Information Exchange and Behavior

Copenhagen
Business School
HANDELSHØJSKOLEN

Information Exchange and Behavior

A Multi-method Inquiry on Online Communities

Nikolaos Theodoros Korfiatis

PhD Series 13.2009

LIMAC PhD School
Programme in Informatics

PhD Series 13.2009

Information Exchange and Behavior: A Multi-method Inquiry on Online Communities

Dissertation submitted in partial fulfillments
for the degree of Doctor of Philosophy

by

Nikolaos Theodoros Korfiatis

LIMAC PhD School
Programme in Informatics
Copenhagen Business School, Department of Informatics

Contents

List of Figures

List of Tables

Abstract Executive Summary

This dissertation studies the behavioral characteristics of participants engaged in information exchange in the context of online communities. Online communities are defined as collectives of individuals that use computer mediated communication to facilitate interaction over a shared purpose and/or objective. It is argued that this interaction creates externalities, for example, in the form of codified information that others can use through web search tools. These externalities assemble a virtual form of social capital, a commonly shared resource. The research objective of this thesis is to examine how the behavioral tendencies of the participants in online communities are affected by the way this common resource is formatted, administered and shared.

The dissertation consists of two parts: a theoretical part where the empirical background and the object of research inquiry is highlighted, and an empirical part which consists of four empirical studies carried out in the context of three online communities, namely, Google Answers, Yahoo!Answers and Amazon Online Reviews. The empirical part of this dissertation starts with a controlled experiment emulating a well known social dilemma: the public goods game. It provides substance as to whether and when participants in online communities behave (un) cooperatively. The next two studies focus on a special case of online communities where participants ask questions and other participants post answers conditionally on social and monetary incentives. The results of these two studies confirm that community participants do care about the contributions of others and engage in incentive compatible behavior. Yahoo!Answers participants exercise effort in the community by posting answers to questions conditionally on benefits provided by other participants. The empirical findings show that contributing participants in an online community receive answers faster, while those that do not contribute much effort are sanctioned in the form of longer response-time to their questions.

In Google Answers this thesis, interactions can be observed that are based on monetary rewards (rather than social rewards in the form of a reputation index as in

Yahoo Answers). Participants make use of voluntarily awarded payoffs (tips) along with stated rewards, in order to motivate those that provide answers (answerers) to provide better quality in their responses. The findings of this study confirm the symmetric effect between monetary rewards and quality. However, this study also identifies cases where social norms have a significant effect on response behavior. When participants seek to get better service with less effort (in terms of total cost), a reputation index which is constructed by the history of their previous interactions supports such an attempt. In other words, reputation history influences information sharing behavior in online communities.

The last chapter of the empirical part focuses on another crucial aspect of information as a shared resource: Clarity and understandability. The study examines online product reviews on Amazon.com. The results suggest that participants do care about the clarity of this codified form of experience which increases a helpfulness index accordingly.

The thesis overall finds symmetric effects between participation in online communities and output of interaction, but also identifies the ability of the participants to interact strategically as they seek to minimize the effort they provide in order to find the information they seek. The results underline the importance of signaling and quality evaluation mechanisms as counter-balancing control that can enhance activity on online communities.

JEL Classification Codes: M3, L13, L14, L86, D40, D43

Dansk Resume

Denne afhandling undersøger informationsudveksling og deltageres adfærdskarakter-istika i forbindelse med online fællesskaber. På baggrund af forskningsundersøgelser karakteriseres online fællesskaber som grupper af individer, der bruger computerme-dieret kommunikation til at lette interaktionen i forbindelse med fælles formål og / eller mål. Det hævdes at dette samspil skaber eksternaliteter, som i dette tilfælde er kodificerede oplysninger der kan anvendes af andre deltagere ved at udnytte søge-funktioner på nettet. Disse eksternaliteter etablerer en virtuel form for social kapital. Ved teoretisk at bestemme social kapital som en delt ressource, er forskningsmålsæt-ningen med denne afhandling at adressere om forholdet mellem deltagernes adfærd er påvirket af måden hvorpå denne fælles ressource er formateret, administreret og delt.

Afhandlingen består af to dele, en teoretisk del, hvor den empiriske baggrund og genstand for forskningsundersøgelsen er fremhævet, og en empirisk del, der består af fire empiriske undersøgelser foretaget i forbindelse med tre online fællesskaber nemlig Google Answers, Yahoo! Answers og Amazon Online Reviews. For at sk-abe en generel forståelsesramme begynder den empiriske del af denne afhandling med et kontrolleret forsøg på at efterligne et velkendt socialt dilemma, The Public Goods game. Denne undersøgelse bidrager med indsigt i, om deltagere i online fæl-lesskaber ønsker at samarbejde eller ej. De næste to undersøgelser fokuserer på et særligt tilfælde af online fællesskaber, hvor deltagere stiller spørgsmål og andre deltagere svarer. Resultaterne af disse to undersøgelser bekræfter, at deltagerne i disse fællesskaber er interesserede i hinandens bidrag og udformer deres adfærd i ov-erensstemmelse hermed. På Yahoo! Answers gør deltagerne en indsats for fællessk-abet ved at svare på spørgsmål, men får samtidig gavn af den indsats, der leveres af andre deltagere. De empiriske resultater viser, at deltagere som yder en større ind-sats for onlinebrugere, ved at bidrage med svar på de øvrige deltagere spørgsmål, får svar hurtigere, mens dem der ikke yder en stor indsats i samfundet bliver sanktioneret

således, at der går længere tid, før de modtager et svar på deres spørgsmål.

På Google Answers, hvor interaktion er baseret på monetære belønninger (snarere end sociale belønninger i form af omdømmeindeks som i Yahoo Answers) gør deltagerne brug af frivilligt tildelte udbetalinger (tips) sammen med explicitte belønninger, med henblik på at motivere dem, der kan give svaret (svarerne) til at levere bedre kvalitet i deres svar. Resultaterne af denne undersøgelse bekræfter den symmetriske virkning mellem monetære belønninger og kvalitet, men identificerer også andre tilfælde, hvor sociale normer kan have en betydelig virkning. Et særligt tilfælde er, når deltagerne søger at opnå bedre service med mindre indsats (målt i samlede omkostninger), ved at opbygge et omdømmeindeks, der bygger på deres tidligere interaktioner. Svarerne interesserer sig for vurderingen af deres svar, omdømme og historie og tilpasser deres adfærd i overensstemmelse hermed.

Det sidste kapitel i den empiriske del fokuserer på en anden egenskab ved oplysninger som en fælles ressource: klarhed og forståelighed. Baggrunden for undersøgelsen, der benyttes i dette tilfælde, er online anmeldelser på Amazon.com. Resultaterne antyder at deltagerne er interesserede i klarheden af denne kodificerede form for erfaringer og belønner (med et hjælpsomhedsindeks) i overensstemmelse hermed.

Afhandlingen konkluderer overordnet, at der er symmetrisk effekt mellem deltagelse i online communities og output for interaktion, men peger også på deltagernes evne til at interagere strategisk, idet de søger at minimere den indsats de yder for at finde de oplysninger de søger. Resultaterne understreger vigtigheden af signal- og kvalitetsevalueringsmekanismer som modvægtskontrol der kan øge aktiviteten i online praksisfællesskaber.

Acknowledgements

This dissertation marks the "ithaca" to which I endeavored when I started working as an undergraduate research assistant for the Electronic Trading Research Unit (ELTRUN) at the Department of Management Science and Technology at Athens University of Economics and Business. I would like to thank Professor George Doukidis for taking the risk of offering to a second year undergraduate student at that time, the opportunity to become an active member of a fruitful research environment and participate in the research activities of the group. Angeliki Poulymenakou had been a great support during my introduction to the Information Systems research methodology working with the Organizational Information Systems group. Diomidis Spinellis had been a great mentor and teacher always pointing me to seek the practical and the efficient. My deepest gratitude goes to Miltiades Lytras for encouraging me to pursue research as well as for listening and supporting me in any personal and education related matter. All the time I spent as a student in Athens was a nice memory because of him.

My research working experience with Ambjörn Naeve and his research group in Royal Institute of Technology (KTH) in Stockholm as well as Mathias Palmer of Uppsala Learning Lab in Uppsala University had been a breakthrough towards my decision to pursue a PhD degree. I would personally like to thank also all the members of the Information Engineering Research Unit in the University of Alcalá in Spain for being supportive during my research visits. I would like to specially thank my long term colleague and collaborator Miguel-Angel Sicilia as well as Daniel Rodríguez-García, Salvador Sánchez-Alonso and Elena García-Barriocanal for making my research and course visits in Spain a memorable experience.

I would like to thank all the people of the Informatics Department at Copenhagen Business School: Jan Damsgaard for giving me the opportunity to be part of the PhD programme. Ioanna Constantiou for guiding my first steps in Denmark and giving her support in all the matters that might arise for a foreigner in a new country. The working experience in the department of informatics has been most enjoyable. I still remem-

to my parents Theodoros and Athena for their love and support and

to Katerina for always waiting me

Part I

Theoretical background

Introduction

1.1   Introduction to this dissertation

One of the most influential economists of modern times and a Nobel Prize laureate, Hayek, suggested that a key problem of society is the coordination of dispersed knowledge — a problem a central planner would be unable to address [Hayek, 1945]. Many things have changed since then, including the fundamental ways information is transferred between individuals in a digital context. Nonetheless Hayek's key perspective on that problem of society still remains central in a digital age, where the World Wide Web does increasingly facilitate collaboration between individuals. Wikipedia, the online encyclopedia, is a case in point: On many occasions it contains much more content than a centralized system can handle [1]. If collaboration cannot rely on rules administered by a central planner, how then do people participating in online collaboration self-organize not only their interaction but also the incentives inducing online behavior?

---

[1]Although for some, the accuracy of Wikipedia still remains a controversial issue

The thesis aims to contribute to the understanding of the motivational factors that affect participation in virtual/online communities, collectives of social interactions on the web, from a participant behavior point of view. For clarity reasons, we usually refer to the participants of online communities as users, since online communities are based on software that is accessible through the World Wide Web. Most of the current academic research on online communities (outlined in the next chapter) approaches the subject from rather a posterior perspective, treating it ex ante as a living body (e.g., a mailing list) rather than an ongoing formation with profound behavioral characteristics [Barak, 2008]. Inevitably, exchanging information plays a key role behind the motivation for participation of an individual user, especially because in online communities it is the primary medium of exchange and codified output. However, a key question that is tackled in this dissertation is what constitutes the nature and the driving forces that are behind this desire for information. Do users consider their desire for information as the key reason for participating in online communities? Is their participation affected by other characteristics that have to do with behavioral properties which are attached to this desire for information.

In order to achieve this object of research inquiry, this dissertation encompasses a theoretical development outlined in this and the subsequent chapter (Part I) where the research context of user interactions in online communities is discussed together with the related theories. The empirical part of this dissertation (Part II) encompasses four studies that are grounded on user's interactions captured in four datasets used in this dissertation. The overall goal of the empirical part is to demonstrate the issues discussed in this and subsequent chapters, as well explain the connection with theory presented on the following chapter using a mixed research design approach.

The sequence of the studies presented on the empirical part follows a top down approach. While the third chapter studies the impact of the behavioral characteristics on a controlled environment (dictated by a quasi experimental design) the next three chapters provide an analysis on an online context. Chapters 4 and 5 provide an

analysis where cooperation is an important element of the study context (community managed question answering systems), while the Chapter 6 provides an analysis of an environment where self interest is evaluated by an online community mechanism.

In particular, the studies presented in this dissertation are as follows:

- The third chapter provides a study on cooperation in relation to online communities. Cooperation is an important factor for the sustainability of online communities since it affects the outcome of the interaction between the users. In this chapter the relation between cooperation and online social interaction characteristics is made clear using a public goods game. We first explain the public goods game and its game theoretical assumptions and then describe the experimental procedure. We use two distinct framings in relation to the presence of a subject in an online setting, where: (a) the game contributes to the common good (b) it receives benefits from it. The framing is distributed into two distinct treatments with an extra treatment acting as a control of offline participation. The subjects are then presented with a sequential version of the Public Goods Game where each decision (with the exemption of the unconditional choice) is given at once. Results indicate that participants in online communities indeed also show a high degree of cooperation both on the contributing and the benefiting framing, conditional on the contribution provided by the others.

- The fourth chapter examines the effect of activity on service posture (measured by volume and time) as expressed by user contributed effort and user received benefit in an online community facilitated by users of the Yahoo!Answers service. Yahoo!Answers operates a question-answering community of users who post questions and receive answers on various topics. We describe how the ration of contributed effort and received benefit has an effect on service posture (volume and time). By programming a web crawler to store a random sample of questions posted over a period of one month, we use a set of time series, random effects and logistic regression models which confirm, to a large extent, our

formed hypotheses. In particular, we find that users who contribute more effort in the community than received benefit get a question answered by more users in less time than users who receive more benefit than the effort they contribute to the community.

The fifth chapter tests whether a particular type of voluntarily awarded monetary rewards (tips) are paid for strategic reasons in a quasi-experimental setting. The context of study is Google Answers. Google Answers was a marketplace of information inquiries in which any *asker* can post a question along with a price to be paid for a satisfactory answer. One researcher from a closed group of *answerers* answers the question, usually by providing reference to authoritative sources. Upon receiving an answer, the asker rates the quality of the answer obtained; if satisfied with the quality, the asker pays the price and additionally pays a voluntary tip. We investigate tipping behavior before (when strategic considerations can play a role) and after (when they cannot) the announcement of the shutdown of the answering service. To disentangle a motive of the strategic nature of tipping from other (reciprocal or norm-driven) motives of tipping, we analyze pre-announcement tipping behavior. The empirical results suggest that askers use tips to induce better (e.g., in terms of better promptness) service in the future, and that answerers respond to tipping by providing services more promptly to those with a better history. We particularly show that a class of users relies on repeated interaction in order to receive better service with less cost.

