THE CENTRE FOR MARKET AND PUBLIC ORGANISATION

The Centre for Market and Public Organisation, a Research Centre based at the University of Bristol, was established in 1998. The principal aim of the CMPO is to develop understanding of the design of activities within the public sector, on the boundary of the state and within recently privatised entities with the objective of developing research in, and assessing and informing policy toward, these activities.

CMPO

Centre for Market and Public Organisation
University of Bristol
Department of Economics
Mary Paley Building
12 Priory Road
Bristol BS8 1TN

Tel: (0117) 954 6943
Fax: (0117) 954 6997
E-mail: cmpo-office@bristol.ac.uk

Why Voluntary Contributions? Google Answers!

Tobias Regner

January 2005

Working Paper No. 05/115

ISSN 1473-625X

CMPO Working Paper Series No. 05/115

Why Voluntary Contributions? Google Answers

Tobias Regner

January 2005

Abstract

We study the pricing and tipping behaviour of users of the online service `Google Answers'. While they set a price for the answer to their question ex ante, they can additionally give a tip to the researcher ex post.

We develop a model that is based on reciprocal theories of social preferences pioneered by Rabin (1993) and extended by Dufwenberg and Kirchsteiger (2004). The predictions of our model are empirically tested with the field data we obtained.

The reasons for leaving a tip are analysed. A significant amount of users are motivated by social preferences. We also find strong support for reputation concerns. Moreover, researchers appear to adjust their effort based on the user's previous tipping behaviour.

We conclude that an endogenous incomplete contracts design encourages people to contribute voluntarily. This is motivated by reciprocity when people are socially minded, but also generally by strategic behaviour to build up a good reputation. Efficiency is increased when contracts are left open deliberately as high effort is sustained.

Keywords: social preferences, reciprocity, moral hazard, reputation, internet

JEL Classification: C24, C70, C93, D82, L86

Acknowledgements

We thank the Leverhulme Trust and the ESRC for funding this research and I am grateful to Maija Halonen-Akatwijuka and Sebastien Mitraille for valuable discussions and to David Winter, Juergen Bracht and Osiris Parcero for their econometric advice.

Address for Correspondence
Department of Economics
University of Bristol
12 Priory Road
Bristol
BS8 1TN
tobias.regner@bristol.ac.uk

CMPO is jointly funded by the Leverhulme Trust and the ESRC

1 Introduction

There are literally hundreds of laboratory experiments showing that individuals consistently deviate from the neo-classical self-interest hypothesis. The first ultimatum games have been conducted in 1982 and the contradiction of their results with the assumption of purely self-interested individuals has been confirmed by all kinds of variations of the standard experiment. This has been generally explained with other-regarding behaviour of individuals and attempts have been made to design a broader theory than pure self-interest, one that incorporates psychological insights from behavioural economics, namely social preferences.

Skepticism of the results and its related criticism of the self-interest hypothesis are mainly based on the potential framing effects of experiments. Participants of lab experiments know they are under observation and might behave different than in reality. The results in doubt of self-interest are questioned based on these alleged blurrings that could come with experiments. We believe this is a valid point for such a delicate issue like other-regarding behaviour and therefore want to take the discussion of what drives people's behaviour to another playing field.

We collected field data about the pricing and tipping behaviour of 'Google Answers' users. In this online service (which is part of but not to be confused with the Internet search engine Google) users can post questions and set a fixed price for the answer. They can also leave a voluntary tip to the researcher who found the answer. Our new data set contains 6,853 answers (collected over half a year) with a number of additional explanatory variables thanks to the online availability of past questions of the service. The average price for one answer is about \$20. The service is a small but significant part of Google's business and cannot be considered online nonsense. This rich data set puts us in a position to test the relevance of social preferences in real life instead of observing through the lab microscope. We avoid the potential framing effects of laboratory conditions as 'our subjects' are simply concerned about their own cause with no outside effects or bias possible as it could be the case in the lab.

About 25% of all answers have been tipped and the three main conclusions from our empirical analysis follow. The more questions users ask over time the more likely they are to tip. Thus, there are reputation concerns that make people tip. It seems that they aim to build up a fine reputation in order to encourage good effort of researchers for future questions. Naturally, individuals have this concern no matter what their preferences are. Moreover, even single users tip, though: around 18%. This clearly deviates from self-interest. Hence, social preferences of some kind have to be involved. Finally, tipping – motivated by social preferences or out of reputation concerns – 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.

Our theoretical model describes the moral hazard relationship between user and researcher. We model social preferences by incorporating reciprocity into the utility function, something first done by Rabin (1993). In particular we follow the concept of sequential reciprocity of Dufwenberg and Kirchsteiger (2004). The moral hazard aspect of our model is inspired by the literature that started with Fehr, Gächter and Kirchsteiger (1997). Labour relations are studied where the worker's effort is not contractible. They describe the benefits of contracts that give a mutual opportunity to reciprocate. The firm can reward or punish the worker ex post. These endogenous incomplete contracts encourage reciprocity the most and they reach higher effort levels than stricter contract options.

The paper will discuss three possible motivations for the tipping of users and we will test empirically, if these reasons explain the behaviour of Google Answers users. First of all, the tipping can be to conform with a social norm as it is the case in restaurants, for instance. Moreover, social preferences can motivate users to leave a tip and our model explains this in detail. Finally, users may decide to tip, because they benefit in the long-run. A good reputation means high effort answers and more value for the user. We also explain this possible motivation of frequent users in our model.

Tipping is widespread in many service professions (most popular in restaurants) and so we should consider the economics of tipping in our analysis. Azar (2003) surveys the history of tipping and covers the question why people tip in the usual service situation of 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. We will test empirically, if this is the tipping behaviour of Google Answers users.

Social preferences relax the neoclassical assumption of pure self-interest. Individuals get utility from their own payoffs, but also from other sources. Fehr and Schmidt (2001) and Camerer (2003) survey the existing theoretical and experimental literature of other-regarding behaviour. The numerous variations of experiments are described as well as different theoretical models to incorporate concerns for fairness and reciprocity. Charness and Rabin (2002) test the different strands of theoretical models of social preferences by conducting 29 slightly different games. They find that concerns for social welfare (agents like to increase social surplus not just their private one), reciprocity (a desire to raise or lower others payoffs depending on how nice or not these behaved) and indifference aversion (agents want to reduce differences between their and others payoffs) all play a significant role to explain other-regarding behaviour. For simplicity reasons we will focus only on reciprocity in our model. Using natural experiments to test social preferences is a fairly new technique. List (2004) collected data from a television game show in the public goods game context, for instance. However, our analysis appears to be the first natural experiment that examines social preferences in the particular context of a bonus game.

The importance of reputation, particularly in online environments, is another possible motivation for frequent users to leave a tip. In online auctions buyers and sellers do not know each other. Naturally, the biggest online auction house eBay has been the topic of a number of recent studies regarding the role of the seller's reputation.[1] Cabral and Hortacsu (2004) model the reputation concerns of eBay sellers theoretically. Buyers react to the feedback (reputation) changes of sellers and sellers anticipate that. Reputation seems to be very important for online transactions. This seems plausible since users of online service do not know and see each other when they transact. Our analysis of Google Answers takes that into account and tests for reputational concerns.

In the following section we give a thorough overview of the pitch of our natural experiment as we describe the service Google Answers in detail. Section 3 contains our model of users and researchers. Section 4 describes our data set, while section 5 analyses it. The conclusions are in section 6.

2 The Online Service Google Answers

2.1 The Service

Google has become a household name in the Internet. It 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. The service Google Answers (www.answers.google.com) provides help to Internet users whose requests are too specific for a standard Google search. It offers assistance from researchers with expertise in online searching.

After registering anonymously with the service users post a question to Google Answers and specify how much they are willing to pay for an answer. A researcher will search in the Internet to obtain the information and when they find it, they will post it to Google Answers. Users are only charged for their question if and when an answer is posted. 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. They can also rate the quality of the answer. The average rating of a researcher is easily accessible and should affect the reputation of the researcher towards users and their employer Google. Moreover, users can give a tip to the researcher who answered. The tip goes fully to the researcher in contrast to the price of a question where Google takes a 25% cut.

There is a non-refundable listing fee of $0.50 per question plus the price set which can be as low as $2 or as high as $200. Once a researcher decides to search for an answer, a question will get 'locked' (for 4 hours if the price is

[1] See Bajari and Hortacsu (2004) for a survey of empirical studies of the eBay reputation mechanism.

below \$100, for 8 hours if above). This means a question is actively worked on by a researcher and no other researcher can try to answer it.

According to Google all researchers are tested to ensure that they are expert searchers with excellent communication skills. Some of them also have expertise in various fields. Thus, questions may be answered by an expert in a particular field or by an expert searcher. Additionally, answers are edited by Google to ensure quality. Researchers are independent contractors who are paid for posting answers to the site.

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.

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

2.2 The relation between users and researchers

Google Answers users ask questions and Google Answers researchers try to answer them in return for a fixed price and a possible tip. Users are looking for a piece of information and they value the answer to their question in some way. However, their cost to find the information themselves is relatively high.[2] They set a price below that cost and post the question.

