tip)
first	4354	0.21	21.94	8.25
2nd to 9th	1945	0.33	22.59	11.62
10+	554	0.34	15.26	10.05

When we consider the respective tip history of each user at each question number we find that there is a large split between tipped and untipped questions. Naturally, the tip history is 0 at question number 1. In the intermediate range of question numbers users who did not tip had an average tip history of just

[10] Or the seller's reputation on eBay.

19%, while users who left a tip had one of 61%. We can observe a very similar split in the high range of question numbers as shown in Table 7.

Users who tip an answer clearly had a tendency to do so in the past as well. On the other hand, users who did not give a tip have a rather low tip history. Users appear to have preferences or a strategy to tip (high quality answers) and they stick to it, which confirms hypothesis 3.

Table 7: Question no. and tip history

question no.	avg. tip history	/ without tip	/ with tip
first	–	–	–
2nd to 9th	0.32	0.19	0.61
10+	0.31	0.16	0.60

If researchers do in fact update their beliefs about the chances to get a tip for high effort work, then they should anticipate that and make their effort decision based on this belief. They should put in low effort when they face a user with a poor tip history and they should exert high effort when they meet a user who has tipped in the past.

Table 8 relates the question number to the effort of the researcher (the word count-based 'effort1'). When a user asks the first question no tip history exists and the effort decision cannot be based on the user's past. Comparing questions with and without tip shows that effort is slightly higher for the tipped ones. This is just what we should expect since we know that effort explains the tip. We have already seen that users tend to stick to their tipping pattern. The tip history appears to be a good indicator of the user's type/strategy. Researchers can update their beliefs about the user's tendency to tip, once previous questions are available. They can make an educated effort decision as the updating tells them whether the user is likely to reward high effort with a tip or not. While the split between questions with and without tip is similar in the intermediate range of question numbers (Mann-Whitney test between tipped and untipped samples, 5% significance level), the gap clearly widens in the high range. With ten or more questions of tip history available the tipped questions have been worked on with significantly more effort than questions without tip (Mann-Whitney test, 1% significance level). The average effort is also higher compared to earlier questions that were tipped (Mann-Whitney test between tipped samples of occasional and frequent users, 1% significance level).

Table 8: Question no. and effort1

question no.	average effort1	/ without tip	/ with tip
first	58.65	56.75	65.82
2nd to 9th	61.32	58.49	67.03
10+	57.50	48.58	74.73

It seems that in deed researchers update their beliefs based on the tip history and that they make their effort decision according to that belief. Moreover, users stick to their behaviour type/strategy and they reward high effort, if they are sufficiently motivated by reciprocity or reputation. Effort and tip history are correlated in the frequent user sample (Spearman correlation coefficient, 5% significance level). This confirms the fourth hypothesis.

Finally, the question is whether something can be said about the efficiency of this open contracts design and the fact that it provides opportunities to reciprocate for users and researchers. Does it pay off for researchers to put in high effort, when they work on questions of users who are known for tipping? We can express the rent of researchers in dollars received per 100 words. It equals price plus tip divided by word count and multiplied by 100.[11]

Table 9: Rents of Researchers

question no.	average rent	/ without tip	/ with tip
first	6.731	5.936	9.73
2nd to 9th	7.333	5.828	10.374
10+	7.039	6.962	7.19

We have seen that researchers work harder when the available tip history of a user is promising. The uncertainty regarding a possible tip is reduced. Now we see that they do get rewarded for that extra effort. The rents for researchers when tips are given are consistently higher than when no tips are given (Mann-Whitney test between samples without and with tip, 1% significance level for single and occasional users, no significance for frequent users). Users known for tipping get higher effort answers than new users, but they also reciprocate and apparently let the researchers participate in the gain from a high quality answers by returning some of the surplus and leaving a high tip.

However, once there is a substantial tip history available the researcher's effort for untipped answers drops down so much (see Table 8) that their rent reaches the level of tipped answers (around $7 per 100 words). It appears they adjusted their effort decision since their updating tells them a tip is very unlikely. Then researchers do as well as when they put in high effort and get a tip (Mann-Whitney test between tipped and untipped samples, no significant difference).

The open contract design increases the effort level and the efficiency. It seems that it encourages socially-minded users to reciprocate (tipping high effort answers) and that it makes self-interested users consider building up a good reputation (in order to motivate future high effort answers). Through belief updating the researchers are able to match their effort decision better to the user types. Consistent high effort answers are possible in contrast to a more complete contract that does not allow a tip. Such a strict contract type is simulated, when users reveal that they are not going to tip (long enough low tip history). Then researchers update their beliefs accordingly and put in relatively low effort, which results in less quality. Hence, we can confirm hypothesis 5.