- The sixth chapter investigates how users perceive interactions that affect their decision to buy and, in particular, whether their evaluations are related to communication issues as in the case of how readable the submitted reviews are in relation to the usefulness ratio that is attached each review. The unit of analysis in this chapter is online product reviews. Online product reviews are an important resource for consumers of experience goods in online marketplaces because they provide a useful source of support information during the purchase of a

good. Furthermore, in some online marketplaces consumers have the opportunity to evaluate how helpful a review is by using a binary evaluation interface provided by the online marketplace. This results in a usefulness score of a review which is calculated as a fraction of helpful votes over the total votes that this review has received. The results indicate that the usefulness score of a particular review is affected in a significant way by the qualitative characteristics of the review as measured by readability tests applied to a large dataset of reviews collected from the U.K section of the popular online marketplace Amazon.

Having outlined the objectives of the studies presented in the empirical part of this dissertation, we continue to provide the theoretical background, as well as the unit of analysis addressed in this thesis.

1.1.1 Web enabled production and use of information

As aforementioned, the specific empirical context of the research encompassed in this dissertation is Online Communities. Preece [2000] defines online communities as a collective of individuals that interact socially with other individuals by using computer mediated communication and adhering to a set of policies imposed by tacit assumptions and protocols that guide their interaction.

Online communities are a particularly interesting context to study online social interaction, where interaction is theorized as a subset of computer mediated behavior. This is the unit of analysis of this research. In particular, the thesis aims not only to study behavior in online communities under different perspectives imposed both by context and content, but also to theorize the underlying patterns of behavior that are evident in these contexts. To this end, the research presented in this thesis makes the following assumptions.

- Exchange of information takes place in codified form and is the result of social interaction. By implication online communities leave traces of recorded information

behavior. This online context facilitates observation of information behavior by providing automatic classification of the recorded sessions (for example a thread in an online forum).

- Social interactions in online communities are rule based. This thesis is particularly interested in the interaction rules and how they affect contributions in online communities. Interaction rules can be either explicit (e.g., interaction policy rules as posed by the online community administrators) or tacit (in the sense of moral codes) which leads us to the third assumption.

- Special modes of social interaction (e.g., the adherence to standards of social behavior) have to be examined to see whether they are applicable in online settings where anonymity prevails. This is particularly important in the context of which social interaction takes place and the perceived value that the individuals consider this interaction to have (e.g., a review provided that concerns a specific product).

To this end, an online community should not be considered exclusively only as a space where interaction between individuals is framed on the exchange of information, but the way behavioral characteristics affect the outcome of interactions between individuals in a specific context (where the nature of interaction becomes more or less important).

Therefore, the research focus of this thesis takes this case one step further, assuming that the shared purpose is encapsulated through the exchange of information. We consider this exchange of information to have an outcome or externality produced in codified form due to the use of computer mediated communication tools. Therefore, this thesis considers an online community to be an environment where participants can form bi-directional interactions over a shared purpose related to information (e.g., the exchange of knowledge). While other online social environments might fall into that category (e.g., information portals), the primary requirement that we focus on in

this dissertation is that these environments allow their users to interact and to form social interactions that also have specific constraints, such as the requirement for an identity or an alias to be supplied (usually through a registration system), in order for other participants to be able to observe the actions and participation history of the members of an online community.

From a research context point of view, this dissertation is centered on the study of social interactions on the World Wide Web (WWW) using it as a focal point to analyze the way people seek and contribute information. This is used as an approach of measuring a particular aspect of behavioral characteristics related to motivation to participate and expressed by information exchange.

In other words, this dissertation does not study the nature of information per se; rather, it focuses on the social context under which production of information takes place, such as the one evident in the context of an online community. With the development of new ways of collaboration over the web, an ongoing development of the scientific literature has addressed the essence of harvesting the potential that technology provides for institutions and individuals to interact over information artifacts. This is essentially important for organizations, for example, where by harvesting this rich social environment on the web, it is possible to either innovate [von Hippel, 2007] or diversify the existing customer and/or user base [Godes et al., 2005].

1.1.2 Research background and motivation

Social capital in online communities

One particular aspect of the research issue described above, resides in how to harvest information from all these individuals and make them cooperate and collaborate on the exchange of information. In other words, the question can be framed on how to enhance cooperation towards a specific goal. Interestingly enough, this research question has been framed in other parts of the research literature, particularly in the research stream that has to do with the intellectual capabilities of the firm as an

information processing unit [Barney, 1991]. The research is published in Journals such as the Academy of Management Review, as well as in exclusively related journals with organizational and economics related issues of information systems, i.e., Information Systems Research and Management Information Systems Quarterly [McKnight et al., 2003, Jr et al., 2008, Majchrzak et al., 2005, Wasko and Faraj, 2005, Pinsonneault and Kraemer, 1993] Research appearing in these publication venues summarizes these issues as problems related to the impact of social capital in various areas of online activity ranging from trust building in organizations to knowledge exchange on the internet.

Social capital relates with the research endeavor of this thesis for the following reasons:

- First it has as its focus the relationships between the individuals in a social structure [Burt, 2005]. Regardless of the implicit or explicit nature that these relations might have, social capital provides a framework for the incorporation of these relations into the understanding of the dynamics of a social structure such as in our case an online community.

- Second, it considers individuals' attributes as the primary factor that makes their relations sustainable [Erickson, 2001] and

- Third it emphasizes the existence of social resources (e.g. in our case information and knowledge) as the sole factor that creates hierarchies over the structure of these relations.

With those three elements in focus, let us revisit the definition of the online community that we defined in the very beginning of this thesis. In particular we consider an online community formed around individuals (I) who share a common purpose (III) and adhere to a set of protocols (II and III), with their adherence being subject to their arguments (II). To this end we consider the concept of social capital as the ideal framework for studying online communities since it provides a conceptual pathway to

embed all the above three elements in one perspective.

Although the term "social capital" has appealed to many scholars in the research literature, there is a strong debate related to the formalization of social capital as a resource that can be measured and exploited. Table 1.1 summarizes the most acknowledged definitions of social capital. Bourdieu in his work *"Forms of Capital"* [Bourdieu, 1986] recognizes social capital as the third element of a triple consisting of economic, cultural and social capital which runs as an enabler for the formation of social interactions. These interactions are centered on the exchange of resources and construction of the necessary social cooperation for the purpose of the creation of these resources. Coleman [1988] extends the sociability characteristic of social capital to capture anything that can enable social interaction to happen, which is generated by collective action, reciprocity or vast networks of relationships. Another approach by Putnam [1995] in the 90's through the initiative of a World Bank research program, addressed the concept of social capital as a multifaceted artifact that encapsulates the ability of a social structure to generate value through collaboration, requiring enablers such as trust to be present. Putnam's approach emphasizes the importance of trust, social norms and social networks for improving the efficiency of the social structure that possesses them (e.g. a firm) and on a macro level the society itself.

Although Putnam has extended the concept of social capital to a macro level, Bourdieu's definition is still intriguing from the perspective that it theorizes social capital as "*an attribute of an individual in a social context*"; This attribute can be acquired depending on the ability to employ the nature of the social connections around the individual. This definition, along with the definition provided by Coleman, follows a structural perspective on social capital and its nature. The later adds that social capital cannot be evaluated without the presence of mechanisms that enable social interactions to take place. For example, one cannot assess how cooperative an individual is without providing an environment where cooperation can be formulated. Such cooperative structures constitute forms of social capital where the social connections

play the roles of the transformers of the relationships to something that provides a value for the generation of other forms of capital.

Although social capital has made an enormous impact on the management and organizational theory literature [Adler and Kwon, 2002], there is criticism from other fields, especially economics. An important critique from the viewpoint of economics relates with the imprecision of the term from a resource based perspective. In particular, Arrow [2000] addressed the issue that social capital on the one hand is not capital per se in the form that economists asses it due to the imprecision of its definition and on the other hand even if we accept that social capital is indeed the driving force for the performance that some social structures have, there is no clear evidence to that.

Table 1.2 summarizes some of the major criticisms towards social capital. As mentioned earlier, the arguments against social capital come mainly from the economic literature and particularly in the way that social capital literature imprecisely asserts social capital as a new form of resource. Arrow's argument against the concept of social capital is that it does not resemble capital in a standard tangible form so that it can be transferred from entity to entity. But even if it can be considered as an asset for an organization, it is difficult to find evidence of how important it is. Solow's argument [Solow, 1995], for example, proposes that when social capital is of an individual nature that is also affected by culture, there is no evidence of contribution of social capital to economic activity (e.g., in the form of trust) across different nations where culture could play a role. Coleman's definition of social capital highlights the importance of dense relations in a social structure as a mechanism that enforces cooperation moving the discussion of social capital as a resource to social capital as a mechanism for the generation of these resources.

Social capital is also connected with the behavioral characteristics of the individuals. Foley and Edwards [1999], for example, criticize the context dependent nature of social capital, which is attached to several different aspects of social activity and which therefore cannot be theorized as a distinct concept. In social psychology we

can observe similar situations due to the fact that behavior in a societal structure is very much dependent on the context.

Behavioral characteristics play a key role in defining what social capital is and its relation to social interactions. The fundamental axiom is that humans as social beings have the ability to manipulate their behavior conditional to the environment that they are in. This manipulation can happen either consciously or unconsciously, depending on the presence or absence of several factors which are axiomatically accepted to cause a change on the behavioral patterns of an individual. A very early study by Allport [1935] provided the first insight on the reasons why the change of a human subjects' behavior is attributed to the presence of the others. In this and subsequent studies, behavior in terms of sociability is dealt with as a resultant of three basic elements: incentives (factors that push behavior to a certain direction), structure (the way behavior is affected by the presence of the others) and the setting that in which this behavior is observed (off-line or physical environment, online or virtual environment).

According to Deci and Ryan [1985], incentives can be on the one hand extrinsic or exogenous in the form that the subject[2] receives a measurable compensation for his/her effort. On the other hand, incentives can be intrinsic, as subjects might also be motivated by intrinsic or endogenous means of motivation where the compensation is not measured with standard utility yardsticks.

Coming back to the research context of this dissertation, as mentioned earlier, the research that has been undertaken for the development of the web has been extensively on the issues of technical realization and evolution of technical standards. Although the technical issues regarding the retrieval of information from web sources have been well addressed and well challenged by the information retrieval community by developing computerized methods for better information reference and retrieval [Brin and Page, 1998] there is an undermining of the social potential that the world

---

[2]With the term subject we characterize those social agents that participate in a type of interaction (social, economic etc)

Definition	Source
Social Capital as the aggregate of the actual or potential resources which are linked to possession of a durable network of institutionalized relationships of mutual acquaintance and recognition	Bourdieu(Bourdieu, 1986).
Social Capital consists of a variety of entities with two elements in common: they all consist of some aspect of social structure and they facilitate certain actions of actors within the structure	Coleman (Coleman, 1988).
Social capital refers to the collective value of social structures /networks and the tendencies that arise from these networks in two perspectives: bonding (between homogeneous groups) and bridging (between heterogeneous groups)	Putnam / World Bank [Putnam, 1995].
Social Capital as a combination of structural, relational and cognitive abilities in an organizational structure.	Nahapiet and Ghoshal [1998]

Table 1.1: Definitions of Social Capital

Argument	Source
Social capital does not resemble a standard form of capital in the way it can be transferred from one owner to the other.	Arrow [2000]
There is no evidence that social capital contributes to economic activity, especially if you compare studies across different societal structures.	Solow [1995]
Social capital is based on premature concepts encompassing several different aspects of social activity and therefore cannot be perceived as a distinct entity.	Mondak [1998]
Social Capital in terms of social structures, norms, trust and reciprocity cannot be theorized due to the context dependent nature of its value.	Foley and Edwards [1999]

Table 1.2: Arguments against social capital

wide web offers for the production of information. This later trend known as Web 2.0 [O'Reilly, 2005] focuses on scenarios where individuals can use the web to find information, and then in cooperation with other individuals can structure and define this information better using the benefits of collective action. Concerning the later, web sites of collective action, such as Wikipedia[3], provide an example of how the web can harness the collective wisdom of individuals and transform it to a dynamic artifact where the quality of resources becomes better and better.

Virtual forms of social capital and online communities

Thus far, the discussion provided in the previous section concerning social capital considers it to be a form of capital that takes place in offline settings where social interactions are formed in a physical form (either by affiliation e.g., participation in a club or a community group or spontaneous by context dependent settings such as the workplace). But how can social capital be addressed in a virtual form? Is there an infrastructure that permits the creation of social capital in a virtual setting? Can forms of social capital be found and studied on the Internet?

In the perspective of this thesis, this is an important research question first from the conceptualization of social capital itself. This is because, as mentioned earlier, social capital addresses the importance of individual characteristics and their individual attributes over the access of shared social resources. However, an important issue is that the research literature that we discussed earlier approaches the *offline* definition of social capital. The Internet, however, is an environment that has well grounded social mechanisms; for example, sanctioning (in cases of antisocial behavior) is difficult to be imposed, thus making the adherence to social norms a difficult-to-moderate issue. None forbids an individual who participates in an online community to create a new alias and to behave in a similar manner as before (in case he gets sanctioned for antisocial behavior). However, the question still remains as to what the reaction

---

[3]http://www.wikipedia.org

of the other participants will be, that is, will they react the same? This is more or less related to whether the focus of analysis is the individual or the group. In that sense, it is important to study forms of social capital on the Internet to see whether there are any similarities with the theoretical arguments that have to do with offline settings.

In this dissertation we approach social capital primarily from a collective action perspective. A particular case of this research issue is the case of collective action around information artifacts in the principles of the original model of the web [Bimber et al., 2005]. Groups of individuals are provided with a platform where information exchange between them can be facilitated [Turner et al., 2005]. This in fact can be seen as a mode of collective action. As Olson [1971] states in his fundamental work around collective action *"groups of individuals with common interests are expected to act on behalf of their common interests" (Olson, The logic of Collective Action, pp: 149).*

Following Olson's original definition, collective action is dependent on collective behavior. As aforementioned, collective behavior is a type of behavior that can be defined as coordinated action among a specified population. One characteristic of collective action as discussed in the literature is that it occurs as a result of temporal collective behavior under a specific context [Gurven and Winking, 2008]. For example the coordination of crowds in sporting events is an aggregate of the collective behavior of individuals that exists only during the context of the sporting event that they attend [Bartel and Saavedra, 2000].

Theoretical research around the characteristics of collective behavior can be generally classified into two theoretical perspectives. The first perspective occupies the view that collective behavior is a result of the social environment and its settings (e.g., already defined hierarchies and social structures); the second one advocates that behavior is a result of a context specific social action that acts for a specific outcome. According to this perspective offline cases of collective behavior such as rumor spreading can be explained due to the fact that the action has a specific outcome.