There is a pool of researchers who compete to answer questions. Researchers will have an idea of the minimum amount of money they want to earn in a given time span to make answering questions worthwhile. This reservation utility can be based on the average return of their previous answers or simply on their alternative source of income outside of Google Answers. If the expected utility from answering a given question on offer is lower than their reservation utility, then the researcher will decide not to answer. If answering the question is not attractive to any researcher out of the pool it will expire after 30 days.

The first researcher who believes the expected utility of the question exceeds his reservation utility will lock the question and will start working on the answer. The actual quality of the answer depends on the effort of the researcher. Consequently, the value of the answer to the researcher depends on the quality of the answer and therefore on the effort put in by the researcher. Without

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

a doubt luck plays also a role in finding the requested information online and finding what the user particularly values. Nevertheless, the researcher's effort positively affects his answer's value to the user.

If this generated value does not even reach the price the user paid, she will likely ask for a clarification of the answer and possibly give a bad rating (if she bothers to). She might even reject the answer and demand a refund,[3] if the answer is very disappointing. If the value produced by the answer exceeds the price, the user reaps a benefit. She might still ask for a clarification of the answer and tends to give a positive rating to acknowledge the good work of the researcher.

Moreover, users can give a tip to researchers to show their appreciation. Naturally, this will only happen, if there is a benefit out of the transaction: The actual value must be higher than the price. Only then might users be willing to give a tip. Whether they share the surplus created or not depends on their preferences. A strictly self-interested user will clearly keep the surplus to herself. However, a user with social preferences might give a tip. She has concerns for social welfare instead of pursuing only her private interests and wants to participate the researcher in the surplus, particularly since the researcher created the surplus in the first place. She also cares about reciprocity and returns kind behaviour.

What else makes users tip apart from social preferences? Nothing really, if we only consider "one-shot game" relationships. However, users might ask more than just one question over time. For these frequent users reputation concerns could matter. Frequent users of the service could have an interest in building up a reputation of appreciating good effort and acknowledging it with a tip. This way their questions should attract able researchers who recognise them as generous and will deliver good work. They would aim to encourage researchers to deliver good answers. Thus, they benefit from high quality work.

On the other hand, single users or very infrequent users would not have these incentives to tip out of reputation concerns. There is no network effect for them to take into account or it is possibly too small.

Both types of users do care about the outlined reputation concerns. However, only the type with social preferences will tip out of social welfare considerations or reciprocity. However, they will only do so, if the answer creates a surplus or to acknowledge kind behaviour. Naturally, the purely self-interested type does not have these concerns and will never tip because of that.

3 Model

We model the relationship between a user and a researcher in a Principal-Agent framework with moral hazard. The user posts a question and the researcher answers it. The price is set ex ante by the user. The quality of the answer depends on the effort of the researcher, which is not observable. The user's

[3]However, answers have to be very faulty that a refund is granted. They are very rare. Our sample of 6,853 observations contains only 8 cases where the price was returned.

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.

In the standard approach of contract theory the principal implements efficient, second-best contracts based on incentive compatibility and participation constraints. The set up of Google Answers allows for an alternative contract design. Endogenous incomplete contracts are deliberately left open on both sides. The two agents have the opportunity to respond to the action of the other. Thus, both sides are encouraged to reciprocate. Fairness and reciprocity can also be regarded as the enforcement device of this contract. The fact that in our model a tip can always be given adds this feature to the contract design.

Therefore, our model consists of three stages. First, the user u posts her question and sets the price p. Then, the researcher r who accepted the question chooses his effort level e, works on the answer and posts it back to the user who gets the value $v(e)$ out of the answer. Finally the user has the opportunity to reward the researcher with a tip t. He can also punish him by rejecting the answer which is regarded as the fine f.

We follow Fehr, Gächter and Kirchsteiger (1997) in their approach. They 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 treatment and a contract that gives the opportunity for mutual reciprocity was found to improve efficiency. Fehr, Klein and Schmidt (2001) is another paper that applies endogenous incomplete contracts.

In order to explain social preferences we do not give up the assumption that individuals maximise their utility. We merely allow their utility to reflect social concerns as well. It matters to them as well how much other individuals receive. A few theoretical approaches exist and they are surveyed in Fehr and Schmidt (2003) and Camerer (2003): altruism, inequality aversion and intentions-based approaches. We focus on concerns for reciprocity to model the behaviour of individuals with social preferences. However, this is just one possible theoretical approach and it appears fair to assume that they all play their part to explain social preferences. A more general model would have to combine aspects of the different theories. This has also been done, but for the scope of our model we concentrate on the intentions-based approach. We integrate reciprocity based on the seminal work of Rabin (1993) for normal form games and its extension Dufwenberg and Kirchsteiger (2004) for extended form games.

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.[4] For agent j it is given by:

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

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:

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

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

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

Incorporating kindness and the beliefs about it gives us 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. Once an action of a player has taken place, beliefs involving randomisation about this action are replaced by pure choice beliefs. The individuals apply Bayesian updating. We will simplify notation in this case by skipping the index, e.g. b_{ij} becomes b, if j has already made his choice.

The price is pre-set by the user in our analysis and we focus on the choice of effort and tip. Two factors determine the effort level of the agent: the time he spends on answering the question and the number of words in his answer. The higher the effort, the less time it takes him (for a given task) and/or the more words the answer contains (for a given time). The cost of effort is linear and is simply expressed by the effort level itself. We simplify and allow only a limited discrete choice of effort. No effort at all means the researcher shirks ($e^0 = 0$), a low effort level is denoted by e^l and a high effort level by e^h. We assume that there is no randomness in our model. The value of an answer depends solely on the effort level with the following functional properties: $v = v(e), v' > 0, v'' < 0, v(e^h) - e^h > v(e^l) - e^l > v(0)$. This guarantees that a high

[4] We use the average in our analysis 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.

effort level is efficient. As explained earlier the agent knows his income from an alternative activity (within or outside of Google Answers). He will compare this reservation utility to his expected utility from answering the question, which depends negatively on the effort and positively on the price and the expected tip. We assume a binary choice of the tip. It is zero if no tip is given and it is τ when there has been a tip.

Assumption 1 The effort level of the researcher can be low, denoted by e^l, or high, denoted by e^h. No effort is denoted by e^0.

Assumption 2 There is either no tip given by the user or a tip denoted by τ.

These are the 'material' payoffs of the sequential game for the user and the researcher. Its structure is equal to the sequential prisoner's dilemma game. Recall that researchers receive only 75% of the actual price of a question:
low effort, with tip:

$$\Pi_u^{lt} = v(e^l) - p - \tau, \ \Pi_r^{lt} = \frac{3}{4}p + \tau - e^l \tag{5}$$

low effort, no tip:

$$\Pi_u^{lnt} = v(e^l) - p, \ \Pi_r^{lnt} = \frac{3}{4}p - e^l \tag{6}$$

high effort, with tip:

$$\Pi_u^{ht} = v(e^h) - p - \tau, \ \Pi_r^{ht} = \frac{3}{4}p + \tau - e^h \tag{7}$$

high effort, no tip:

$$\Pi_u^{hnt} = v(e^h) - p, \ \Pi_r^{hnt} = \frac{3}{4}p - e^h \tag{8}$$

The following assumptions are necessary to keep the payoffs of our game in line with the sequential prisoner's dilemma.

Assumption 3 When low effort is put in and no tip is given the user's payoff equals zero: $v(e^l) - p = 0$

Assumption 4 With high effort the payoff of the user is positive even when a tip is given: $v(e^h) - p - \tau > 0$

Assumption 5 Putting in high effort is profitable for the researcher when he receives a tip: $\tau - e^h > -e^l$

The general participation constraint the principal has to consider follows. Price plus expected tip minus effort has to exceed the reservation utility. The probability with which the researcher expects a tip is denoted by μ_0.

$$U_r(\Pi_r) = \frac{3}{4}p + \mu_0 \cdot \tau - e > U^R \tag{9}$$

3.1 Single Users

3.1.1 Benchmark model with self-interested individuals

We first analyse the case of single users in a benchmark model under the self-interest hypothesis. The exogenous fine f enforces a low effort level. The rational researchers will not put in more effort than necessary, just enough to avoid getting fined. The fine occurs with probability q when the researcher is lazy and his effort is zero.

The researcher's effort is determined by the following incentive compatibility constraints.

Researchers prefer putting in low instead of no effort in order to avoid getting punished with the fine:

$$\Pi_r^l = \frac{3}{4}p - e^l > (1-q)\frac{3}{4}p - qf - e^0 = \Pi_r^0 \tag{10}$$

When they put in high effort, researchers expect a tip with probability μ_0. However, if researchers expect users to be selfish μ_0 equals 0 and researchers will always put in low effort:

$$\Pi_r^l = \frac{3}{4}p - e^l > \mu_0(\frac{3}{4}p + \tau - e^h) + (1-\mu_0)(\frac{3}{4}p - e^h) = \Pi_r^h \text{ if } \mu_0 = 0 \tag{11}$$

The participation constraint of the principal is equal to the incentive compatibility constraint:

$$\Pi_r^l = \frac{3}{4}p - e^l > U^R \tag{12}$$

Selfish users will never tip as this only reduces the profit of the user. The resulting utility for a user with self-interest preferences is her 'monetary' payoff:

$$\Pi_u = v(e^l) - p \tag{13}$$

3.1.2 Mixed population with self-interested and socially minded individuals

We now include individuals with social preferences in the user population. As explained earlier we focus on concerns for reciprocity to model the behaviour of socially minded individuals. Their utility function increases not only in their material payoffs but also in the psychological payoffs which depend on their kindness and their beliefs about the kindness of the other individual towards themselves.