[11] We do not calculate researchers' payoff here but instead their rent to keep differently-priced questions comparable.

6 Conclusions

We collected a rich data set of Google Answers questions to investigate the real-life pricing and tipping behaviour of individuals. The motivation for the paper was to use this data set to check the external validity of experimental studies of social dilemma games. In particular, we were interested in the underlying motivations for the occurring voluntary payments and the efficiency of such an open contracts design.

Our empirical analysis shows that the tip can be explained by social preferences proxies (effort, rating and answer clarification) and reputation proxies (frequency of use). In addition to the regression analysis we found evidence for social preferences, once we controlled for reputation concerns by focusing on single users. Moreover, we were able to separate the reputation concerns within the data and found further evidence for strategic considerations among users. The higher tip rates of frequent users are in line with the experimental findings of Gächter and Falk (2002). The effect of reciprocal behaviour and repeated game incentives appear to be complementary. The data from Google Answers also confirms the positive effects of an open contracts scheme on the effort level as found in Fehr, Gächter and Kirchsteiger (1997).

We relate the findings to the theory of sequential reciprocity of Dufwenberg and Kirchsteiger (2004). Tipping takes place even among single users, if they are sufficiently sensitive to reciprocity. Frequent users may have an incentive to imitate this socially-minded behaviour. They tip in order to create a reputation of rewarding high effort of researchers, thus attracting high effort answers in future transactions. Once researchers realise this through updating of their beliefs about the user's type, they will put in high effort in anticipation of a tip. The uncertainty about whether a user will tip is reduced when a sufficient tip history is available and researchers can update their beliefs reliably. They can make an educated effort decision. High effort is matched to rewarding users, low effort is matched to users who do not tip. The open contract design can be seen as a virtuous circle that increases efficiency.

7 Appendix

References

[1] AMEMIYA, T. (2002). "Tobit Models: A Survey", *Journal of Econometrics* 24: 3-61.

[2] AZAR, O. (2004). "The History of Tipping - from Sixteenth-century England to United States in the 1910s", *Journal of Socio-Economics*, 33: 745-764

[3] CAMERER, C. (2003): "Behavioral Game Theory: Experiments in Strategic Interaction", Princeton University Press, Princeton

[4] CRAGG, J. G. (1971): "Some Statistical Models for Limited Dependent Variables with Application to Demand for Durable Goods", *Econometrica*, 39, 829-844.

[5] DUFWENBERG, M., and G. KIRCHSTEIGER (2004): "A Theory of Sequential Reciprocity", *Games and Economic Behavior*, 47, 268-298.

[6] FEHR, E., S. GÄCHTER, and G. KIRCHSTEIGER (1997): "Reciprocity as a Contract Enforcement Device: Experimental Evidence", *Econometrica*, 65, 833-860.

[7] FEHR, E., and K. M. SCHMIDT (2003): "The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories", *Handbook of Reciprocity, Gift-Giving and Altruism*

[8] GÄCHTER, S. and A. FALK (2002): "Reputation and Reciprocity: Consequences for the Labour Relation", *Scandinavian Journal of Economics, 104(1), 1-26*

[9] GEANAKOPLOS, J., D. PEARCE, and E. STACCHETTI (1989): "Psychological Games and Sequential Rationality", *Games and Economic Behavior*, 1, 60-79

[10] KREPS, D., MILGROM, P., ROBERTS, J. and WILSON, R. (1982). "Rational Cooperation in the Finitely Repeated Prisoners Dilemma", *Journal of Economic Theory* 27, 245-252

[11] LYNN, M. (2005): "Tipping in Restaurants and Around the Globe: An Interdisciplinary Review" in: "Foundations and Extensions of Behavioural Economics: A Handbook", edited by M. Altman, M. E. Sharpe Publishers

[12] RABIN, M. (1993): "Incorporating Fairness into Game-Theory and Economics", *American Economic Review*, 83, 1281-1302

[13] REGNER, T. (2005). "Why Voluntary Contributions? Google Answers", *University of Bristol working paper 05/115*

[14] REGNER, T. and J. BARRIA (2005). "Do Consumers Pay Voluntarily? The Case of Online Music", *working paper*

[15] SEINEN, I. and A. SCHRAM (2005). "Social status and group norms: Indirect reciprocity in a mutual aid experiment", *European Economic Review*