Turner and Killian [1987] in their study of collective behavior among human subjects provided a classification of three types of collective behavior, namely *the crowd*, *the public* and the *social movement.* According to this study, this classification is very much based on the social setting in which the collective behavior takes place. This is due to the socio psychological perspective that collective behavior is not a pathological phenomenon, but is very much dependent on social change (e.g., the environment and the social norms that characterize it). Their model acknowledges several social properties that characterize collective behavior: (a) the existence of emergent norms, (b) feasibility of the action (c) timelines – the time setting in which the action occurs and (d) the preexisting groups and networks.

Another well known research paradigm that is often adopted in studies of collective behavior is the one developed by Smelser [Smelser, 1962]. In this model Smelser summarizes a set of conditions that need to be present in order for collective action to occur. These conditions are classified in (a) Breakdown of social control, (b) Structural Conclusiveness and (c) precipitating incidents or triggers that occur before the emergence of collective behavior. In particular precipitating incidents are vital for an online community due to the asynchronous mode of communication among the participants [Ravid et al., 2004].

As will be discussed later, in principle, collective action as a result of collective behavior, occurs only under certain conditions, namely, *Scope* and *Interests*. From the perspective of the research question tackled in this thesis, collective action provides that individuals form groups which have as an objective function the addressing of the compilation of information sources either by doing it explicitly with a certain objective (e.g., a Wikipedia article or the development of a new software) or implicitly (by deliberately posting information in an Online Forum or in an Internet newsgroup). Olson [1971] takes this approach one step further, arguing that effective collective action (with no individuals taking advantage of the effort of other individuals, and thus becoming free riders) leads to a production of commodity that is known in the literature

17

as public good [Samuelson, 2000]

Following this perspective, our perceived nature of information as a **public** good becomes that of a codified artifact, which under a context provides value for those that use it, though retaining its public nature and not losing its value. Nonetheless the view of *information* as an artifact is still limited around its consumption. Individuals tend to use it, consume it or produce it when their knowledge is limited on the domain that in which they are active. However an issue remains on *how individuals communicate and how they contribute information?*

In that way the incremental adoption of the web as a communication channel has resulted in a broad variety of online communities which can be conceptualized as groups of individuals with a dense number of social interactions over the Internet. The later embellish a significant role into several application domains (e.g., opinion forums, online auctions, etc.) with potential applications in other areas, such as enhancing trust for electronic transactions. Particularly in this diversity of communities, there are cases where online social interactions are not only a way of communication, but act as an enabler of transactions (e.g., in the case of online auctions) where no contractual enforcement is present [Dellarocas, 2003].

Conversely, unless there is a formal protocol which defines how communication is facilitated, a significant problem of these online communities is the issue of participation, both in terms of membership and activity. Membership deals with the handling of participants in an online community and levels of functionality that the members can employ. For example, in communities where the content discussed is *moderated*, the structure of the members is not flat, but it employs a certain hierarchical structure. Furthermore, the membership has to be retained at all the stages of the community activity in order for it to become sustainable.

Membership in online communities is also a limiting factor in cases where an online community might require participants to register their identity usually through a login system. Some online community systems, such as those provided on the discussion in

online blogs, allow members to participate without registering but require them to use the same identity during their participation. Therefore, in this dissertation the notion of membership is not tied specifically to the registration policy but to the membership monitoring which is very much related to the identity management issue. Deciding whether to have a registration policy in an online community is also important for the attraction of members in the initial stages of development of the community since this might have a negative effect on the nurturing of the community [Preece, 2000]

In particular, online virtual environments require a certain number of members or a critical mass in order to have some activity and thus retain their members. Activity acts as an incentive mechanism to the existing community members to participate and to outsiders to join the community. Nonetheless, although there are profound flexibilities to form interaction (e.g., related to time or space distance), this type of virtual communication is quite difficult to be formed in a non ad-hoc way.

As Finholt and Sproull [1990] indicate, technical solutions that act as enablers of communication over the internet, address only the infrastructural solution to this problem. Preece [2000] adopts another perspective to this issue by addressing online community participation by using two pillars: the usability and sociability of the community mechanism. It's commonly accepted that technology solutions per se cannot guarantee participation of individuals in order to assert an on-line (virtual) social activity. A particular need for understanding the social mechanisms that highlight the participation and social interactions on online communities arises as the potential of these communities has been addressed in the literature both from a theoretical [Quan-Haase and Wellman, 2004, Wellman et al., 1996] as well from a practical viewpoint[Godes et al., 2005].

The impact of incentives for this type of collective behavior is an important issue in online community research. In the literature there are several studies that try to outline what the incentives are for participation and thus explain the behavior of individuals that participate on these online social groups [Jones et al., 2004, Kollock,

19

1999].

However as stated in the study by [Wang and Fesenmaier, 2003] there is little empirical evidence regarding the nature of the incentives that affect behavior in online communities and their contribution to the contagion. The study by [Ling et al., 2005] for example examined contributions from a collective perspective concluding that users will contribute more in an online community if they perceive their contributions as important for the group outcomes. The basic assumption taken in this dissertation is that incentives affect behavior to a way which is expressed with future action. We categorize incentives into two major groups, namely those related with social or psychological factors and those that have to do with economic behavior. The social and behavioral category of incentives deals with the cases where behavior is affected by endogenous social factors related with the social context and the position of the individual in it. Social incentives study the way group interaction patterns are formed by taking a holistic view of the interaction structure and behavior under certain viewpoints (e.g. contributed effort, activity and commitment).

The other category of incentives studied in this dissertation, and particularly in Chapter 5, relates economic incentives in terms of compensation which can be either monetary rewards or elements of value which users consider to be important (e.g., non contractually stated rewards such as tips). In most cases, economic incentives or other extrinsic forms of motivation try to explain behavior by theorizing a rational agent model of the participant.

That is, in the case where an individual's objective function is to seek relevant information, in a way, it maximizes his/her utility by participating in a community. Nonetheless, empirical evidence may contradict this direction. One could argue that since members receive no profound compensation for their participation, they have a high opportunity cost. For example, an expert who participates in an online community (e.g., a forum of computer programmers) and devotes a significant amount of time for answering complex questions might have a high opportunity cost depending

20

on his off-line activities and the compensation that he receives by doing them.

Similar to the later, one of the much cited problems in the case of communication activity is the factor of the membership size [Butler, 2001]. As a club or a union, an online social structure, in order to operate appropriately, needs a critical mass of members. Related to this is the problem of activity. While due to design settings, people are obliged to become members of a virtual community in order to participate, there are several cases where activity that is not obligatory (e.g., in terms of a community facilitating transactions such as e-bay) is not directly affected by membership size.

This phenomenon has been placed in computer mediated communication literature as lurking [Preece et al., 2004, Ravid et al., 2004]. Lurking characterizes the behavior of individuals who while participating formally in the community, are not active. An online community with a high number of lurkers has an activity problem which results in a low quality of social interactions between members. Furthermore, as Cummings et al. [2002] point out, a significant problem is the quality of those social interactions and the nature of the relational ties that are formed through them, with respect to the rest of the participating individuals. All these concepts have contributed to the viewpoint of the emergence of virtual forms of social capital taking place in the realm of online communities where individuals interact with each other by coordinating actions (e.g., online petitions) or by simply contributing information.

While the growing focus that firms give to the cultivation of their social capital potential is evident in the management literature, online communities also provide a significant space for their interaction with customers. Armstrong and Hagel [2000] argue that a significant benefit for the nurturing of online communities by firms can be customer loyalty. This can be attributed to the network effect that might become evident when, for example, a community of customers of a specific product reaches the critical mass. As it is in offline settings where an adoption of a product in a market depends on a critical mass of consumers, so it is in online settings where the network

21

effects in terms of awareness become exponential. However, the latter applies to those firms that consider information to be an important element for preserving customer loyalty, while in offline settings other issues, such as brand management and market segmentation, play a significant role as well [Kotler and Bliemel, 2000].

Thus far, we have reviewed some of the characteristics of social capital that resemble a virtual form of social capital in the context of an online social environment or an online community. But **are there any other characteristics that we should take into account when studying an online community?** Is there a special connection between the collective action that occurs on a virtual setting and the motivational factors that affect it ? We provide a review of some of the characteristics of online communities in the subsequent section. A more extensive review is provided in the second chapter of this dissertation which deals exclusively with the research that has been undertaken in the field of online communities in relation with the motivational factors that affect their sustainability and success.

Contributions to social Capital in Online Communities

Although the term "Online or Virtual" provides the same semantics it still holds some imprecision when it comes to addressing the case of an online community. Lave and Wenger [1991] have described a community of practice as an *activity system* which includes individuals that are united in action and in the meaning that action has for them on the larger collective. The concept of Online or Virtual community has been attributed to the work documented by Rheingold [2000] on the creation of a virtual social environment using the early infrastructure of the internet and in particular USENET[4]. Much of the research done in online communities has been theorized on the context of computer mediated communication (CMC)Walther [1996]. Computer mediated communication considers the case of communication models between individuals that

---

[4]USENET was an electronic mail exchange facility where it provided the means for bulletin boards of electronic messages that appeared publicly forming online discussions as thread of related electronic messages. It is still in use today and its archive is accessible via Google Groups

are facilitated by computer and technological means. According to Baym [1995] frequency of this type of communication can lead to the formation of a community with distinct social characteristics.

Another perspective, for example, the study by Bryant et al. [2005], tackles the social mechanism of an online community as a facilitator of *new ways of organization* when it comes to a specific context and a purpose (in that study the case of the Wikipedia). This study provides for example the case that when individuals communicate by using technology then organizational hierarchies are very difficult to be formed, and thus a more flexible organizational structure can be shaped. This comes to significance when communities are formed to address a specific goal. For instance the Apache HTTP server project started by an online community of web technology professionals that wanted to contribute to a type of software that was regarded by many as a commodity and started sending pieces of codes (patches) to the mailing list [Rigby et al., 2008].

Furthermore, Rheingold's work approaches online communities as an extension of continuous online interaction - interactions that are supported by the use of electronic tools such as email. This perspective encompasses a broader definition of community than the one which are adopted by sociologists. For Rheingold, an online community will exist as on meta-level of interaction where individuals have already established a communication channel. The face-to-face definition of community as it is known in sociology, however, requires that for individuals to join in a community, they have to first establish a communication channel, that is, by passing the *social barrier* of first establishing a contact, and then participating in the community. .

Whittle [1997] argues that one of the main socio-psychological characteristics that make the virtual community a broader space than the offline one, is the notion of *Equalized Status*. The distributed nature of the Web, as can also be observed today, provides everyone who has access to a simple technological appliance (such as a network connected computer with a web browser) an equal opportunity to signal his/her

Factor	Online Communities	Off-line Communities
Type of Interaction	Non-personal	Personal
Output	Codified	Non-Codified
Facilitation of Cooperation	Ad-hoc	Determined
Attribution of Social Status	Linear	Hierarchical

Table 1.3: Online Communities and Offline Communities

presence and express his opinion. In most online communities, such as in Internet newsgroup, everyone regardless of (offline) social status, economic condition, gender, etc., can participate and express personal views. Unlike in offline communities where participants share a common view and hierarchical structure is predominant, the participation in online communities generally follows a linear structure where participants possess an equal level status.

Nonetheless, with the development of software that allows more advanced ways of interaction and communication between participants (e.g., open source online forum software), mechanisms, such as moderation, become easier to be deployed, in which case the acceptance of moderation from the subjects becomes an important issue. Anderson and Kanuka [1997] in an early paper discuss the role of the moderator as essential for creating and supporting a stimulating environment. This can be seen both from a technical perspective (where the moderator has the power to delete or edit posts provided by participants which can be of offensive nature) and from a stimulation perspective where he/she can guide the interaction towards more meaningful outcomes. In such a case the moderators' status clearly becomes an instrument of influence. More and more online communities allow participants to become moderators as a result of their frequent participation. Moderator status results in high power and influence over other participants.

This becomes important in the way, for instance, that a community member might influence others. Social influence is always observed in off line settings where apart

from communication, other social characteristics such as those mentioned above (e.g., appearance) might play a role. In online communities, social influence might be attributed to communication skills, persistence and quality of ideas and the level of technical knowledge that someone has. The latter is important for two reasons; first, for providing the ability to communicate by using technical means (an email software application or the capability to navigate to web pages by using a web browser), and second, the technical knowledge that might be of importance to the context of activity in which the community is taking part.

Rheingold [2000] provides a deeper way on the advantages and disadvantages that an online community has. These advantages can be summarized in three different dimensions: (a) *Break through Constraints, (b) Social Presence, and (c)* the ability to form *several levels of Social Connections.*

The *break through constrains* advantage of an online community is built on the broader conception of an online community as an extension of online communication. Following the basic assumption that online communities are grounded in online communication, the standard communication variables of space and time become absent due to the fact that co-location (same place) and synchronization (same time) become less significant to the formation of a community as it might stand in off-line settings.

The level of *Social Experience* is a meta-stage of the *break through constraints* characteristic of an online community due to the fact that members gain a higher level of exposure to an environment full of different cultural settings. The later contradicts the *homophyllic*characteristics of the off-line communities, the case that participants form connections due to common characteristics such as language and culture.

The case of social connections raises another significant property of online communities. As has been mentioned, online communities are established by extending the individual-to-individual communication to a group level without constraints. Rheingold argues that social connections act as bonds and become stronger the more integrated the online communication is [Rheingold, 2000].

25

Nonetheless, online communities suffer some problems which are a consequence of the non-personal type of interaction that is formed and the codified output that they produce (most of the times an archive of the discussions taking place). According to Mynatt et al. [1997], this is connected to two major issues namely the *Lack of Accountability* and the *Privacy Issues*.

- The *Lack of Accountability* in an online community is a major characteristic that is derived from the use of communication tools that do not reveal personal information due to the use of pseudonyms or aliases for establishing contact and signaling presence. This often makes the users un-accountable for their actions and expressions due to the fact that no off-line world details are revealed and individuals feel free to express views that otherwise they wouldn't express in offline settings. Whittle [1997] refers to this issue as a major disadvantage of an online community which might create problems for the sustainability of the community due to the fact that self control over the views expressed is hindered.