The equitable payoff of an individual is the average of his best and worst outcome based on the choices of the other individual. In the case of the user u it is given by:[5]

$$\Pi_u^e(a_r, b_{ru}) = \frac{1}{2}(\max\{\Pi_u(a_r, b_{ru})\} + \min\{\Pi_u(a_r, b_{ru})\}) \tag{14}$$

[5] The equitable payoff of u depends on the actual effort choice of r (a_r) and whether r believes u tips or not (b_{ru}).

The best payoff for the user is the result of a high effort choice of the researcher, the worst payoff results when r puts in low effort. The average of the two is the equitable payoff of the user. This serves as a reference point for how kind the researcher is to the user. Recall that assumption 3 says $v(e^l) - p = 0$:[6]

$$\Pi_u^e(a_r, b_{ru}) = \frac{1}{2}((v(e^h) - p - t_{ru}) + (v(e^l) - p - t_{ru})) = \frac{1}{2}((v(e^h) - p) - t_{ru} \quad (15)$$

We calculate the equitable payoff for the researcher r in the same way. He receives his highest payoff, if the user gives a tip and his worst when she does not:[7]

$$\Pi_r^e(a_u, b_{ur}) = \frac{1}{2}((\frac{3}{4}p + \tau - e) + (\frac{3}{4}p + 0 - e)) = \frac{3}{4}p + \frac{1}{2}\tau - e \quad (16)$$

In order to determine how kind or not individuals are, we relate the actual payoff they give to the other player's equitable payoff. The kindness functions of u and r towards each other are:

$$\kappa_{ur}(a_u, b_{ur}) = \Pi_r(a_u, b_{ur}) - \Pi_r^e(a_u, b_{ur}) \quad (17)$$

$$\kappa_{ru}(a_r, b_{ru}) = \Pi_u(a_r, b_{ru}) - \Pi_u^e(a_r, b_{ru}) \quad (18)$$

Similarly u's belief about the kindness of r to u is:

$$\widetilde{\kappa}_{uru}(b_{ur}, \widetilde{b}_{uru}) = \Pi_u(b_{ur}, \widetilde{b}_{uru}) - \Pi_u^e(\widetilde{b}_{uru}) \quad (19)$$

Finally the perceived kindness of u by r is:

$$\widetilde{\kappa}_{rur}(b_{ur}, \widetilde{b}_{rur}) = \Pi_r(b_{ru}, \widetilde{b}_{rur}) - \Pi_r^e(\widetilde{b}_{rur}) \quad (20)$$

Incorporating kindness and the beliefs about it gives us the following utility function with the material payoff Π as the first term and the reciprocity payoff in the second term that is weighted by α, the individual's sensitivity to reciprocity.

$$U_u(a_u, b_{ur}, \widetilde{b}_{uru}) = (v(e) - p - t) + \alpha_u \cdot \kappa_{ur}(p, t, e) \cdot \widetilde{\kappa}_{uru}(p, t_{uru}, e_{ur})$$

$$U_r(a_r, b_{ru}, \widetilde{b}_{rur}) = (\frac{3}{4}p + t - e) + \alpha_r \cdot \kappa_{ru}(p, t_{ru}, e) \cdot \widetilde{\kappa}_{rur}(p, t_{ru}, e_{rur}) \quad (21)$$

The utility of user and researcher does not only depend on their material payoff. The reciprocity payoff is added. Essentially utility is increased by reciprocity when the sign of an individual's kindness κ matches the sign of the perceived kindness of the other individual ($\widetilde{\kappa}$). Both are negative when the individuals behave unkind to each other. This nasty or negative reciprocity equilibrium of the game is when the researcher decides to put in a low effort

[6] We now replace the researcher's action by the chosen effort level (e^h or e^l) and his belief about u's action (tipping or not) by t_{ru}.

[7] Now b_{ur} – the user's belief of the researcher's action – is already known and we only write e. The user's action a_u is either τ or 0.

and the user subsequently gives no tip. This is also the outcome when purely self-interested individuals play as shown above.

Does a positive reciprocity equilibrium exist and what are the conditions for that? Note that the equitable payoff of r when he puts in high effort is:

$$\Pi_r^e(a_u, b_{ur}) = \frac{1}{2}((\frac{3}{4}p + \tau - e^h) + (\frac{3}{4}p - e^h)) = \frac{3}{4}p + \frac{1}{2}\tau - e^h \quad (22)$$

Thus, the user's kindness when she tips (and the researcher put in high effort) is therefore:

$$\kappa_{ur}^t(p, t, e = e^h) = (\frac{3}{4}p + \tau - e^h) - (\frac{3}{4}p + \frac{1}{2}\tau - e^h) = \frac{1}{2}\tau > 0 \quad (23)$$

While if she does not tip (and the researcher put in high effort) the kindness of u is:

$$\kappa_{ur}^{nt}(p, t, e = e^h) = (\frac{3}{4}p - e^h) - (\frac{3}{4}p + \frac{1}{2}\tau - e^h) = -\frac{1}{2}\tau < 0 \quad (24)$$

In order to determine how kind u believes r is after putting in high effort, we need to specify u's belief of what r believes is u's choice after the high effort decision. This second order belief $\widetilde{b}_{uru}$ can be either 0 or τ and we assign the probability $\widetilde{\theta} \in [0; 1]$ to u's belief of "tip". The payoff of u resulting from r's low effort choice is assumed to be zero (the nasty equilibrium where no tip follows low effort) and the payoff u believes r intends to give to u when he chooses high effort is:

$$\Pi_u^h(a_u, b_r, \widetilde{b}_{uru}) = \widetilde{\theta} \cdot (v(e^h) - p - \tau) + (1 - \widetilde{\theta}) \cdot (v(e^h) - p) \quad (25)$$

The equitable payoff for u is the average of these two payoffs (Π_u^h and the certain payoff of zero in the nasty equilibrium):

$$\Pi_u^e(a_u, b_r, \widetilde{b}_{uru}) = \frac{1}{2}((\widetilde{\theta} \cdot (v(e^h) - p - \tau) + (1 - \widetilde{\theta}) \cdot (v(e^h) - p)) + 0) \quad (26)$$

Therefore, the believed kindness of r towards u after choosing high effort is:

$$\widetilde{\kappa}_{uru}(b_{ur}, \widetilde{b}_{uru}) = \frac{1}{2}(\widetilde{\theta} \cdot (v(e^h) - p - \tau) + (1 - \widetilde{\theta}) \cdot (v(e^h) - p)) \quad (27)$$

We are now in a position to calculate the utility of u when she tips and when she does not.

$$U_u^t(a_u, b_{ur}, \widetilde{b}_{uru}) = \Pi_u^t + \alpha_u \cdot \kappa_{ur}^t(p, t, e) \cdot \widetilde{\kappa}_{uru}(p, t_{uru}, e_{ur}) \quad (28)$$

$$U_u^t = (v(e^h) - p - \tau) + \alpha_u \cdot \frac{1}{2}\tau \cdot \frac{1}{2}(\widetilde{\theta} \cdot (v(e^h) - p - \tau) + (1 - \widetilde{\theta}) \cdot (v(e^h) - p)) \quad (29)$$

$$U_u^{nt}(a_u, b_{ur}, \widetilde{b}_{uru}) = \Pi_u^{nt} + \alpha_u \cdot \kappa_{ur}^{nt}(p, t, e) \cdot \widetilde{\kappa}_{uru}(p, t_{uru}, e_{ur}) \quad (30)$$

$$U_u^{nt} = (v(e^h) - p) + \alpha_u \cdot (-\frac{1}{2}\tau) \cdot \frac{1}{2}(\widetilde{\theta} \cdot (v(e^h) - p - \tau) + (1 - \widetilde{\theta}) \cdot (v(e^h) - p)) \quad (31)$$

The condition for the existence of a fair, positive reciprocity equilibrium is:

$$U_u^t(a_u, b_{ur}, \widetilde{b}_{uru}) > U_u^{nt}(a_u, b_{ur}, \widetilde{b}_{uru}) \tag{32}$$

This is fulfilled if:

$$\alpha_u \cdot \frac{1}{2}(\widetilde{\theta} \cdot (v(e^h) - p - \tau) + (1 - \widetilde{\theta}) \cdot (v(e^h) - p)) > 1 \tag{33}$$

Since in equilibrium beliefs must be correct, it follows that the condition must hold for $\theta = 1$, if in equilibrium r puts in high effort. $\theta = 1$ means the belief is a tip will be given which has to be fulfilled in equilibrium.

$$\alpha_u > \frac{2}{v(e^h) - p - \tau} = \overline{\alpha}_u \tag{34}$$

Conversely the condition must not hold for $\theta = 0$ so that in equilibrium u decides against giving a tip. $\theta = 0$ implies that r does not believe a tip will be given. The user will have to prefer not giving a tip ($U_u^{nt} > U_u^t$) so that in equilibrium beliefs are correct.

$$\alpha_u < \frac{2}{v(e^h) - p} = \underline{\alpha}_u \tag{35}$$

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 $\alpha_u < \underline{\alpha}_u$ corresponds to the nasty equilibrium.