- *Privacy Issues* raise another concern mainly for the way that individual participation in a virtual community might be affected. Due to the fact that online communication always produces a codified output participant's views, it can be reviewed and evaluated in most cases by outsiders. Furthermore, those that deliberately extend their alias to personal information in order to receive a higher social experience might face a privacy issue. The continuous development of information retrieval mechanisms provides information in online communities that can be easily retrieved and evaluated by those who are interested in their activity. Gross and Acquisti [2005] argue that privacy issues pose a threat to the activity of the community and thus may provide a barrier to individual to express freely themselves. A particular example can be the case of the online social network Facebook[5] where privacy issues are an ongoing stream of debate.

---

[5] http://www.facebook.com

de Souza and Preece [2004] provide another dimension on the analysis of online communities, that of information noise or spam. In their analysis, they emphasize the importance of policies of communication between members as an important factor that can enhance the quality of an online social space. With the recent increase of SPAM as a threatening factor to the sustainability of online social spaces [Brown et al., 2008], the importance of communication policies in the form of becomes a factor that should be considered from the perspective of an online community. However, moderation can be difficult in cases where increased activity leads to a high volume of communication. Distributed moderation, such as the one provided by the Slashdot online community, can be a solution where the interaction between members is also fostered and participants who are status seekers might be highly incentivized to participate more and provide more valuable content [Lampe and Resnick, 2004]. This, in turn, results in the diversification of the types of online social interactions that can be formed in an online community (e.g., direct interaction between members or indirect interaction in the case that the participants interact by exchanging opinions in a thread or a discussion).

In this dissertation we generally define online social interaction as a social interaction that takes place in an online setting, such as the WWW. As it has been provided by social psychologists, social interaction is a dynamic set of social actions between individuals in a specific context. For example, in the context of an online community, there might be a sequence of the basic events: post-reply-repost provides a block of social interaction that takes place in an online setting, which is a research issue from a privacy point of view [Acquisti and Gross, 2006, Ellison et al., 2007].

1.1.3 The problem of Motivating Contributions to virtual forms of social capital

Having discussed in the previous section why online communities resemble a virtual form of social capital, we now clarify the particular research question that this disser-

tation tackles. Kollock [1998] summarizes the problem of online cooperation in the context of a social dilemma. In general, social dilemmas represent cases where private interests collide with collective interests; this is because of the social nature of most participants who weigh their (short term) personal interests more heavily than the (long-term) interests of the group. Clearly, contributions in an online community can be seen as a social dilemma due to the fact that they have to be incentivized in order for some activity to occur, as well as pose impositions on the collective nature of the participation. In other areas of the research literature, the classical social dilemma is the case of public (non rival and non excludable) goods, where non participation might lead to the well known case of the tragedy of the commons [Hardin, 1968]. This is true in social dilemmas in situations where individuals who are driven by their own self interest may cause a decline of the collective value of a shared resource (e.g., low activity might lead to the fact that the information shared in an online community becomes less valuable). This example is contradicted by the long term interest of the individual, something which certainly is not to degrade the value of the online community, but to receive a benefit by becoming a member. A similar case with the tragedy of the commons can be seen in the example of lurking, where individuals observe but do not participate, and with this behavior contribute to the low activity of the community. Armstrong and Hagel [2000] argue that this is directly connected to the value that the community provides for its members.

However, as the aforementioned incentives can be of exogenous, endogenous or combined form, their role is difficult to distinguish. On the other hand, how do the participants themselves consider their participation in an online community? What is the importance of these motivational factors?

One might consider these questions to be of some sort of psychological nature. Indeed, the research problem is essentially a study of behavior. However, what makes it an interesting case from an information systems point of view is the context. In psychology and especially the area that is involved with the analysis of behavior in social

Function	Research Objective
Value Expressive	A,C
Utilitarian	A,C
Social Adjustive	B
Knowledge Seeking	B

Table 1.4: Clusters of Functions and research objectives addressed using the framework of Snyder and Cantor [1998]

groups, there is the dominant principle of the functionalist perspective which perceives behavioral outcomes to be dependent on different mental states and sensory inputs. In other words, individuals are expected to perform, based on the environment and the input that they receive from it, as well as the functions in which they are expected to perform [Block, 1980].

Snyder and Cantor [1998] position these functions into four different groups: (a) Value Expressive: where individuals express their values out of altruistic concern (b) Utilitarian: where individual seek rewards from the external environment (c) Social adjustive: where individuals do a certain thing in order to fit better with the group and (d) Knowledge Seeking: when an individual may be performing a task in order to get a new learning experience that he might use later. But can this functionalist perspective be fitted in the context of an online community? This is the research issue that we are interested to examine in this dissertation. In particular, what are the motivational factors that affect the participants in an online community and make them to become more engaged?

Table 1.4 summarizes the above functions and the research objectives pursued. Taking this classification as a guide, we provide the three basic research objectives of this dissertation in the section that follows.

1.2 The research objectives

Having provided thus far the theoretical background and the motivation for the research issue that this thesis is targeting, we now frame a set of three research objectives based on the following arguments:

1. The research question should address the importance of social capital (in its virtual form) as an important factor for understanding participation in online communities. Therefore the research objectives should take into account both the behavioral properties of the participants, their interactions as well as the nature of the shared resource that these interactions are formed around.

2. The extrinsic and intrinsic form of the outcomes that the users expect to accomplish by participating in an online community. This is of particular importance if we take also into account the opportunity cost that a participant might have by participating in an online community and

3. The importance of social mechanisms that facilitate these types of online (or virtual) interaction should be highlighted among the individuals that form an online community.

For the above reasons we chose the following three research objectives which are described in the following paragraphs.

Research Objective A: What is the nature of the incentives that affect contribution in online communities?

In Organizational Science there has been considerable literature that discusses the effectiveness of incentives as a mechanism for empowering individuals in several contexts, especially as a key component of agency theory in organizations [Eisenhardt, 1989]. Gibbons [1998],for example, has studied the impact of incentives in an

organizational setting and provides evidence that incentives promote effort and performance. When it comes to the psychological interpretation of the term, incentives are seen as psychological *stimuli* or influence factors, responsible for modifying the behavior of the individual under a specific setting. Behavior is often operationalized as a blend of several sources of action which can be seen by the reaction of individuals in different settings (e.g., a buyer's behavior might change as a reaction to the increase of the price of a good). In the framework presented by Snyder and Cantor [1998], this can fit with the Value Expressive and Utilitarian cluster of functions since this captures both extrinsic and intrinsic forms of incentives.

In our case we study incentivized behavior by using a special model of social interactions that takes place in a class of social interaction structures called exchange networks[Cook and Whitmeyer, 1992]. Cook and Emerson [1987] approach the definition of an **exchange network** as a network structure where the directional relationships imply exchange of resources either of material nature (e.g. economic and business relations) or of a non economic relations such as power relations (where individual has an authority or power to influence another individual). According to Cook and Emerson [1978], exchange networks are formed by the mapping of exchange activity among different individuals into dyadic relations. Unlike standard network theory where individuals are concerned to interact with the whole set of social structure (e.g. the society or the market) an exchange network aims to analyze the reciprocal advantage that an individual might draw from the engagement of an individual to a form of transaction. As in any other type of network analysis [Borgatti and Foster, 2003], in exchange networks the fundamental unit of observation is the dyad (the reciprocal connection between the minimal actors required for an exchange). Two dyadic relations in an exchange network are connected in the situation where the one relation can be *contingent* on exchange (or no-exchange) on the other relation. The particular nature of this connection can be either *positive* or *negative*. A positive connection provides that the one exchange relation is contingent on exchange with the other

while a *negative* exchange relation is non-contingent on exchange with the other. In greater detail Chapter 3 provides a controlled study with a large set of participants where this particular research question is partially explored as well as in the case of Chapter 6 where apart from the incentivized contribution we look in to the nature of the contribution itself.

Research Objective B: How do participants of online communities perceive the nature of their contribution?

The second research objective of this thesis relates to the participants' perceived nature of the contributions to the online community. In order to carry out this research, we consider behavior as a variable of latent nature, which implies that in order to consider behavior as a unit of observation (or the dependent variable), this has to be expressed as a set of other factors-variables which can be operationalized by methodical observation (either in a controlled setting or a field setting). We approach the case of contributions by using three preliminary types of contributions in an online community based on the approach of studies of collective action discussed in the previous sections. We categorize this approach into three different forms: (a) The case of contributions as *contributions to the public good*, (b) the case of contributions as a reciprocal action and (c) the case of contribution as a compliance with the social norms that characterize the group

A contribution to the public good is a dominant form of perceived nature of contributions where collective action occurs also in offline settings. What makes this perspective interesting from a research point of view is *how participants perceive the public good nature of the content of the online community*. As will be argued in the next chapter, contributions in online communities can be seen as contributions to the public good since the consumption of the outcome is available to anyone; as long as it remains free, there is no rivalry among consumers for the consumption of the information (as in the form of a good in a tangible form). This resembles the characteristic of a

public good as possessing the nature of non rivalry and non excludability [Samuelson, 2000].

On the other hand, some of the community participants are not incentivized by the connection of the community outcome as a public good due to the fact that they do not perceive the production of the collective action as a public good. Following the metaphor of exchange between two parties, reciprocity is in principle an exchange of the same or equivalent resources that the one party has given to the other. An exchange relation is reciprocal when participants perceive the value of an item to be exchanged to be of roughly equivalent value to the one received (in case of a dyadic exchange). When they are forced by other participants into the terms of the exchange, it means that, instead of seeing the contribution to the online community as a contribution to the public good, they are reciprocating interaction instantiated by other community participants. In such cases, there is usually a trigger event which can be in the form of a message around a particular subject that intrigues the participant to contribute (e.g., the case of an article in Wikipedia which contains no accurate information).

There can also be the situation where community participants perceive their participation as something that is obligatory, and this is very much related to the cohesiveness of the online community, as in the offline settings [Gross and Martin, 1952]. This suggests a dichotomization of the factors related to the previous research question into those provided by the social environment or the consequences of social activity, and those provided by exogenous factors, as in the case of a contract enforcement mechanism which will be discussed in 4 and 5.

Research Objective C: Do participants of online communities care about the contributions of others ?

The discussion focusing on the case of collective action considers the cooperation of participants to be an important element that characterizes the outcome of this

cooperative effort. However, there might be a situation where some participants rely on the effort of others in order to get the same benefit but with less effort. In other words, participants might consider the level of activity of others in order to participate or contribute effort in an online community. The purpose of this research objective is to evaluate whether participants do care about the level of activity of others or are agnostic in the way that they do not care about the activity level of the others (e.g., the contributions that they have already provided to the online community) or their status.

In these environments, on the other hand, it is difficult for the participants to clearly observe this due to the high amount of interaction taking place. However, since this interaction is codified, participants have the ability to trace the activity of the other participants and compare it with their own. Therefore, we need to be able to answer whether such a social comparison process exists and what the output is of these processes.

Having provided the problem formulation and the research objectives of this thesis, we present the methodological approach that was employed to study the above stated research objectives.

1.3 Research approach and methodology

When it comes to addressing the research questions defined above, we need to be able to clarify (a) the nature the research inquiry that we address (b) the research methodology that we are going to use in order to handle them and (c) ways to assess the external and internal validity of the findings that we will obtain by pursuing the stated research questions.

1.3.1 Online communities and Information Systems research

In order to comply with the main elements of the scientific method, a research inquiry has first to be defined within the scope or the "epistemology" of a research field. The epistemology of Information Systems (IS) addresses the study of technology oriented phenomena and its relation between people and organizations. As a field, it is established around the socioeconomic implications of the use of Information Technology in both Micro (individual) and Macro (Organizational) levels. Thus, we need to clearly define the unit of observation, after which we will also define a set of constructs that will be used to address the research questions that we pursue. This is in contrast to the social science perspective which provides that the unit of observation is clearly the individual. The network based perspective focuses on the characteristics of the relations rather than the individuals themselves (such as personal or behavioral characteristics).

Following the network perspective, we need to emphasize the assumption that actors of a social system form relational ties which can be interpreted differently under certain contexts. For example, in a social system there are often several different types of relations which have a different consequence for individuals that are part of them. For example, Padgett and Ansell [1993] studied the network of different relations among Florentine families in the early renaissance in order to explain the rise of the *Medici* family as rulers in renaissance Florence. This research example, although it can be considered to be distant from the research subject pursued in this thesis, it does have a common viewpoint on the research perspective: the network structure implies different access to resources for individuals that form relational ties. The authors' research results provide an understanding of the relation between ties of different strengths (case.g., family ties through marriages and business ties).

1.3.2 The positivist view in IS research

Methods of scientific enquiry start by making a basic distinction between two funda-
mental concepts, namely the **epistemology** and the **methodology** aspect of the
research that is undertaken. On the one hand epistemological aspect of a research
inquiry defines the process of how we come to know the properties of the research
inquiry such as the extent to which the outcomes of the scientific inquire are still valid
and are safe to be used to address a problem. In other words the epistemology of the
fields defines the boundary between the beliefs and the scientific truths that can be
extracted by addressing the research question formulated.