A tip follows high effort, if the value of the answer is $\frac{2}{\alpha}$ more than price plus tip. In other words, if the fairness weight α is very small it seems very implausible that a tip will be given and even if an individual has reciprocity concerns the generated value of an answer has to exceed the price paid by more than just the tip. On the other hand, no tip will be given in equilibrium, if the value of the answer exceeds the price paid by less than $\frac{2}{\alpha}$.

For intermediate values $\frac{2}{v(e^h)-p} < \alpha_u < \frac{2}{v(e^h)-p-\tau}$ the user will randomise with probability $\widehat{\theta} = \frac{v(e^h)-p}{\tau} - \frac{2}{\tau \cdot \alpha_u}$. The utility of giving a tip (U_u^t) is equal to the utility of not giving a tip (U_u^{nt}), when u choses to tip with probability $\widehat{\theta}$ for intermediate values of α_u.

Since we have established conditions for u to give a tip once r has put in high effort, we now have to analyse whether r will ever put in high effort in the first place. He knows that the user will never tip when $\alpha_u < \frac{2}{v(e^h)-p}$ and therefore he will never put in high effort in that case.

He also knows that u will act reciprocally once her sensitivity to reciprocity α_u is large enough. That means he assumes u will reward a choice of high effort with a tip and will reply to a low effort with no tip. The equitable payoff of the user is therefore the average of the high effort plus tip (Π_u^{ht}) and the low effort without tip (Π_u^{lnt}) payoffs:

$$\Pi_u^e(a_r, b_{ru}) = \frac{1}{2}((v(e^h) - p - \tau) + (v(e^l) - p)) = \frac{1}{2}(v(e^h) - p - \tau) \tag{36}$$

The choice of high effort means the kindness of researcher to user is:

$$\kappa_{ru}^{h}(a_r, b_{ru}) = \frac{1}{2}(v(e^h) - p - \tau) \tag{37}$$

Accordingly the researcher's kindness to the user with low effort is:

$$\kappa_{ru}^{l}(a_r, b_{ru}) = -\frac{1}{2}(v(e^h) - p - \tau) \tag{38}$$

Once again we have to specify second order beliefs to calculate how kind r thinks u is towards him. R's belief of what u believes is r's choice is $\widetilde{b}_{rur}$. We assign the probability $\widetilde{\eta} \in [0;1]$ to r's second order belief of high effort ($\widetilde{\eta} = 1$ means the belief is high, $\widetilde{\eta} = 0$ means it is low). In order to find r's equitable payoff we take the average of the best and worst outcome for him. The best is when the user always tips following the two options of the researcher (low and high effort weighted by η), the worst is when the user never tips.

$$\Pi_r^e(a_u, b_{ur}, \widetilde{b}_{rur}) = \frac{1}{2}((\widetilde{\eta} \cdot (\frac{3}{4}p + \tau - e^h) + ((1 - \widetilde{\eta}) \cdot (\frac{3}{4}p + \tau - e^l)) + \\ (\widetilde{\eta} \cdot (\frac{3}{4}p - e^h) + ((1 - \widetilde{\eta}) \cdot (\frac{3}{4}p - e^l))) \tag{39}$$

Hence, the belief of r about u's kindness from choosing in a reciprocal way is the actual[8] minus the equitable payoff of r:

$$\widetilde{\kappa}_{rur}(b_{ur}, \widetilde{b}_{rur}) = \frac{1}{2}(\widetilde{\eta} \cdot (\frac{3}{4}p + \tau - e^h) + (1 - \widetilde{\eta}) \cdot (\frac{3}{4}p - e^l) - \\ (1 - \widetilde{\eta}) \cdot (\frac{3}{4}p + \tau - e^l) - \widetilde{\eta} \cdot (\frac{3}{4}p - e^h)) \tag{40}$$

$$\widetilde{\kappa}_{rur}(b_{ur}, \widetilde{b}_{rur}) = \widetilde{\eta} \cdot \tau - \frac{1}{2}\tau \tag{41}$$

The utility of the researcher when he puts in high effort (and expects the tip) is:

$$U_r^h(a_r, b_{ru}, \widetilde{b}_{rur}) = \Pi_u^{ht} + \alpha_r \cdot \kappa_{ru}^h(a_r, b_{ru}) \cdot \widetilde{\kappa}_{rur}(b_{ur}, \widetilde{b}_{rur}) \tag{42}$$

$$U_r^h(a_r, b_{ru}, \widetilde{b}_{rur}) = (\frac{3}{4}p + \tau - e^h) + \alpha_r \cdot \frac{1}{2}(v(e^h) - p - \tau) \cdot (\widetilde{\eta} \cdot \tau - \frac{1}{2}\tau) \tag{43}$$

When he chooses low effort (and expects no tip) it is:

$$U_r^l(a_r, b_{ru}, \widetilde{b}_{rur}) = \Pi_u^{lnt} + \alpha_r \cdot \kappa_{ru}^l(a_r, b_{ru}) \cdot \widetilde{\kappa}_{rur}(b_{ur}, \widetilde{b}_{rur}) \tag{44}$$

$$U_r^l(a_r, b_{ru}, \widetilde{b}_{rur}) = (\frac{3}{4}p - e^l) + \alpha_r \cdot (-\frac{1}{2}(v(e^h) - p - \tau)) \cdot (\widetilde{\eta} \cdot \tau - \frac{1}{2}\tau) \tag{45}$$

[8] This is the reciprocity equilibrium with high effort plus tip or low effort and no tip.

The condition for researchers to make the high effort decision is:

$$U_r^{e^h}(a_r, b_{ru}, \widetilde{b}_{rur}) > U_r^{e^l}(a_r, b_{ru}, \widetilde{b}_{rur}) \tag{46}$$

This is fulfilled if:

$$\alpha_r \cdot (v(e^h) - p - \tau) \cdot \widetilde{\kappa}_{rur}(b_{ur}, \widetilde{b}_{rur}) > e^h - \tau - e^l \tag{47}$$

Since in equilibrium beliefs must be correct, it follows that the condition must hold for $\eta = 1$, if in equilibrium r puts in high effort. This is always true since all terms on the left hand side are non-negative and the right hand side is negative. The sensitivity to reciprocity α is non-negative by definition, the second term is positive based on assumption 4 and $\widetilde{\kappa}_{rur}$ – the perceived kindness of u – equals $\frac{1}{2}\tau > 0$ for $\eta = 1$. Finally, it follows from assumption 5 that the right hand side is negative. The intuitive equilibrium of high effort (and beliefs about that) followed by a tip results.

Conversely the condition must not hold for $\eta = 0$ so that in equilibrium r decides to put in low effort:

$$\alpha_r > \frac{\tau + e^l - e^h}{(v(e^h) - p - \tau)\frac{1}{2}\tau} \tag{48}$$

This condition outlines the possibility of a "self-fulfilling expectations" equilibrium that is reached when the researcher's sensitivity to reciprocity is large enough.[9]

Our analysis of the researcher's decision has now covered the case when the user's sensitivity to reciprocity is not high enough ($\alpha_u < \underline{\alpha}_u$ means the user will never tip and therefore the researcher will always put in low effort) and the case when she is sufficiently motivated by reciprocity ($\alpha_u > \overline{\alpha}_u$ means the user will reciprocate and the researcher will therefore put in high effort unless his sensitivity to reciprocate is too high). The case of intermediate values of $\underline{\alpha}_u > \alpha_u < \overline{\alpha}_u$ requires further analysis. However, there are no qualitatively new results to obtain. If the user's randomising probability of tipping θ is high enough, the researcher will put in high effort when he is sufficiently motivated by reciprocity. "Self-fulfilling expectations" equilibria are possible once again when the user's sensitivity to reciprocate is too high.

The model explained when single users leave a tip. Social preferences are necessary which we incorporated into the utility function with a reciprocity payoff. Once reciprocity gains outweigh the material loss of paying the tip, users will prefer to give a tip. However, users and researcher have to be sufficiently motivated by reciprocity for this to happen. Our first hypothesis follows which will be empirically tested with our data.

[9] When r thinks that u believes he puts in low effort ($\eta = 0$), he assumes u will not tip in that case (since he assumes u reciprocates). That makes r believe u is unkind. He would put in low effort then.

This can only happen when r's reciprocity motivation is really high as it needs to offset the material gain plus the gain from u's reciprocal behaviour when r puts in high effort.

Hypothesis 1: The population of users is indeed mixed and contains individuals with self-interest and as well with social preferences. Therefore, the average tip of single users is significantly greater than 0.

3.2 Frequent Users

We now turn our attention to users who repeatedly ask questions. The total amount of a user's questions is finite and we denote it with T. The current question number of the user is marked with the index n. We assume that users know the total amount of questions they will ask in the future and are thus able to calculate future profits.

Moreover, we assume that researchers can observe the previous tipping behaviour of users. They can also evaluate whether the effort of other researchers has been high or low in order to tell if giving no tip by the user was justified or not. That means researchers will update their beliefs about the previous (and now realised) tipping behaviour of the user they face. We denote the tip history of a user at question number n by μ_n. Users are not able to observe the effort history of the researcher.[10]

When there is no tip history available, researchers have an uninformed belief μ_0 regarding the chance a specific user will tip. It is determined by the number of users r assigns to the four different groups that are possible: The ones with social preferences (group A), self-interested single users (group B), self-interested frequent users who decide to tip to build up or maintain their reputation (group C) and self-interested frequent users who decide not to tip, because investing in reputation does not pay off (anymore) (group D). Users from groups A and C will tip, users from groups B and D will not. Therefore, the uninformed belief is the ratio of users who will tip (given high effort):

$$\mu_0 = \frac{A + C}{A + B + C + D} \tag{49}$$

When researchers observe that their user has tipped the previous question, they know she either has social preferences (and will always reciprocate high effort answers) or she is a frequent user who tipped out of reputation concerns. But this frequent user could change strategy and switch from tipping to not tipping, because she has reached her break even point in the meantime. Researchers take that into account by updating their belief in the following way:

$$\mu_n^t = \frac{A + C}{A + C + D} > \mu_0$$

When the researcher observed his user has tipped the last answer, the user cannot be of the non-tipping group B type. The updated belief of the probability

[10] We assume that the previous effort of researchers is not observable to users. If users could take this into account the reciprocity effects we explain would only be larger. However, it is not realistic to assume that users would make the effort to check the previous answers of a researcher in order to tip him for that.

a user will leave a tip increases. On the other hand, when he observes the user did not tip he concludes she has changed strategy and stopped tipping or has never tipped at all.

$$\mu_n^{nt} = 0$$

3.2.1 With self-interest preferences

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. If a researcher meets a user who has tipped each of her previous questions he could fairly assume that a tip will be given again if the quality of his answer is right. On the other hand, if he faces a user who is not known for rewarding high effort in the past, he will probably not choose high effort, since chances are very low that he will be tipped based on the user's previous behaviour. In particular, he will not believe the user is going to tip, if she did not leave a tip for the last (high effort) answer. The decision rule for the researcher follows, he chooses e^h if and only if:

$$\Pi_r^h = \mu_n(\frac{3}{4}p + \tau - e^h) + (1 - \mu_n)(\frac{3}{4}p - e^h) > (\frac{3}{4}p - e^l) = \Pi_r^l \tag{50}$$

$$\mu_n > \frac{e^h - e^l}{\tau} = \overline{\mu} \tag{51}$$

The Bayesian updating of users' past tipping behaviour reduces the uncertainty the researchers face. If they are able to inform themselves about the user's past 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 (by self-interested users who take reputation into account or by users with social preferences). They will put in low effort, if chances for receiving a tip from the user they face are too low. When the updated belief μ_n is greater than the critical value $\overline{\mu}$, the researcher will decide to put in high effort. In case of $\mu_n^{nt} = 0$ equations (50) and (51) do not hold and the researcher chooses low effort.

If rational, selfish users expect to use the service frequently in the future, they take the researcher's updating into account and might tip out of reputation concerns. By tipping high quality answers they create a good reputation and encourage high effort answers in the future. This might pay off in the long run, even if they give away money in the short run. They will compare their future payoffs when they consistently tip and are more likely to receive high effort answers to the future payoffs of low quality answers and no tips given. A user posts T questions in total and each one can be with a different researcher. A rational user who knows that she returns for more questions will take this into account.

We assume that the users' tip history μ is the prior belief of the researcher they face. If the prior belief regarding the probability of whether u tips exceeds the critical value of the researcher ($\overline{\mu}$), then he will put in high effort. Thus, researchers are expected to realise that they are getting encouraged to put in high effort and are getting rewarded for it ex post. In other words, they must update their beliefs about the tipping behaviour of users and decide to put in high or low effort based on this information. Only then does this strategy work for users.

The profit in the current period ($\Pi_{u,n}$) depends on whether u tips the high quality answer or not. The profit in the following periods is also affected by the tipping decision. If a tip has been given and the belief of the tip history exceeds the critical value $\overline{\mu}$, the user benefits from high effort answers. She receives $\Pi^h_{u,m} = v(e^h) - p - t$ as payoff in period $m = n+1$ where the tip t can be zero or τ. This happens with probability $\gamma^t_n = \Pr(\mu^t_{n-1} > \overline{\mu})$. Otherwise, she only gets a low effort answer with payoff $\Pi^l_{u,m} = v(e^l) - p = 0$. When the previous answer has not been tipped, the user cannot expect high effort answers anymore. The probability for that is $\gamma^{nt}_n = \Pr(\mu^{nt}_{n-1} > \overline{\mu}) = \mu^{nt}_{n-1} = 0$. Therefore, the decision to tip depends on (future profits are discounted by the factor $0 < \delta < 1$):

$$\Pi^t_u = \Pi^t_{u,n} + \sum_{m=n+1}^{T} \delta^{m-1}(\gamma^t_m \cdot \Pi^h_{u,m} + (1 - \gamma^t_m) \cdot \Pi^l_{u,m}) \tag{52}$$

$$\Pi^{nt}_u(\Pi^{nt}_{u,n}, t, T, \mu_t) = \Pi^{nt}_{u,n} + \sum_{m=n+1}^{T} \delta^{m-1}(\gamma^{nt}_m \cdot \Pi^h_{u,m} + (1 - \gamma^{nt}_m) \cdot \Pi^l_{u,m}) \tag{53}$$

The second term in the summation drops out in both equations since the low effort profit Π^l is zero. Future profits after no tip has been given get eliminated as well since $\gamma^{nt}_t = 0$. The user decides to tip as long as her discounted profits from tipping exceed the present profit if she does not tip:

$$\Pi^t_u = (v(e^h) - p - \tau) + \sum_{m=n+1}^{T} \delta^{m-1} \cdot \gamma^t_m \cdot (v(e^h) - p - t) > (v(e^h) - p) = \Pi^{nt}_u \tag{54}$$

$$\sum_{m=n+1}^{T} \delta^{m-1} \cdot \gamma^t_m \cdot (v(e^h) - p - t) > \tau \tag{55}$$

Tipping is profitable if and only if the increase in future payoffs exceeds the one-off tip. Essentially, leaving a tip has to generate enough good reputation that sufficiently many questions in the future are answered with high effort. The frequent user's decision to tip or not depends on the number of remaining questions, the discount factor and the surplus from high quality answers. How many questions remain and how much these potential future profits are valued by the user affect the size of the discounted profits. When the user expects to ask many more questions in the future, she will benefit often from high quality answers. Then tipping makes sense as long as she does not discount these future profits too much. Naturally, the relation between the surplus from high quality answers and the tip matters to determine profits, but this ratio is exogenous in our model.

The decision to tip or not might change over time. If the current question number n approaches the total amount T a user intends to ask, then the remaining questions might not be enough to justify tipping any longer. The incentive to encourage future high quality answers decreases with the number of questions remaining and at some point the cost of tipping is too high. A rational strategy for the user is therefore to build up a good reputation by tipping high quality answers, but to stop tipping, when she plans to post only few more questions in the near future. This strategic element makes frequent user tipping different from tipping motivated by social preferences.

We can make the following further hypotheses from our model:

Hypothesis 2: Self-interested frequent users do anticipate the benefits from establishing a good reputation and the resulting high quality answers in the future and will use the tip as a strategic contract device. Therefore, the tip rate weakly increases with the number of questions a user asks.

Hypothesis 3: Researchers inform themselves about a user's tip history and update their belief about the probability with which she might tip. That affects their effort decision. Consequently – after enough observations to establish a reputation – the questions of users with a high tip history are answered with higher effort.

3.2.2 With social preferences

It is easy to see that also frequent users with social preferences will tip as long as the conditions from the previous section are fulfilled: The effort of the researcher has to be high, otherwise the user will reciprocate negatively. Moreover, the sensitivity to reciprocity of the user has to be high enough ($\alpha_u > \overline{\alpha}_u$).[11] Since material payoffs are as well part of the users' utility, the reputation concerns outlined in the previous paragraph matter to users with social preferences, too.

Researchers are able to observe the previous tipping behaviour of users and they can also evaluate, whether a tip was not given due to low effort. As explained before that means researchers will update their beliefs about the previous (and now realised) tipping behaviour of the user they face. They will take the kindness of 'their' user towards other researchers into account once they are also motivated by indirect reciprocity.[12] Then the researcher's belief about the kindness of the user $\widetilde{\kappa}_{rur}$ is updated based on the user's previous actions and the researcher will chose high effort, if the user has a good enough track record of tipping and rewarding high effort answers.

The condition for researchers to make the high effort decision is:

$$U_r^h(a_r, b_{ru}, \widetilde{b}_{rur}(\mu_n)) > U_r^l(a_r, b_{ru}, \widetilde{b}_{rur}(\mu_n)) \tag{56}$$

[11] It is possible to imagine that users will tip even though 'their' researcher has put in low effort, if this researcher has put in high effort in his previous answers which could be regarded as kind by the user. However, we do not cover that in the model.