On the other hand the methodological aspect of scientific inquiry deals with the
more practical perspective of the epistemological view of the field. It is not only con-
cerned with *"how we come to know"* which provides the background knowledge about
it, but it focuses mainly on the specific ways of scientific inquiry-methods - that pro-
vide a tool for understanding and validating the background knowledge known about
this scientific inquiry.

rom this perspective, the positivist view of reasoning in IS research suggests that
the researcher analyzes results concerning a technology oriented phenomenon based
on statistical and formal reasoning, rather than on perspectives theorizing about in-
dividuals' actions such as in action research [Baskerville and Wood-Harper, 1996].
However in order for a positivist research study to be able to provide a meaningful
set of concluding remarks a set of steps need to be undertaken. Figure 1.1 provides a
summary of the stages in a positivist research based inquiry in the field of information
systems. The basic step in those procedures is the selection of the type of research
that will be conducted. This can be either *exploratory* or classical/*confirmatory.*

Exploratory research approaches data analysis from a perspective that it permits
the researchers to employ an analysis which allows the data itself to reveal an under-
lying structure [Tukey, 1977]. This approach can be generally seen as an inductive
view of the research inquiry that is not based on any preexisting model or theoreti-

cal pathways in order to test or confirm a theory. By employing this approach, the researcher gets an overview of the data and the research issues that can be initially deducted from a preliminary data analysis. This is grounded in the assumption that the more a researcher knows about the nature of the data, the better these data can be used to construct a model in order to test or define a new theory.
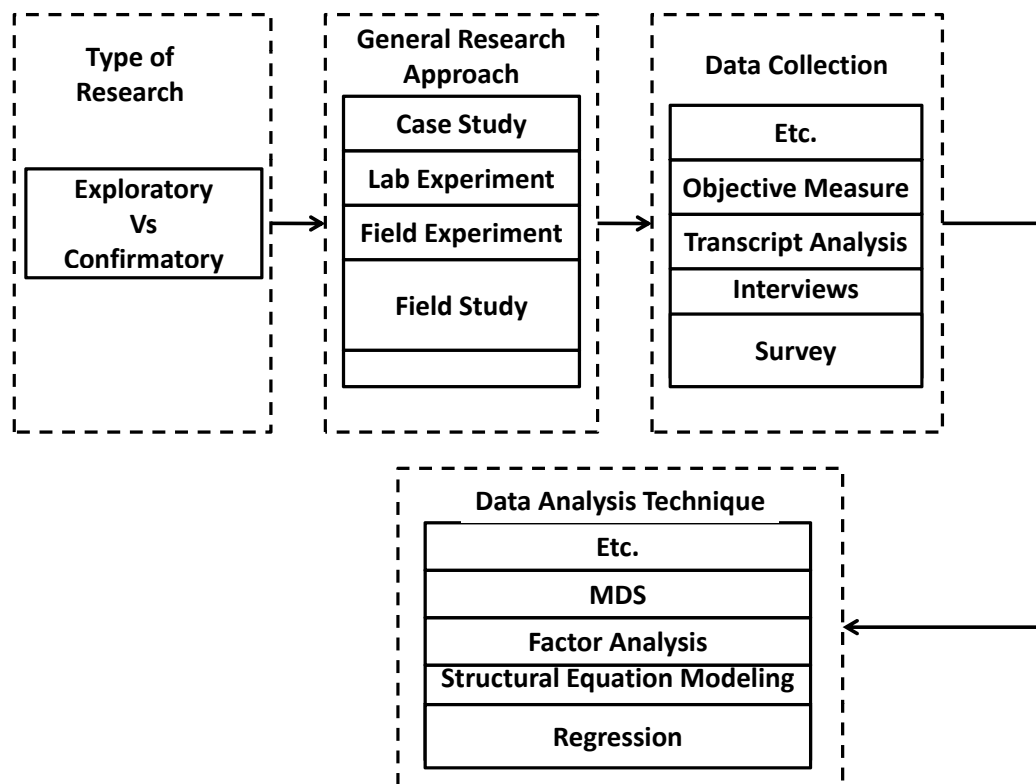
Figure 1.1: The stages of positivist research in information systems research. Adopted from Boudreau et al. [2001]

Exploratory data analysis employs a heavy use of descriptive statistics in order to provide some basic descriptive measures that can give researchers an idea about what the dataset describes and how the different variables are connected. For example, in the case of a dataset consisting of several variables, before the researcher can use a statistical technique such as regression analysis, it is always better to have an overview of the correlation between the variables for an overview of which dataset variables are useful for inclusion to the model or not. Visual representations of the

relations between the variables (e.g., scatter plots) can also provide a meaningful way of getting an overview of the properties of the dataset.

Hartwig and Dearing [1979] provide an overview of exploratory data analysis techniques and address two basic principles in which the researcher should ground the inquiry. The first has to do with *openness of the empirical observations*. This perspective provides that analysts should be open to address unexpected results from the preliminary data analysis and reformulate their assumptions accordingly. The second has to do with the case of *skepticism* from the side of the researcher. This is due to the fact that the exploratory data analysis techniques might conceal aspects of the empirical observations that are important for the research questions to be addressed.
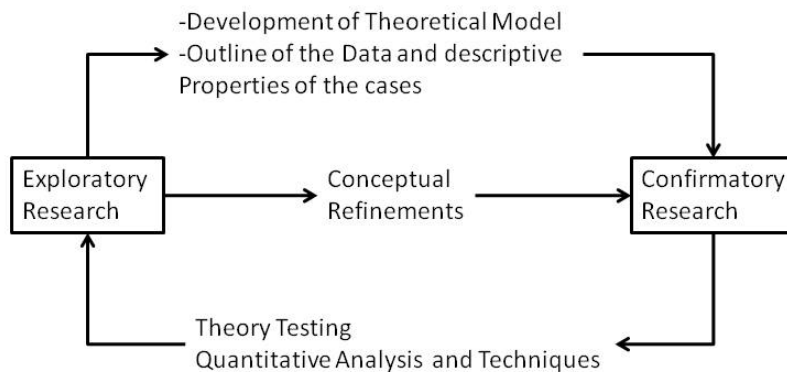
Figure 1.2: Confirmatory vs. Exploratory Research. Adopted from Straub [1989]

Confirmatory research, on the other hand, adopts a deductionist perspective in order to test whether an already developed theoretical paradigm is appropriate to explain a phenomenon or not. Following this perspective confirmatory research is relying basically on inferential statistics (e.g. hypothesis testing) in order to test a model that explains the phenomenon under observation.

As can be seen in the figure confirmatory research has several differences with the exploratory data analysis based on the following:

- Confirmatory data analysis starts with an already specified model that uses the

Empirical Study	Research Approach	Data Analysis Type
Chapter 3: Public Goods Experiment	Experiment	Confirmatory
Chapter 4: Yahoo! Answers	Field Study	Confirmatory
Chapter 5: Google Answers	Field Study / Field Experiment	Exploratory
Chapter 6: Amazon Reviews	Field Study	Exploratory

Table 1.5: Research and Data Analysis Approach for the chapters covering the empirical part of this dissertation.

empirical observations of the dataset collected in order to evaluate the fit of the model for this particular case. On the other hand exploratory data analysis starts with the analysis of the data in order to construct a theoretical model that corresponds to the empirical observations gathered.

- Confirmatory data analysis relies heavily on formal methods in order to evaluate a model such as regression and analysis of variance (ANOVA), whereas exploratory data analysis is more grounded in the relations between variables and fewer formal instruments that display the importance of these variables to the effect that is studied.

- In contrast with the explanatory data analysis where no or few assumptions are made in the beginning, the use of formal methods employed by confirmatory research (for instance, the use of inferential statistics) requires strong assumptions related to the internal validity of the process and the interpretation of the results obtained.

- On the other hand exploratory data analysis produces more general results but is less sensitive whereas confirmatory based approaches have more predictive

power over the phenomenon studied.

With respect to the above two perspectives, we used the concepts of induction and deduction in order to demonstrate the differences between the two research methods, which rely on *Inductionist* and *Deductionist* approach respectively. However one must consider that depending on the research question studied a combination of other perspectives might be used as well (e.g., the case of triangulating the research results with other approaches).

Induction is the logical process of generalizing about a phenomenon based on observations that a researcher has had from a particular instance of it. Thus *Inductionism* is used in order to produce theory from empirical observations and an Inductionist perspective contributes to the development of a new theory or the extension of an existing one [Lee and Baskerville, 2003]. However is not always feasible to have an Inductionist viewpoint in a research setting due to issues related with the *external validity* of the empirical observations which is very much dependent on the research design itself.

Deduction on the other hand is a process of reasoning that uses existing knowledge to "deduct" statements about a particular phenomenon. The deductionist perspective uses already developed theories and methods in order to address a specific instance of a phenomenon [Lee, 1989]. For example, in this thesis a deductive reasoning would suggest that an existing theory would be used in order to address a research question raised through the theoretical grounding of this thesis. Deductionism, as a research paradigm, is very much connected to positivist perspectives in information systems research. When it comes to analyzing empirical observations related to the research question that we pursue, it is very common to ground statements on hypotheses which are later use in accordance with statistical instruments in order to falsify (reject) or approve them (accept). When we test hypotheses using the deductionist perspective and if the empirical observations support them, the theory is either confirmed or rejected.

According to Boudreau et al. [2001] the cases of Inductionism and Deductionism

represent the exploratory and confirmatory part of a research study. However as it can be observed in Figure 1.2, there is the need in a research study for conceptual refinements on the exploratory part taken from what we receive from a confirmatory study, in which case there is the need to use triangulation as a way of combining both results. Essentially, triangulation in a research study refers to the use of several approaches into the scientific inquiry raised by the research study in order to raise the level of confidence on the results that the study produces. Research approaches in social sciences often disagree on the use of methods (e.g., Quantitative Study vs. Qualitative Study) and often offer limitations on the level of addressing the research questions formed. Usually these limitations are derived from the disadvantages of the research method or the limited use of the method due to factors such as the range of the data or the sampling procedure that is followed.

1.3.3 Methodology

The use of experimental methods

One particular quantitative method that is used in this dissertation (of confirmatory approach) is the use of controlled and field experiment methods. Apart from chapter 4 which is a controlled experiment, the next three chapters can be also seen from an experimental viewpoint since they represent artefactual field experiments (studies based on intergroup evaluation grounded on a particular characteristic).

Merriam-Webster defines the term experiment as *"a procedure carried out under controlled conditions in order to discover an unknown effect or law or to test or improve an existing hypothesis".* As can be observed in Figure 1.3 experimental methods are the bridge between the epistemological concepts of theory and empirics. Theory is set of well structured and consistent statements that suggest an empirical way to prove them. Empirics define all the theory –related phenomena that can be observed under the context that the theory is describing. The development of empirics and theory is an iterative process. Theory suggests and modifies empirics, while empirics

41

test and modify existing theoretical assumptions. Consistent theories use empirics as a tool of verification while emerging theories use empirical observation and testing as a method of improvement [Friedman and Cassar, 2004].
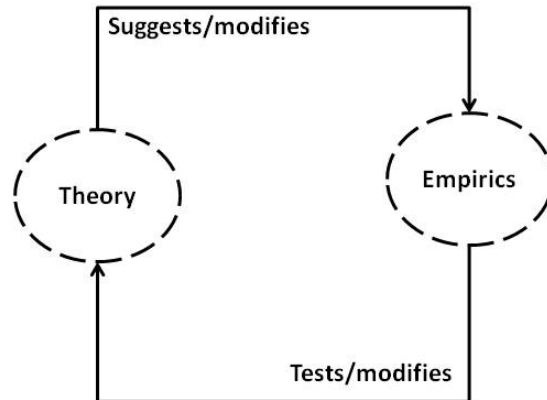
Figure 1.3: The general epistemological framework for the relation between theory and empirical observation. Adopted from Friedman and Cassar [2004]

As encapsulated in the definition, an experiment is a structured procedure. In this procedure we have several inputs (from where the empirics are going to be constructed) and stages that arise through the different alternations of the procedure itself. The different elements that are part of the experimental procedure are:

- **Subjects:** The participants of the experiment whose behavioral characteristics are observed during the experiment.

- **Treatments:** A part of the experiment where the variables relate to the experiment, or the subjects are altered in order to observe the behavior. It should be noted that every experimental procedure is subject to a set of "treatments" where the original procedure might be altered, controlling for other experimental factors.

- **Input:** Usually a set of parameters including subjects or a scenario

- **Output:** The behavioral characteristics obtained through a set of stages (treatments).

Psychology scholars have been the early adopters of experimental methods outside the field of the core natural sciences. Due to the obvious connections with medical and clinical experimental practices experimental psychologists have formulated the basic principles and methodologies for conducting experiments that test and modify behavioral observations Reips [2002]. Through the study of anomalies and framing effects using laboratory controlled settings, the use of experimental techniques have been advocated to other fields of the social sciences, the most significant application being in the field of behavioral and experimental economics [Smith, 1982, Tversky and Kahneman, 1974].

Empirics constitute the outcome of the experiment.Friedman and Sunder [1994] provide a classification of data characteristics regarding empirics depending on whether they are controlled or not. According to this classification, empirical data can be classified in the following categories:

- **Happenstance data:** which are gathered and produced under non controlled settings. This kind of data is used in the studies presented in Chapters 4,5 and 6.

- **Experimental data:** which are a product of a scientifically designed setting (e.g., a task is given to subjects to be accomplished). This class of data is used in the study presented in Chapter 3.

Depending on the setting that these data have been collected we might have either *laboratory data (e.g. from an experiment carried out in a computer lab)* or *field data* that have been gathered from a field experiment (experimental setting out of the context of a laboratory) or a controlled data collection. Data from field experiments differ in that they are not strictly controlled by the laboratory which may result in some inconsistencies with the treatments. On the other hand, field experiments resemble, in a higher degree, the realistic principles under which the experimental setting is tested (e.g., a trust scenario etc).

The key characteristic of any set of experimental data is the **ceteris paribus** con-

dition that is imposed at any experimental stage (*treatment*). The **ceteris paribus** condition is important because we can use it to identify the consequences of each individual effect to the whole outcome of the experimental treatment.

Let us consider a process X that takes inputs A and B and produces an outcome Y. Suppose that we want to measure what was the effect of B in the process X–>Y. To do that we consider two treatments: the first (t_1) is the treatment with all the elements normal, the second (t_2) considers an alternation of B to B* so the process becomes X->Y*.
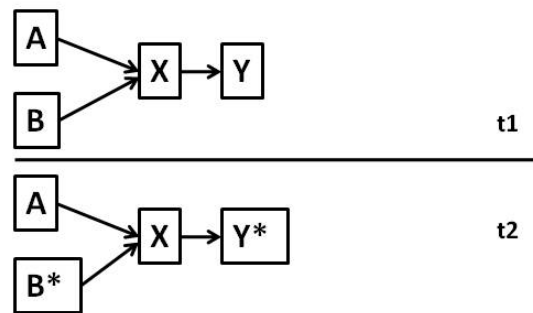
Figure 1.4: Two treatments under the ceteris paribus condition

Under the ceteris paribus condition we will demand the following:

A (t_1) = A (t_2)

X (t_1) = X (t_2)

This means that the setting is controlled under ceteris paribus condition if and only if under alternation of B in treatment 2 the input /variable A remains the same in both treatments and the process transformation is exactly the same. After that and in order to measure the effect we consider the Δ (B, B*) with Δ(Y, Y*).

The ceteris paribus condition complements the two basic assumptions that are used in any empirical study. The first has to do with the *hypothetical isolation* or the case where we hypothesize that the factors or variables that we focus on in a study solely describe the unit of analysis that is defined. There might be other factors as well (endogenous or exogenous), however, this assumption permits us to study an effect of
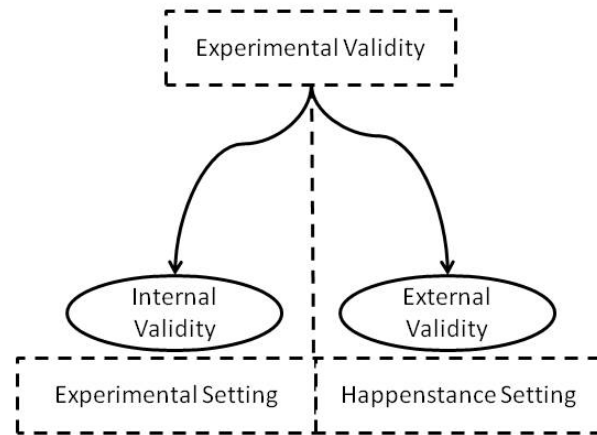
Figure 1.5: Aspects of Validity for Experimental Results

one situation on another, as they were completely exogenous to the other. For example, the impact that the observed series of actions of a user will have an impact on the way the other participants perceive this user as an important participant in an online community or not. However, since empirical studies as the ones that are contained in the thesis also consider complex environments where the hypothetical isolation of factors can mislead to different results, another situation is when the ceteris paribus condition provided is the case of temporal isolation where we consider that other factors (e.g., the community size) change slowly during the research study. While ceteris paribus is not the only condition that a research study can assume, this is a safe way for the interpretation of the results that provide a discussion on the approximate effects, rather than safeguarding effects based solely on assumptions (something that can be dangerous for social studies in general).

An experiment as every other empirical method needs to have some validity measures that can express how safe is to induce results and validate hypotheses. In principle *experimental validity* is an indication of the validity that the results of the experiment can advocate. There are two aspects of validity namely *internal validity* and *external validity*. Internal validity resembles how reliable the independent variables are to support the behavioral observation. A conclusion drawn from an experiment

has verified internal validity when an independent replication of the same experiment provides the same likelihood of results. *External Validity* refers to the happenstance resemblance of the validity obtained by the experiment in the real world. In other words the external validity of an experiment describes how valid are the conclusions of the experiment to the real world.

Although internal validity is observable and can be assembled by statistical techniques and tools, external validity is very difficult to obtain due to the fact that the settings facilitated in the experiment are not easy to apply to the real world. Apart from this there are several cases of experiments where the context is different than that of the lab where external validity is difficult to induce due to the high complexity from the environment.

Criticism and validity

For the aforementioned reasons the positivist perspective on information systems research provides a more sufficient ground for pursuing the research question framed on this chapter. However on the other hand, scholars often cite as a major criticism of the positivist perspective in information systems research the fact that it tries to account the unpredictable complexity of the human nature by employing the use of isolate models that incorporate a set of static variables [Myers, 1997]. Furthermore it is argued that positivists look for patterns around the case that they examine thus reducing the complexity that is provided by the human dynamics of the actors. To some extend this case deals with the operationalization of variables related with the research inquiry that is undertaken by the scholars. The employment of latent models for the description of variables is a standard practice that is used by behavioral scientists such as psychologists in order to comply with this criticism.

Another major criticism on the perspective of positivism and positivist research in Information systems research *is whether the instruments that are used for reasoning are reliable or not*. Straub [1989] in an early study examined the use of various tech-

niques of quantitative research in mainstream information systems research journals and reported that only the 17% of the articles reported reliability of the scales used. In other ways only few of these articles were validating the constructs that were used to assess the research questions posed. Boudreau et al. [2001] provided a follow up to this study which examined a broader collection of articles in information systems research journals. The results indicated an average of 55% of all the quantitative studies provided in this set of journals was reporting measures of validity both in pre-liminary phases (pilot test) and after. The findings suggested that positivist view is moving slowly to address the issues raised by Straub's suggestions. Furthermore the development and validity of instruments based on structural equation modeling for example [Chin, 1998] has provided a better ground for positivist research in informa-tion systems.

Structural instruments[Jobson, 1992] very often provide two perspectives on the way that they emphasize the output of the model of study. The first is the formalized or econometric perspective which is merely focused on prediction and the significance of the factors that characterize the model. The other perspective which is widely used in information systems is emphasized on psychometric analysis which models con-cepts as latent (unobserved) variables. The late provides that the units of study are not operationalized by a direct observation but are indirectly inferred from several ob-served variables which in the literature are often referenced as indicative or manifest variables.

In the empirical part of this dissertation we make extensive use of estimation mod-els based on ordinary least squares (OLS), as well as logistic regression models (TOBIT, PROBIT). This was done in order to provide (a) an evaluation of the significance level for the coefficients both from the perspectives of size and significance level, (b) an evaluation of the case where common measurement issues such as homoscedasticity might affect the significance level, and most importantly, (c) defining upper and lower limits for the dependent variable since in the studies we were theoretically informed

about the actual range that the dependent variable was having. Furthermore, distribution assumptions, such as normality, might also violate the estimator's trustworthiness if the dependent variables are not normally distributed. A logistic estimator can become independent of that issue and complement the case where the coefficient's size is significant only upon this assumption. The same case applies to the comparison between groups since the assumption of normality entails the danger of statistical significance being present at a wrong level. This is avoided by testing for normality (using the Kolmogorov-Smirnov test), as well as using non parametric tests such as the Mann-Whitney U test.

1.4   Structure of this dissertation

Having provided the research questions and method selection approach for this dissertation, we now supply an overview of how this dissertation is structured in order (a) to demonstrate the connection with the empirical part that covers the research questions provided above and (b) to provide an overview of the data used in the empirical part.

1.4.1   Chapter structure

As can be seen in Figure 1.6 the dissertation consists of two parts: a theoretical part consisting of two chapters (this introductory Chapter and Chapter 2), where the research question is underlined and an empirical part (Chapters 3, 4, 5 and 6) where the connections with theory are highlighted in four different application settings. To this end, the empirical part provides a collection of four chapters structured according to the aforementioned research objectives to support conclusions related to the research questions formed in this and the subsequent chapter.

Chapter 7 presents the conclusions of this dissertation and retrospectively to the research results summarized in the previous chapters, provides a more analytical view-

point regarding the practical implications and contributions of this research.
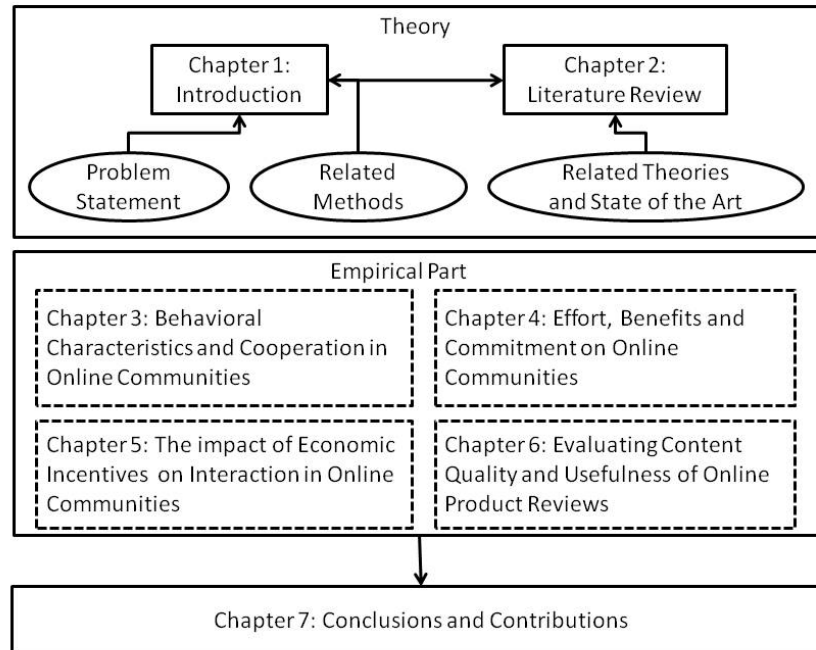
Figure 1.6: The structure of this dissertation

To rest of this dissertation is structured as follows.

Chapter 2: Models and Theories for Understanding and Motivating Contributions in Online Communities

This chapter provides the background of studying online communities and the social interactions that characterize the activity of their members. In particular, an attempt is made to summarize the findings of resent research studies and methodological frameworks that are in relation to motivating contributions to online communities. The chapter approaches the research context and theories related to research on online communities by adopting a social interaction based approach in order to illustrate the rich social environment that is encapsulated in an online community.

Chapter 3: Behavioral Characteristics and Cooperation in Online Communities: An Experimental Investigation

In this chapter we present the results of a large Internet experiment that was run to evaluate whether factors such as trust and fairness play an intuitive role in the formation of attitudes towards contributions to online communities. The chapter describes the experimental procedure and the theoretical expected results and then continues by providing an analysis of the data and a discussion of the findings. This chapter is important for the empirical part since it presents a non-contextual study, such as the studies presented in the next three chapters, but evaluates attitudes and behavioral characteristics of subjects that already participate in online communities. The findings of the study presented in this chapter are also used in Chapter 7 to evaluate the findings from Chapters 4, 5 and 6 in order to provide the conclusions and contributions that can be drawn out of this thesis.

Chapter 4: Effort, Benefits and Commitment on Online Knowledge Communities: An empirical study on Yahoo!Answers

This chapter provides an evaluation of interactions taking place in a widely used online community system operated by Yahoo! and branded under the name: Yahoo!Answers. We provide an analysis of the interactions and the activity taking place in order to evaluate whether contributed effort as motivated by exclusively intrinsic forms of motivation is affected by the level of individual interaction of those participating in the platform. The chapter argues, based on empirical findings, in favor of a "cyber-public goods" theory taking place in the realm of the Yahoo!Answers service which is based on the perceived utility of the answers that users get from the service. Users then assess how much they participated in Yahoo!Answers based upon what they receive. The chapter argues in favor of the ??latter as a way of enforcing participation and motivating contributions with an effect into an increased user activity.

Chapter 5: The impact of Extrinsic Rewards on Strategic Interaction in Online Communities: An analysis on Google!Answers

Interaction in an online community system called Google Answers which has been operated by Google for a period of four years is explored in this chapter. We study extrinsic forms of motivation as depicted from the monetary incentives that the platform uses to motivate its exclusive list of participants to provide high quality answers to questions posted by users in the platform. This particular chapter is closely connected to Chapter 4 from an interaction context perspective, the basic difference being a concentration on the nature of incentives that characterize the participants. In the context of this chapter, participation is motivated by extrinsic (monetary based) and information demand side factors which means that the mode of operation is almost the same (online answers community).

Chapter 6: Evaluating Content Quality and Usefulness of Online Product Reviews

This chapter provides a different study context compared to the study contexts provided in the previous two chapters by addressing a case of online communities where the objectives of the users can contradict each other. An online community of reviews is formed around a specific product on the Amazon.co.uk website. In order to evaluate this approach, we use a set of two measures: an objective measure as measured by the aggregation of the individual preferences of the community members provided in the website, and a quantitative based evaluation approach based on content analysis.

Chapter 7: Conclusions and Retrospect

The final chapter provides the conclusions and the connection of the findings by connecting the theory with the results obtained from the empirical part. As aforementioned we address the findings from Chapters 4, 5 and 6 in relation with the findings

from Chapter 3 and the theoretical background that we provided in this chapter and Chapter 2. The related contributions to theory and the practical implications of this dissertation are highlighted, and an outlook for future research in the subject is offered.

We now provide an analytical overview as well the description of the datasets used in order for the four chapters of the empirical part to provide the analysis presented in each of these chapters. We summarize the dataset and the findings of each individual chapter in the following two sections.

1.4.2 Datasets used in the empirical part

For the purpose of the analysis provided in the empirical part of this dissertation a set of four datasets were collected in various periods of the PhD process. With the exception of the Public Goods Experiment Dataset that was made accessible due to the researchers' own involvement in the preparation of the experiment, the other three datasets summarized here and used in the studies presented in the Chapters 4, 5, and 6 which are a result of the programming effort of gathering the datasets. Each of the datasets is structured as a panel with an identifier attached to each individual interaction session. For the Public goods experiment the interaction section is linked to a user (subject) since the interaction data are unique for each subject. For the Yahoo!Answers, Google Answers and Amazon Reviews dataset the panel structure is keyed on the interaction section (Question in the case of Yahoo and Google Answers and review for the case of Amazon Dataset).

Table 1.6 summarizes the qualitative characteristics of these datasets. The number of interactions is used as the main descriptive variable in order to demonstrate the volume of social activity that was recorded in each of these datasets.

Bellow, we provide a description of each of the datasets used in the empirical part.

Dataset Name	Category	Number of Interactions	Dataset Description
Public Goods Dataset	Controlled Experiment	3200	Dataset from an online Internet Experiment
Yahoo!Answers	Field Study / Online Community	65000	Partial Data collected from the Yahoo!Answers online Service
Google Answers	Field Study / Online Community /Information Market	150000	Complete data from the GoogleAnswers online Service
Amazon Reviews	Field Study / Online Community	12000	Partial Data collected from the books section from the UK store of Amazon

Table 1.6: Datasets compiled for the purpose of this dissertation and used in the Chapters 3, 4, 5, 6.

The public goods experimental Data

The data for this chapter were collected from 22 May to 15 June 2008, as the result of a large scale internet experiment with paid subjects. The experiment was a part of the iLEE project which is a project funded by the Carlsberg Foundation[6] and is coordinated by the Department of Economics at the University of Copenhagen in collaboration with Denmark's National office of Statistics and the Danish Institute for Local Government Research[7]. A random sample of the Danish population was picked by Denmark's statistical agency and a total of 21.052 invitation letters were sent to subjects covering all the geographical districts of Denmark. The subjects who responded to the invitation logged in to the online platform developed for this purpose and were then randomly assigned into groups of four members. As will be elaborated upon in Chapter 3 the subjects played a public goods game where they had to make decision about how much of the endowed amount that was given in a later stage they wanted to keep and how much they wanted to contribute to an investment that would provide double return with equal contribution by all group members.