[12] See Seinen and Schram (1999) for an experimental study of indirect reciprocity where observed records of cooperativeness of a player induce others to cooperate with him.

3.3 Summary

In this section we have developed a model that describes the behaviour of users and researchers in an endogenous incomplete contract environment. The basic model with self-interested, single users was extended to integrate social preferences and frequent repeated use. We have analysed all four cases separately in order to show what drives the individual's behaviour in detail. We can also combine them in a general case. The user's decision depends on:

$$U_u^t(\Pi_{u,n}^t, n, T, \mu_t, \alpha_u, \kappa_{ur}^t, \widetilde{\kappa}_{uru}) = \Pi_{u,n}^{ht} + \sum_{m=n+1}^{T} \delta^{m-1} \cdot \gamma_m^t \cdot \Pi_{u,m}^h + \alpha_u \cdot \kappa_{ur}^t(p,t,e) \cdot \widetilde{\kappa}_{uru}(p, t_{uru}, e_{ur}) \quad (57)$$

$$U_u^{nt}(\Pi_{u,n}^{nt}, n, T, \mu_t, \alpha_u, \kappa_{ur}^{nt}, \widetilde{\kappa}_{uru}) = \Pi_{u,n}^{hnt} + \sum_{m=n+1}^{T} \delta^{m-1} \cdot \gamma_m^{nt} \cdot \Pi_{u,m}^h + \alpha_u \cdot \kappa_{ur}^{nt}(p,t,e) \cdot \widetilde{\kappa}_{uru}(p, t_{uru}, e_{ur}) \quad (58)$$

We get the special case of purely self-interested users when we set $\alpha_u = 0$ as then the reciprocity payoff disappears. The material payoff remains which can be affected by reputation concerns. In the case of single users $T = 1$ and future payoffs have no impact on the decision. Only the payoff of the current period matters then.

Hypothesis 4: Users tend to stick to their strategy or preference respectively. The decision to tip depends mainly on the quality of the answer, but only users with social preferences and users who take reputation effects into account do consider tips. Therefore, the distribution of tips is not equal over all users, but concentrated on these two groups.

Hypothesis 5: Endogenous incomplete contracts improve the efficiency of the user-researcher relationship over time. After sufficient observations to reveal the behaviour of users, efficiency increases as researchers put in more effort (which means more quality to users) and they are getting rewarded with a tip.

4 Description of the Data Set

The 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 very specific 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 be random and would therefore not affect our sample.

Our data set starts in July 2003 and ends in January 2004. Within this period of time we collected 13,948 questions, only 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.[13] 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 researcher ID of the person 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 that help to explain the relations. 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 the time it took to answer as well as the word count of the 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	1.8049	0.20	0.03	4.6647	0.0014	29.9917
word count	6853	581.9041	360	94	719.0199	1	11482
answer clarification	6853	0.2934	0	0	0.4554	0	1
effort1 (word count)	6853	59.3155	35.4	25	80.7159	0.0667	1657
effort2 (time)	6853	146.7479	52.25	90	463.3866	0.0667	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 mini-

[13] Since the focus of our analysis is the tipping aspect we decided to deliberately truncate the data set and we consider 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.

mum and maximum price of the sample. The average price conditional of the question being answered (6,853 observations) is $21.59. As one should expect, the average price of the 7,095 questions that have not been answered (and thus expired after 30 days) is lower. It is only $19.23 and this supports the argument that researchers found them less attractive. Median and mode of the price distribution are $10.

As mentioned before 1,745 questions out of the total have been tipped. Again, minimum and maximum values of the tip are pre-set by the system. There is an upper limit for the tip of $100. The mean of the distribution is $8.94 and its median and mode equal $5.

The following diagram shows the distribution of price and tip. Both are very skewed as many observations have lower values than the mean. In order to compare the two distributions we have computed the deciles. The x-axis shows the deciles, the y-axis shows price and tip in percent of the maximum value. Around three quarters of all prices and tips are below 10% of the respective maximum ($200 for the price, $100 for the tip) and the highest decile stretches from approximately 20% to the maximum.

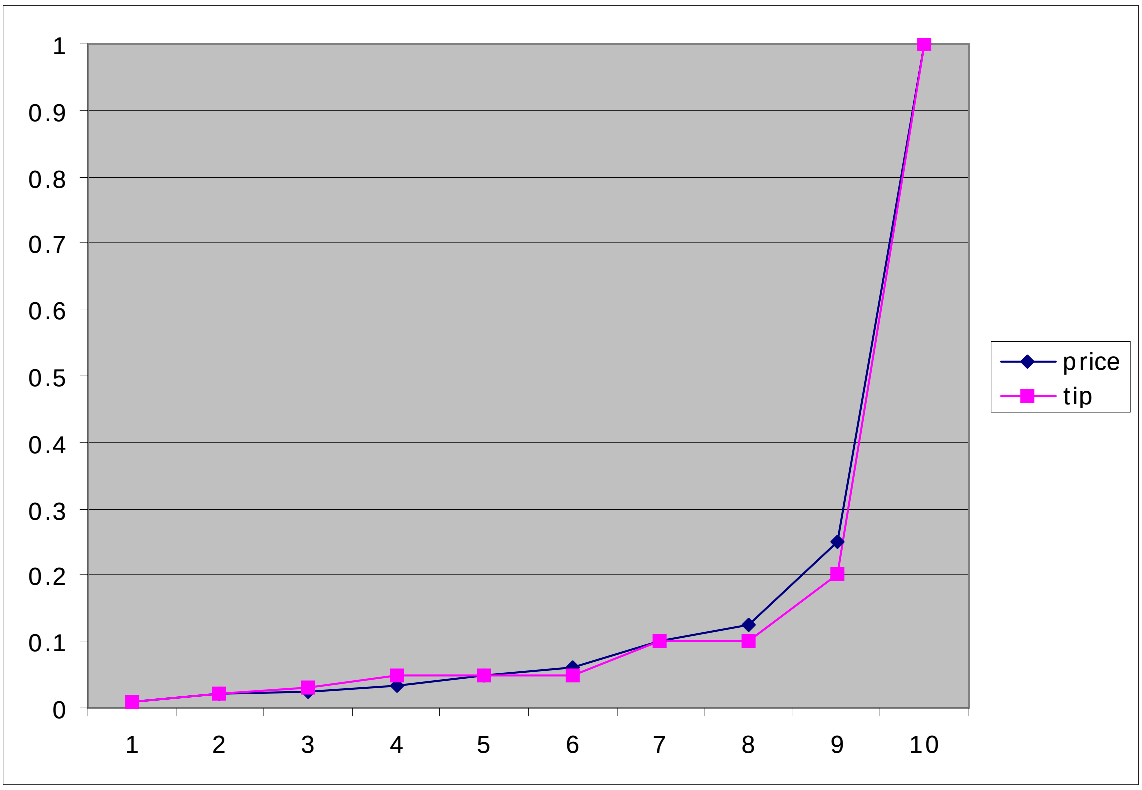

Figure 1: Price and tip deciles

The ratio of tip to price expresses the size of the tip compared to the price of the respective question. It is zero for many observations as no tip was given. On average the tip is 20% of the price, but the variance is very high. The maximum ratio was a tip of 50 times the price. This highest ratio possible (lowest possible price, $2, and highest possible tip, $100) was observed three times. 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, by the way with $65 tip). A rating has been given for 4,359 answers, roughly two thirds of the total. The range is from 1 to 5, with 5 being the top rating. If users do give a rating, they do 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 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 has their effort been. The average words per dollar are 59.3155 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 has their effort 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, unfortunately. 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. This is the case for many questions. On the other hand, 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 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. The pricing and tipping behaviour in the categories appears to be quite diverse. About 31% of all observations in 'Arts & Entertainment' or 'Sports and Recreation' have been tipped. There was the lowest generosity in the 'Business & Money' category as only 21.68% of these questions have been tipped. The highest average price can be found for this category ($34.32). The tip rate of the other categories is fairly close to the overall average of 25.46%. The significance of the category dummies will be tested in Section 5.

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 Set

Before running a regression of the entire data set we focus the analysis on the possible motivations for tipping that we have outlined: conforming to a social norm, because of social preferences or out of reputation concerns. We also cover the main points of the model: the behaviour of single users, the role of reputation and the relationship between updating, effort decision and efficiency.

5.1 Social Norm

Tipping out of a social norm follows a very clear pattern. Everybody who conforms to the norm tips a fairly equal amount for a specific service. In restaurants people would tip roundabout the same percentage of their respective bill. Different social norms have been established as it is common to tip significantly more in the U.S.A. than in Europe for instance, but people generally conform to the norms of the society they live in.

If a similar social norm is therefore the motivation for the tipping in Google Answers we should observe a similar pattern. Many users should tip and the size of their tip 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. It does not seem to be a social norm to do so, at least it is not one that many people follow. It could be argued that the online environment and the fact that there is no direct contact makes less people conform to the norm of tipping, but that it is still the social obligation why they 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. We can exclude that Google Answers users tip, because they follow a social norm.