The dataset contains, apart from the decision variables, a set of demographic variables and the variables related to the subjects' attribution towards online communication methods, trust and fairness perception (using the Danish version of the world value survey), as well as personality characteristics (using the Standardized 5 elements personality test). For the purpose of the study presented in chapter 3, we use only the decision and demographic variables, as well as the online communication variables contained in the dataset.

The Yahoo Answers dataset

Yahoo!Answers is an online question answering service operated by Yahoo. The dataset that was collected for the purpose of Chapter 4 contains interactions covering three

---

[6] http://www.carlsbergfonden.dk
[7] http://ilee.econ.ku.dk

months between users that ask and answer questions in the platform (the period was from 01/06/2008 until 01/09/2008). The dataset was collected by programming an automated crawler that visited the platform every day for this period and registered the data to a relational database for further processing. Due to the anonymized setting of the Yahoo!Answers service interface, the dataset contains only interaction data (user *A* answered the question of user *B*) and no socioeconomic and background data, such as age, country, etc.

The Google Answers dataset

Google Answers was an online information marketplace deployed as a beta software project[8]. This chapter explores a unique opportunity, resulting from the shutdown of the online service Google Answers, to test if tips were paid for strategic reasons in a quasi-experimental setting. Google Answers was a marketplace of information inquiries in which any "asker" could post a question along with a price to be paid for a satisfactory answer. One researcher from a closed group of "answerers" would answer the question, usually by providing reference to authoritative sources. Upon receiving an answer, the asker rated the quality of the answer obtained and, if satisfied with the quality, paid the price and a voluntary tip.

We investigate tipping behavior before (when strategic considerations can play a role) and after (when they do not) the announcement of the shutdown of the answering service. To disentangle the strategic motive of tipping from other (reciprocal or norm-driven) motives of tipping, we analyze pre-announcement tipping behavior. The empirical results suggest that users use tips to induce better (e.g., high prompted) service in the future, and that answers respond to tipping by providing services more promptly.

---

[8]The term beta describes software that although is functional is still in testing phase and implies that users use the software under their own risk. Usually a beta version of a software product precedes the final.

The Amazon Reviews dataset

Amazon operates an interesting web service which is part of the product or item description page, which contains reviews submitted by customers who have used/ or have an opinion about the specific product. By using Amazon's web service interface (REST) we were able to retrieve data for products listed in a random order for a specific section and then direct specific queries to retrieve the customer submitted reviews for this item/book listed in the specific section. For the research purposes covered in Chapter 6, we used only a specific section of Amazon's online service, namely, Amazon.uk (to avoid issues related to language competency, etc). The dataset contained interaction data, as well data that were generated for the research questions addressed in Chapter 6. The data contained item specific variables, such as the category of the product (all the items were books) and rating / review specific data such as the position of the review that we were examining in the review page of the item (third review on the first page, etc).

Bibliography

A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. *Lecture Notes in Computer Science*, 4258:36, 2006.

P. S. Adler and S. W. Kwon. Social capital: Prospects for a new concept. *Academy of Management Review*, 27:17–40, 2002.

G.W. Allport. Attitudes. *Handbook of social psychology*, 2:798–844, 1935.

T. Anderson and H. Kanuka. On-Line Forums : New Platforms for Professional Development and Group Collaboration. *Journal of Computer-Mediated Communication*, 3(3), 1997.

A. Armstrong and J. Hagel. The real value of online communities. *Knowledge and communities*, pages 85–95, 2000.

K.J. Arrow. *Observations on social capital*, pages 3–5. The World Bank, 2000.

Azy Barak. *Psychological Aspects of Cyberspace: Theory, Research, Applications*. Cambridge University Press, 1 edition, May 2008. ISBN 0521873010.

J.B. Barney. Firm resources and sustained competitive advantage. *Journal of Management*, 17(1):99–120, 1991.

C. A. Bartel and R. Saavedra. The collective construction of work group moods. *Administrative Science Quarterly*, 45:197–231, 2000.

R.L. Baskerville and A.T. Wood-Harper. A critical perspective on action research as a method for information systems research. *Journal of Information Technology*, 11: 235–246, 1996.

N.K. Baym. The emergence of community in computer-mediated communication. *Cybersociety: Computer-mediated communication and community*, pages 138–163, 1995.

B. Bimber, A. J. Flanagin, and C. Stohl. Reconceptualizing collective action in the contemporary media environment. *Communication Theory*, 15:365–388, 2005.

N. J. Block. *Readings in Philosophy of Psychology*. Routledge, 1980.

S. P. Borgatti and P. C. Foster. The network paradigm in organizational research: A review and typology. *Journal of Management*, 29:991, 2003.

M. Boudreau, D. Gefen, and D. Straub. Validation in is research: A state-of-the-art assessment. *Management Information Systems Quarterly*, 25:1Ű24, 2001.

P. Bourdieu. *The forms of capital*, pages 241–258. 1986.

S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30:107–117, 1998.

G. Brown, T. Howe, M. Ihbe, A. Prakash, and K. Borders. Social networks and context-aware spam. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pages 403–412. ACM New York, NY, USA, 2008.

S. L. Bryant, A. Forte, and A. Bruckman. Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *Conference on Supporting Group Work: Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, volume 6, pages 1–10, 2005.

R.S. Burt. *Brokerage and Closure: An Introduction to Social Capital*. Oxford University Press, 2005.

B. S. Butler. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, 12:346–362, 2001.

W.W. Chin. The partial least squares approach to structural equation modeling. *Modern Methods for Business Research*, 295:336, 1998.

J. S. Coleman. Social capital in the creation of human capital. *American Journal of Sociology*, 94:95, 1988.

K. S. Cook and R. M. Emerson. Power, equity and commitment in exchange networks. *American Sociological Review*, 43:721–739, 1978.

K.S. Cook and R.M. Emerson. *Social exchange theory*. Springer, 1987.

KS Cook and JM Whitmeyer. Two Approaches to Social Structure: Exchange Theory and Network Analysis. *Annual Reviews in Sociology*, 18(1):109–127, 1992.

J. N. Cummings, B. Butler, and R. Kraut. The quality of online social relationships. *Communications of the ACM*, 45:103–108, 2002.

C. S. de Souza and J. Preece. A framework for analyzing and understanding online communities. *Interacting with Computers*, 16:579–610, 2004.

E.L. Deci and R.M. Ryan. *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer, 1985.

C. Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49, 2003.

K.M. Eisenhardt. Agency theory: An assessment and review. *Academy of Management Review*, 14(1):57–74, 1989.

N. B. Ellison, C. Steinfield, and C. Lampe. The benefits of facebook friends: Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12:1143–1168, 2007.

B.H. Erickson. Good Networks and Good Jobs: The Value of Social Capital to Employers and Employees. *Social Capital: Theory and Research*, pages 127–158, 2001.

T. Finholt and L. Sproull. Electronic groups at work. *Organization Science*, 1:41–64, 1990.

M. W. Foley and B. Edwards. Is it time to disinvest in social capital? *Journal of Public Policy*, 19:141–173, 1999.

D. Friedman and A. Cassar. *Economics Lab: An Intensive Course in Experimental Economics*. Routledge, 2004.

D. Friedman and S. Sunder. *Experimental Methods: A Primer for Economists*. Cambridge University Press, 1994.

R. Gibbons. Incentives in Organizations. *Journal of Economic Perspectives*, 12:115–132, 1998.

D. Godes, D. Mayzlin, Y. Chen, S. Das, C. Dellarocas, B. Pfeiffer, B. Libai, S. Sen, M. Shi, and P. Verlegh. The firm's management of social interactions. *Marketing Letters*, 16: 415–428, 2005.

N. Gross and W. E. Martin. On group cohesiveness. *American Journal of Sociology*, 57: 546, 1952.

R. Gross and A. Acquisti. Information revelation and privacy in online social networks (the Facebook case). In *Proceedings of the 2005 ACM workshop on Privacy in the Electronic Society*, pages 71–80, 2005.

M. Gurven and J. Winking. Collective action in action: Pro-social behavior in and out of the laboratory. *American Anthropologist*, 110, 2008.

G. Hardin. The tragedy of the commons. *Science*, 162:1243–1248, 1968.

F. Hartwig and B.E. Dearing. *Exploratory Data Analysis*. Sage Publications, 1979.

FA Hayek. The use of knowledge in society. *The American economic review*, pages 519–530, 1945.

JD Jobson. *Applied Multivariate Data Analysis*. Springer-Verlag, 1992.

Q. Jones, G. Ravid, and S. Rafaeli. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research*, 15:194–210, 2004.

L. P. Robert Jr, A. R. Dennis, and M. K. Ahuja. Social capital and knowledge integration in digitally enabled teams. *Information Systems Research*, 19:314, 2008.

P. Kollock. *The economies of online cooperation*. Routledge, London UK, 1999.

P. Kollock. Social dilemmas: The anatomy of cooperation. *Annual Reviews in Sociology*, 24:183–214, 1998.

P. Kotler and F. Bliemel. Marketing Management. *Englewood Cliffs*, 2000.

C. Lampe and P. Resnick. Slash (dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 543–550. ACM New York, NY, USA, 2004.

J. Lave and E. Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, 1991.

A.S. Lee. A scientific methodology for MIS case studies. *MIS Quarterly*, 13(1):33–50, 1989.

A.S. Lee and R.L. Baskerville. Generalizing Generalizability in Information Systems Research. *Information Systems Research*, 14(3):221–243, 2003.

K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, X. Li, D. Cosley, D. Frankowski, L. Terveen, and A. M. Rashid. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10, 2005.

A. Majchrzak, A. Malhotra, and R. John. Perceived individual collaboration know-how development through it-enabled contextualization: Evidence from distributed teams. *Information Systems Research*, 16:9Ű27, 2005.

D. H. McKnight, V. Choudhury, and C. Kacmar. Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13: 334–359, 2003.

J. Mondak. Psychological approaches to social capital, special issue. *Political Psychology*, 19:433–439, 1998.

M. D. Myers. Qualitative research in information systems. *Management Information Systems Quarterly*, 21:241–242, 1997.

E. D. Mynatt, A. Adler, M. Ito, and V. L. O'Day. Design for network communities. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217, 1997.

J. Nahapiet and S. Ghoshal. Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23:242–266, 1998.

M. Olson. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, 1971.

T. O'Reilly. What is Web 2.0. *Design Patterns and Business Models for the Next Generation of Software*, 30:2005, 2005.

J. F. Padgett and C. K. Ansell. Robust action and the rise of the medici, 1400-1434. *American Journal of Sociology*, 98:1259, 1993.

A. Pinsonneault and K. L. Kraemer. The impact of information technology on middle managers. *Management Information Systems Quarterly*, 17:271–271, 1993.

J. Preece. *Online Communities: Designing Usability and Supporting Socialbilty*. John Wiley & Sons, Inc. New York, NY, USA, 2000.

J. Preece, B. Nonnecke, and D. Andrews. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20:201–223, 2004.

R. D. Putnam. Bowling alone: America's declining social capital. *Journal of Democracy*, 6:65–65, 1995.

A. Quan-Haase and B. Wellman. How does the internet affect social capital. *Social Capital and Information Technology*, page 113Ű135, 2004.

G. Ravid, S. Rafaeli, P. F. Marty, M. B. Twidale, D. Zeitlyn, F. Barone, A. Ciffolilli, and T. A. Maxwell. Asynchronous discussion groups as small world and scale free networks. *First Monday*, 9, 2004.

U.D. Reips. Standards for Internet-based experimenting. *Experimental Psychology*, 49 (4):243–256, 2002.

H. Rheingold. *The Virtual Community: Homesteading on the Electronic Frontier*. MIT Press, 2000.

P. C. Rigby, D. M. German, and M. A. Storey. Open source software peer review practices: a case study of the apache server. In *Proceedings of the 13th international conference on Software engineering*, pages 541–550. ACM New York, NY, USA, 2008.

P.A. Samuelson. The Pure Theory of Public Expenditure. *Readings in Social Welfare: Theory and Policy*, 2000.

N.J. Smelser. *Theory of Collective Behavior*. Routledge & K. Paul, 1962.

V. L. Smith. Microeconomic systems as an experimental science. *American Economic Review*, 72:923–955, 1982.

M. Snyder and N. Cantor. Understanding personality and social behavior: A functionalist strategy. *The handbook of social psychology*, 1:635Ű679, 1998.

R.M. Solow. But Verify. *The New Republic*, 11:36–39, 1995.

D. W. Straub. Validating instruments in mis research. *Management Information Systems Quarterly*, 13:147–169, 1989.

J.W. Tukey. Exploratory data analysis. *Addison-Wesley Series in Behavioral Science: Quantitative Methods, Reading, Mass.: Addison-Wesley, 1977*, 1977.

R.H. Turner and L.M. Killian. *Collective Behavior*. Prentice-Hall, 1987.

T. C. Turner, M. A. Smith, D. Fisher, and H. T. Welser. Picturing usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, 10, 2005.

A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

E. von Hippel. Horizontal innovation networks–by and for users. *Industrial and Corporate Change*, 2007.

J.B. Walther. Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research*, 23(1):3, 1996.

Y. Wang and D. R. Fesenmaier. Assessing motivation of contribution in online communities: An empirical investigation of an online travel community. *Electronic Markets*, 13:33–45, 2003.

M. Wasko and S Faraj. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *Management Information Systems Quarterly*, 29:35–57, 2005.

B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythornthwaite. Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Reviews in Sociology*, 22:213–238, 1996.

D. B. Whittle. *Cyberspace: The Human Dimension*. WH Freeman, 1997.

Models and Theories for Understanding and Motivating

Contributions in Online Communities

2.1 The social and economic cases for an online community

The previous chapter elaborated on the discussion of an online community at the level of sociability characteristics of its participants. However, we have not thus far provided much background theory related to the actual definition of an online community, and the social properties that an online community has. This is, in essence, the goal of this chapter. First, we explain what an online community is and then provide a discussion on the social properties and the known theory that is centered on the subject that this dissertation tackles. As aforementioned, online communities are of particular importance in organizational settings because they provide the ground for accumulation and dissemination of social capital [Blanchard and Horan, 1998]. It has also been argued that online communities resemble a virtual form of social cap-

ital in an organizational context, which, in turn, gives a strategic advantage for the firm that manages to exploit such a resource [Nahapiet and Ghoshal, 2005, Spender, 1996]. However, theory and research around virtual or online communities tackles this issue in a purpose oriented setting which can be independent of the production of social capital. This suggests that an online community can have several different forms, from an online review system to an Internet newsgroup, with its purpose and functionality being loosely coupled.