5.2 Social Preferences

Based on our model we know that reputation concerns influence the tipping behaviour of users. Therefore, we control for them by looking at single users only. This puts us in a position to test our first hypothesis. A tip would never be given in the benchmark model with self-interested single users. However, tips can occur once we mix in users with social preferences.

Over the entire observation period 18% of all single users did tip.[14] Since our data set ends in January, we treat users wrongly 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 is in fact the case as the tip rate of single users over the observation months increases towards the end of the period. However, the rate of single users who tipped never falls below 15%. In October – when 'our' single users have not asked another question three months before and after – it is 17.28% and too high to be tolerated by the self-interest hypothesis.

Since there is a slight upward trend of the tip rate over the months we calculate the ratio of the single user tip rate to the tip rate of all users to accommodate for that. The ratio is the lowest in October which means that we have the fewest tipping single users in this month – as expected.

TABLE 3 : 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

We proceed to check the statistical significance of single user tips. There are 2,942 single users and 529 of them did leave a tip. The mean of the distribution is 1.472 and the standard deviation is 5.791. In order to prove that the frequency and amount of tips are high enough we test the following null hypothesis:

TABLE 4 : HYPOTHESIS TEST OF SIGNIFICANCE

1) H_0: The average tip of single users = 0; H_1: otherwise	
sample	2942
t-statistics	13.79
p-value	0.00
$\therefore H_0$ is rejected	
2) H_0: The average tip to price ratio of single users = 0; H_1: otherwise	
sample	2942
t-statistics	9.58
p-value	0.00
$\therefore H_0$ is rejected	

Both tests' null hypothesis can clearly be rejected. The frequency and amount of tips are statistically significant. After controlling for the impact

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

of reputation concerns we find that tips are still prevalent at a statistically very significant level. The tipping of quite a number of single users confirms our mixed population model and the effect of social preferences summarised in hypothesis 1.

5.3 Reputation Concerns

In order to test the impact of reputation concerns on the tipping behaviour we split our sample into subgroups. We have computed each user's frequency of use. This is the number of questions he posted during the observation period. Some users will not have a clear idea of how often they will 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 of how much 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 problem where reputation becomes irrelevant as it happens in experiments.

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. They asked only one question during the observation period. Almost 18% of them gave a tip, which is far below the 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 already occasional users 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. In fact, the tip rate increases with the frequency of use, which confirms hypothesis 2 of our model.

5.4 Updating, effort decision and efficiency

Our theoretical model explains how researchers 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 their name as the past average like the rating of researchers is for instance.[15] Table 6 splits the sample into different sub groups regarding the question number asked. Essentially we see that the tip rate increases for users who keep on asking questions.

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 not tipped questions. In the intermediate range of question numbers users who did not tip had an average tip history of just 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 have clearly done so in the past, too. On the other hand, users who did not give a tip have a rather low tip history. By and large users appear to have preferences or a strategy to tip (high quality answers) and they stick to it, just as the model predicts in hypothesis 4.

Table 7: Question no. and tip history

question no.	avg. tip history	/ no tip	/ with tip
first	0.00	0.00	0.00
2nd to 9th	0.32	0.19	0.61
10+	0.31	0.16	0.60

If researchers update their beliefs about the chances to get a tip for high effort work, then they should anticipate that and make their effort decision

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

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 (effort1 which is word count-based). When a user asks the first question there is no tip history. The effort decision cannot be based on the user's past. When we compare questions with and without tip we find that effort is slightly higher for the tipped ones. This is just what we should expect since the tip depends on the effort put in by the researcher. More effort leads to a higher chance of a tip.

We have already seen in Table 7 that users tend to stick to their tipping pattern. The tip history appears to be a good indicator of the user's type. Once there are previous questions the researchers can build their beliefs about the user's tendency to tip based on the tip history. 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, 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. The average effort is also higher compared to earlier questions that were tipped. 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 and they reward high effort, if they are sufficiently motivated by reciprocity or reputation concerns. We can confirm the third hypothesis of the model.

Table 8: Question no. and effort1

question no.	avg. effort1	/ no 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

We can not observe such a clear effect for the second effort variable which is time-based. While for first questions tipped answers are worked on with slightly less effort than the ones without tip, in the intermediate range more effort is exerted for tipped ones. With ten or more questions of tip history available the tipped questions have again been worked on with slightly less effort than questions without tip. Overall, effort expressed in time increases when tip history is available, but there is no clear divergence between the effort levels of tipped and untipped answers like for effort1. A reason for this could be the general drawback of that variable as explained earlier.

Table 9: Question no. and effort2

question no.	avg. effort2	/ no tip	/ with tip
first	135.24	138.16	124.26
2nd to 9th	163.48	157.99	174.56
10+	178.40	179.39	175.91

The final question is whether the endogenous incomplete contracts offered by users – or the mere fact that endogenous incomplete contracts are possible in this design – increase efficiency. Does it pay off for researchers to put in high effort when they answer questions of users who are known for tipping? We express the payoff of researchers in dollars per 100 words. When no tip is left, payoff is simply 100 divided by effort1. In the case of a tip, we can re-calculate the researcher's effort by taking into account the tip they received. Then we transform this into the payoff as described.[16]

TABLE 10: PAYOFFS FOR RESEARCHERS

question no.	avg. p if tip	avg. t if tip	payoff with tip	payoff if no tip
first	22.0782	8.2490	2.0869	1.7621
2nd to 9th	25.5155	11.6153	2.1710	1.7097
10+	18.5889	10.0522	2.0618	2.0585

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 payoffs for researchers when tips are given are rather stable in the three sub groups. When their effort increases, the relative tip (tip over price) goes up as well. Users known for tipping get higher effort – and presumably also higher quality – answers than new users, but they also reciprocate and let the researchers participate in the gain from a high quality answers by returning some of the surplus and leaving a high tip. The overall payoff of tipped answers is surprisingly balanced (always slightly more than $2 per 100 words). Researchers seem to get fairly rewarded for their effort with a tip.

First answers without a tip lead to a lower payoff than tipped ones. They are also worked on with less effort (see Table 8), but the tip more than compensates that. This is also true for the intermediate sub group. However, once there is a substantial tip history available the researcher's effort for untipped answers drops down so much that their payoff reaches the level of tipped answers (slightly more than $2 per 100 words). It appears they clearly 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.

The analysis of the researchers' payoffs reveals that they earn more or less the same with their answer. The exception is when they face a user with an unclear tipping history (no or short and few tips) who turns out to give no tip. Researchers who are able to update their beliefs will be better off.

We found out that users can encourage high effort answers when they consistently tip. They benefit from high quality answers. Due to their preferences or their strategy they will continue to tip and return the surplus. Since the tip they give would not exceed the surplus, they are better off as well. Hence, we can confirm hypothesis 5.

The option of an endogenous incomplete contract increases the effort level and the efficiency. It encourages socially minded users to reciprocate (tipping

[16] Payoff equals 100 times word count divided by price and tip.

high effort answers) and 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. This contract type is simulated when users reveal that they are not going to tip (long enough low tip history), researchers update their beliefs accordingly and put in relatively low effort.

6 Estimations

We intend to explain the tip with three different arguments. First of all, reputation matters. Frequent users of the service have an incentive to build up a good reputation and should regard tipping as a strategic device. Moreover, social preferences make people contribute voluntarily. Users who care for social welfare and reciprocate kind behaviour should give a tip as long as there is a surplus from a high quality answer. Finally, the tip is simply affected by the price of the question. Users should tend to give a high tip for a highly priced question and vice versa.

Our proxy of the concern for reputation is the frequency with which a user asked questions over the observation period. A high frequency of use means the user will 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 reputation concerns. We use the logarithmic value of the frequency of use in our regression.

To take account of behaviour that elicits social preferences we use a set of proxies. The effort involved in a given answer shows how hard a researcher worked for the answer. High effort means a high quality answer is more likely and in turn a high value of the answer to the user becomes probable. 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 without actual need and this should trigger positive reciprocity once more. While we are aware that our variables are rather crude surrogates for what motivates voluntary giving, we do believe that they can help to explain the true cause for social preferences.

We meter effort in terms of time and word count relative to the price of the question. A question that has been answered with relatively high effort is more likely to generate value for the user. An answer of average quality that comes very fast might give increased value to 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, high effort increases the probability of more value to the user. Following our reciprocity-based model users with social preferences tend to give a tip, when their question has been answered with high effort. Another way how social preferences can explain the positive effect of effort on tip are concerns for social welfare. High effort leads to a higher surplus for the user and whenever there is a surplus created users who care about social welfare will participate the researcher especially since the increase in value originated from the researcher.

Once a user leaves a rating, it seems fair to assume that she is not purely 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. The rating of researchers is fairly important to them as it helps them if they receive good ratings. 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 – without having to – and this should trigger reciprocal behaviour of the user. Hence, we take the answer clarification as a 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 rating, the answer clarification dummy, the logarithm of the frequency of use, effort1 (word count-based), effort2 (time-based) and the dummies for the categories. Table 11 lists the variables and their coefficients with respective t-values for our estimations.

We first ran an Ordinary Least Squares regression and the estimates are shown in column I of Table 11. 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 II. 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 III and IV, 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 11: Estimation Results

	I: OLS		II: TOBIT		III: PROBIT		IV: truncated	
Explanatory var.	coeff.	t-value	coeff.	t-value	coeff.	t-value	coeff.	t-value
constant k	-2.668	-9.46	-72.9	-23.23	-3.693	-20.66	-806.6	-4.06
price	0.0762	24.31	0.2039	20.68	0.0043	6.30	1.881	5.41
rating	0.8670	19.56	12.33	20.38	0.6671	18.79	92.93	3.38
D_ANSCL	0.966	4.35	3.698	5.33	0.1734	3.98	61.83	3.72
frequency of use	0.7608	4.13	2.232	3.98	0.089	2.55	31.16	2.70
effort1	0.0044	4.48	0.1243	4.67	0.0012	5.14	0.0068	0.17
effort2	0.0008	3.90	0.0032	0.57	-0.00005	-1.47	0.0004	0.14
D_ART	-0.070	-0.18	1.298	1.07	0.1576	2.09	-55.23	-1.97
D_BIZ	-0.7637	-2.25	-1.718	-1.50	-0.0912	-1.29	15.11	0.79
D_COM	0.1742	0.53	-0.392	-0.37	-0.0532	-0.81	-5.493	-0.30
D_FAM	0.0497	0.09	0.388	0.22	0.0086	0.08	-4.021	-0.11
D_HEA	-0.026	-0.06	0.769	0.54	0.0273	0.31	60.71	2.17
D_REF	0.0027	0.01	0.485	0.40	0.0115	0.15	-9.059	-0.43
D_REL	-0.9635	-1.84	-2.272	-1.35	-0.0767	-0.75	-50.53	-1.28
D_SCI	-0.4395	-0.98	-1.600	-1.11	-0.1298	-1.47	-10.04	-0.39
D_SPO	-0.0553	-0.09	3.092	1.60	0.3527	2.88	-125.6	-2.05
sample size	6853		6853		6853		1745	
fit of regression	adj. R^2	0.146	adj. R^2	0.232	pseudo R^2	0.296		
log likelihood	-		-8737.47		-2736.32		-5401.21	

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. It turns out that the price does affect both decisions. It has a significant and positive impact on the tip. The data also confirms the significance of reputation concerns. The estimators for the coefficient of the frequency of use are significant at least at the 5%-level in all regressions. The effect of the word count-based effort is clearly positive as well. We can observe a difference here between the standard Tobit and the two-equation model. While the Probit estimator for 'effort1' is very significant, it is not significant in the truncated regression. It seems high effort only affects the decision to leave a tip or not, it has no 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 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.

With our dummy variables we can analyse, 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 not significantly positive in the truncated regression. In fact, the coefficient for 'Sports and Recreation' is significant, but negative. 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.

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). This is what we have seen in the previous section. 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.

Alternatively, our set of regressions can be run with the tip rate (tip divided by price) as the dependent variable:

$$\frac{t}{p} = k + bX + \epsilon$$

The results are shown in Table 12:

Table 4.12: Estimation Results 2

	I: OLS		II: TOBIT		III: PROBIT		IV: truncated	
Explanatory var.	coeff.	t-value	coeff.	t-value	coeff.	t-value	coeff.	t-value
constant k	-0.5188	-1.25	-0.0519	-1.25	-3.693	-20.66	-2.51	-4.76
rating	0.7080	10.52	0.0708	10.53	0.6611	18.50	0.4796	4.79
D_ANSCL	-0.0430	-1.29	-0.0430	-1.29	0.2237	5.24	-0.563	-1.01
frequency of use	0.0248	0.89	0.0248	0.89	0.071	2.05	0.2138	4.68
effort1	0.0020	13.83	0.0020	13.85	0.0008	3.69	0.0029	9.07
effort2	-0.00002	-0.51	-0.00001	-0.51	-0.00002	-0.52	-0.0001	-1.67
D_ART	-0.0821	-1.40	-0.0821	-1.40	0.1358	1.81	0.0674	-0.74
D_BIZ	-0.0648	-1.26	-0.0648	-1.27	-0.0444	-0.64	-0.3627	-3.50
D_COM	-0.1271	-2.55	-0.1271	-2.55	-0.0454	-0.69	-0.0889	-1.05
D_FAM	-0.0349	-0.43	-0.0349	-0.43	-0.0005	-0.00	-0.0352	-0.26
D_HEA	-0.1299	-1.97	-0.1299	-1.97	0.0568	0.64	-0.3217	-2.65
D_REF	-1.2262	-2.19	-0.1226	-2.19	0.0154	0.21	-0.1690	-1.71
D_REL	-0.1649	-2.08	-0.1649	-2.08	-0.0723	-0.71	-0.1835	-1.39
D_SCI	-0.1672	-2.46	-0.1672	-2.46	-0.1201	-1.36	-0.1249	-1.07
D_SPO	-0.1226	-1.34	-0.1226	-1.34	0.3319	2.71	-0.2544	-1.67
sample size	6853		6853		6853		1672*)	
fit of regression	adj. R^2	0.046	adj. R^2	0.147	pseudo R^2	0.291		
log likelihood	-		-11199.17		-2756.25		-469.19	

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'

*) Sample size of the truncated regression reduced to achieve convergence.

The regressions with the tip ratio as the left-hand side variable instead of the price on the right-hand side deliver largely similar results. The word count-based effort is very significant in all the estimations as well as the rating. Whereas the frequency of use and the answer clarification dummy are not significant in the OLS and the Tobit regression, they are significant in the more sophisticated two-equation model. Here we see that an answer clarification affects the decision whether to tip (the Probit model in column III), but it has no impact on the size of the tip (the truncated regression in column IV). Moreover, no category dummy is significant for the decision whether to tip. However, the category dummy 'Business & Money' is significant at the 5%-level. Posting a question in this category affects the decision how much to tip negatively. While the fit of OLS and the Tobit regression are quite low, the more appropriate specification of a two-step approach has a good fit.

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 effect of social preferences and reputation on the tip that our model predicts. 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 is sufficient.

7 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 add to the research in behavioural economics. In particular we wanted to analyse field data to avoid possible framing effects that cloud lab experiments' results in favour of social preferences. However, we can only confirm earlier research findings from laboratory experiments: Human behaviour is partly driven by social concerns. We believe our analysis strengthens the position of behavioural economics and it becomes more a question of why and not whether people contribute voluntarily.

We built a detailed model of the pricing and tipping behaviour of users and researchers. It is based on the concept of sequential reciprocity by Dufwenberg and Kirchsteiger (2004) that we adapted for the context of this paper. Tipping takes place even among single users, if they are sufficiently socially minded. Frequent users have an incentive to tip in order to create a reputation of rewarding high effort. Once researchers realise that by updating their beliefs about the user's type they will put in high effort. The uncertainty about whether a user will tip is reduced when researchers can update their beliefs. They can make an educated effort decision and high effort is matched with rewarding users, low effort is matched to users who do not tip. The endogenous incomplete contract design can be seen as a virtuous circle that matches effort and increases efficiency.

The predictions of the model are confirmed in the analysis of our data set. We were able to separate reputation effects within the data and found evidence for reputation concerns among the tipping behaviour of users. Moreover, we also found strong evidence for social preferences once we controlled for reputation and focused on single users. Our estimations show that the tip can be explained by social preferences variables (effort, rating and answer clarification) and reputation variables (frequency of use, effort). The data from Google Answers also confirms the positive effect of an endogenous incomplete contract scheme on the effort level. Researchers inform themselves about users' past tipping behaviour and adjust their effort level accordingly. Users known for tipping in the past receive higher effort answers.

References

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

[2] AZAR, O. "The history of tipping - from sixteenth-century England to United States in the 1910s", *mimeo*, Northwestern University

[3] BAJARI, P. and A. HORTACSU (2004). "Economic Insight from Internet Auctions", forthcoming in *Journal of Economic Literature*

[4] CABRAL, L. and A. HORTACSU (2004). "The Dynamics of Seller Reputation: Theory and Evidence from eBay", *mimeo*, NYU.

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

[6] CAMERER, C. and G. LOEWENSTEIN (2004): "Behavioral Economics: Past, Present and Future", in: "Advances in Behavioral Economics", edited by Camerer, C., Loewenstein, G. and M. Rabin, Princeton University Press, Princeton

[7] CHARNESS, G., and M. RABIN (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117, 817-869.

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

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

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

[11] FEHR, E., A. KLEIN, and K. M. SCHMIDT (2001): "Fairness, Incentives and Contractual Incompleteness," *University of Munich working paper*

[12] FEHR, E., and K. M. SCHMIDT (2001): "Theories of Fairness and Reciprocity – Evidence and Economic Applications", *CEPR Discussion Paper*: 2703 February 2001.

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

[14] LIST, J. A. (2004): "Young, Selfish and Male: Field Evidence of Social Preferences," *Economic Journal*, 114, 121-149.

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

[16] SEINEN, I. and A. SCHRAM (1999): "Social status and group norms: Indirect reciprocity in a mutual aid experiment", *CREED working paper*, University of Amsterdam