Undoubtedly, the definition related to the concept of what an online community is varies in the literature. One of the earliest theoretical papers on the social aspects of online communities by Wellman et al. [1996] elaborated on an online community as an extension of computer networks to those networks where social interactions occur (social networks). This is mainly due to the fact that social theory and economics consider communities as a structure with collective properties where externalities are produced [Shaffer, 1989]. Normally, communities rely on interaction and communication between their members. In a virtual setting, an online community can be defined as an asynchronous communication channel where individuals exchange information following a set of rules which are defined in the communication protocol [Rheingold, 2000].

The communication protocol dictates the way in which the communication endpoints interact. For example, in the case of an Internet newsgroup, the communication protocol provides that a participant posts something and then another participant replies. Usually this type of communication takes place in an asynchronous mode since there is a considerable time window between the time point that someone posts a question (t_{post}) and the time he/she receives a reply (t_{reply}). According to Koch and Wörndl [2001], a community mechanism facilitates the interaction between the individuals by providing the necessary tools for *(a) Identity management, (b) Organization of activities* and *(c) Facilitation of interaction*.

Identity management refers to the infrastructural capability that the community

mechanism offers to an individual in order to be able to manage his/her identity inside the community, as well as providing the capability for the community administrators to be able to manage the community memberships. For instance, Google Groups[1] provides a subscription process which someone must follow in order to register on a specific group. While some communities allow open participation in terms of monitoring of activities (e.g. reading the discussions in a newsgroup), the majority of them require registration in order to actively participate (e.g. to start posting questions).

Most often the participation takes place under an alias or a pseudonym by the community member. For privacy and anonymity reasons, some members prefer to hide their identity using pseudonyms - although in some communities the social properties such as trust are important and the use of pseudonyms might have a social cost at some point. The selection of the pseudonym has been proved that affects this kind of properties and especially trust [Friedman and Resnick, 2001]. For instance, in the case of online auctions (e.g., e-bay) where the primary interaction is transaction dependent, communication between the exchange parts is an important element of the transaction due to the fact that it is an enabler of trust between the sellers and the buyers.

Communities of transactions pose also an interesting role in these cases since much of the literature considers them to be an important factor for the function of electronic markets [Dellarocas, 2003]. Exchange networks can be characterized as a subclass of social networks where actors (whether buyers or sellers) form relational ties that represent exchange relations [Cook and Emerson, 1978]. One particular aspect of the network properties in this type of network is the reputation that an actor possesses in the network. Reputation is a characteristic property in offline settings as well, which can attribute to better network position and thus better opportunities for exchange in the future to those that possess a high degree of reputation in the network of their contacts [Raub and Weesie, 1990].

---

[1]http://groups.google.com

However, reputation is one side of the coin with the other side being trust. It can be argued that in this kind of relational quantification of social networks, trust imposes an important role in the formation of the relational ties. Actors form relational ties influenced by the trust they perceive from the other part of the dyad. In this case, social network metrics, such as centrality, provide a direct quantification of the reputation that the actor possesses in the overall network. For example, if we consider reputation as popularity, then a simple aggregation of the formed relational ties provides an indication of the reputation that the actor possesses in an exchange network (relational ties denote exchange relations). The actor's perceived sense of trust towards another actor is described as benevolence. Benevolence is an essential element in the study of social relations, especially in the case where there is the possibility of exchange networks which might involve trading or institutionalized commitments [Pavlou, 2002].

A portion of the research literature on online communities focuses on the exploitation of online communities for the enhancement of the firm's visibility as well as the improvement of exchange networks between the stakeholders of the firm [Williams and Cothrel, 2000].A particular emphasis is given in the transactions taking place on virtual settings since the Internet has transformed the traditional buyer-seller relationship, thus giving more options to consumers to interact with product vendors [Turban et al., 2000].

This has a significant effect on e-commerce activity through the establishment of exchange networks. Exchange networks of consumer-to-consumer transactions in particular (such as, for example, e-bay[2]), form a significant portion of e-commerce activity on the internet.Nonetheless, it is important for this activity to gain the same formal contractual commitments that can be established on offline settings as well. In this case there is a technological intervention based on community interactions between buyers and sellers which can be met in reputation mechanisms such as the one visible on e-bay. The importance of reputation mechanisms in environments where no

---

[2]http://www.ebay.com

69

contractual enforcement is imposed has been advocated by several researchers [Ba and Pavlou, 2002, Dellarocas, 2003] taking as a base the feedback mechanism of the online auction website eBay and the effect that it has on the enhancement of trust among buyers and sellers [Resnick and Zeckhauser, 2002].

Although relevant research has considered a series of psychological factors and, in particular, trust from the perspective of buyers or sellers, there has not been evidence of research on the analysis of dyadic relations and the network positioning effects in the trust exchange networks that are formed between buyers and sellers, and on the way they affect decisions to buy [Resnick et al., 2006]. Here online communities come to play a role since there is the establishment of a value-expressive and socially adjusting behavior that can also be seen in offline settings where formal establishments are present, such as a contractual enforcement mechanism [Klein and Leffler, 1981]. That is particularly important when the exploitation of the exchange network established in this setting becomes more and more important both for the buyer or the seller [Houser and Wooders, 2006].

However, transactions on formal (offline) establishments mostly consider the subject of exchange to be those goods where their perceived utility can be evaluated by the potential customer. However, on the Internet the type of goods that appear on Internet sites is not easy to be evaluated due to the fact that they are mostly experience goods [Ward and Lee, 2000].

Experience goods are characterized by a high degree of uncertainty [Nelson, 1970]. For instance, someone buys a book but he is not certain if he will like it. In this case, he will seek the advice of others who have an opinion about it. On the other hand, products and markets become much more complex than they used to be. Products and processes, such as Internet transactions, technology products, financial transactions, etc., involve a high level of complexity. Seeking advice from experienced customers/users that are influential in the final judgment is an important factor since the online community formed from the users of that good (e.g., in the case of chapter

6 the reviewers of a particular book) enacts the role of the facilitator of a transaction.
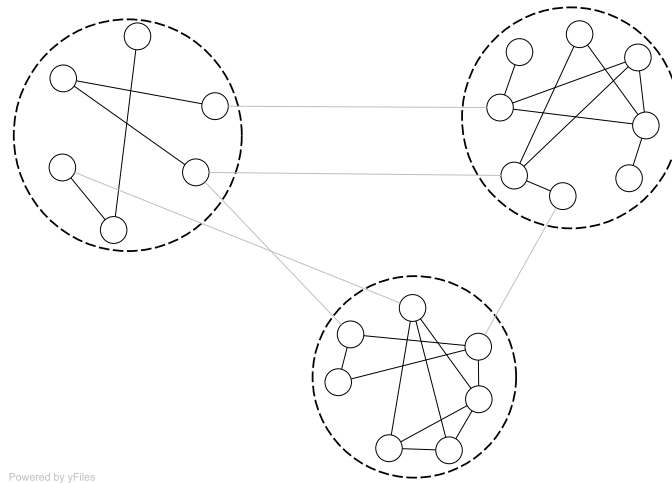
Figure 2.1: A sample representation of a community structure. Communities can be defined as groups of nodes such that there is a higher density of edges within groups that between them.

On the other hand, the part related to the organization of activities addresses the infrastructural side of the community. In particular, it provides the communication protocol between the members. For instance, in the case of an Internet newsgroup the protocol dictates that in order for someone to receive an answer he must first send a request. Online Communities also provide tools that facilitate interaction and make it more usable for the end-users/individuals, such as threading. For instance, a threading feature on an Internet newsgroup is useful because it organizes the information according to the topic or by the date. This part has also to do with the usability issues that someone may come up with when using the community mechanism. Usability issues are also a critical success factor for the development of a virtual community [Mynatt et al., 1997].

Although communication activity in an online community is group-based, most of the interactions that are facilitated are hierarchical or dyadic. As an example, we can revisit the case of the Internet newsgroups. A dyadic hierarchical way of interaction means that someone posts a request and several others reply to the individual and not to someone else, thus creating a discussion thread. Then the individual establishes a
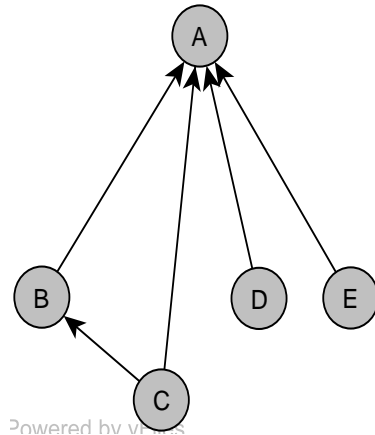
Figure 2.2: An interaction pattern representing a thread of an internet newsgroup. Nodes in the graph represent participants and the direction of the edges implies response to the message.

virtual interaction with each one of the answerers but there might be cases where answerers interact among them during the thread/discussion. For instance, an individual (A) is posting a question in a group and receives answers by four members (B, C, D, and E). As can be seen in the example pattern in Figure 2.2, individual A receives a direct answer by the individuals B, C, D, and then E and forms a dyadic relation with each one of them.

However, individual C also forms a dyadic relation with individual B. There might be the case where C is commenting on B's answer by adding his answer to the message posted by A. Dyadic relations formed among individuals on online communities provide a fruitful way of assessing activity by using established concepts and techniques from the domain of analytical sociology and, in particular, social network analysis. Depending on the nature of the dyadic relation, the activity level varies. Hagel and Armstrong [1997] classify online communities according to their activity level. Following their classification we can classify several types of online communities as follows:

- **Communities of Transactions** consider whether the interaction pattern and the relational ties formed in the community have to do with the exchange of a good or a service. Community mechanisms that enhance a transaction such as

72

feedback forums on e-bay also belong to that category.

- **Communities of Interest** are communities where interaction occurs over a subject and is spontaneous. These communities encompass Internet newsgroups and forums.

- **Communities of Fantasy** have recently evolved in the context of massive multiplayer online games where participants create a persona and maintain it over a context (e.g., a character in the popular online virtual world Second Life)

- **Communities of Relationships** classifies communities where individuals can explicitly create, dissolute and register their relational ties with other participants in those communities. More recently, discussion around this type of communities has evolved to the concept of online social networks In fact, Wellman et al. [1996] hypothesized the evolution of these types of communities which are one of the most popular web applications currently on the web (e.g. Facebook, Linkedin etc).

It is understandable that a classification in one of the above four categories is non excludable, that is, there can be communities that fall into two or more categories. For example, the very popular online social networking site, MySpace, facilitates a community of interest (around music and music production by independent label groups), while still allowing members to form connections (community of relationships).

What is evident from the above classification is that Lave and Wengers' approach considers an online community only in the case where some form of interaction between participants can be formed. For example, the visitors of a web site (either registered or not) do not form an online community unless they form an interaction pattern that can be facilitated by the web site's software (e.g., an online forum inside the website) in a direct (participant to participant) or in-direct mode (taking part in a discussion thread). This is also related to the membership policy that the website or the online community mechanism imposes on its members.

Regarding membership as a factor of online communities, similar to social settings, when members join an online community, their involvement passes several stages [Kim, 2000] usually by being observers of the activity. Lave and Wenger [1991] have studied the transition of stages of involvement on the context of an online community of students. Their study identifies *five* different types of membership trajectories which can be classified in the following clusters:

- **Peripheral:** In the peripheral trajectory a member is a short term visitor that seeks information to cover a specific need. Usually these members are the most bound to lurking since their involvement with the community is random.

- **Inbound:** Here a visitor engages with the community and becomes a regular member. Participation might be low since the frequency is periodical.

- **Insider:** An insider is a regular member of the community which participates very frequently in the activities of the community by posting frequently and helping other community members.

- **Boundary:** The insider becomes a leader of the community by becoming well known by the other members and gaining high reputation.

- **Outbound:** This is the final stage of the membership cycle when leaders who are dissatisfied with the community or having a change of interest drop out of the community.

Lave and Wenger's classification is of particular usefulness due to the fact that it connects the participation level of the community with the status that an individual member possesses. As will be argued later in this chapter, status is an important social incentive for the participation in an online community. Since status is gained through continuous investment in activity, members who gain the status of a frequent contributor on the community tend to invest more on maintaining their status levels.

Another research stream on the study of online communities is the research summarized by Watts and Strogatz [2006] where online communities are approached as structures whose collective dynamics have significant properties. This approach assumes the fact that interaction is not random but is driven by specific ordinal properties. Undoubtedly, there is a growing set of literature related with the discovery of community structures. A very early example is the work of Kleinberg [1999], where the structure of an online community can be classified in two two specific groupings, namely, the hubs and authorities. The objective of this approach is to model nodes on a hyperlinked environment, the hubs and authorities model can be of particular usefulness on understanding authoritativeness in terms of expertise in a networked environment. . The separation of authority is on the basis of hyperlinks or connections. A node is a hub when the number of links/connections departing from that particular node is much higher than the number of links/connections pointing to them. An authority is a node that has the opposite properties from a hub. For instance, if we consider the network of citations where a scientific paper receives the same popularity for two papers, a paper is considered to be an authority if the number of citations it receives is significantly higher than the number of citations that it provides.

A set of several papers present algorithms and methods for extracting community structure from interaction sources, mainly their level of analysis are hyperlinks between web pages [Girvan and Newman, 2002, Huberman and Hogg, 1995, Radicchi, 2004, Wu and Huberman, 2004]. Although these models are of particular interest from an analytical viewpoint, they do not include a rigorous evaluation of behavioral properties derived from the community members' personal characteristics. Furthermore, these models extend the concept of nodes from the information resource to the producer, and therefore are somewhat blind to the social properties of the individuals involved in multiple levels of interactions [Korfiatis, 2007].

This chapter aims to present related theories and models that have emerged in the study of online communities from a behavioral perspective. A particular emphasis is

given to the theories that are used in the studies discussed on the empirical part of this dissertation since the conclusions expand on the results obtained by these empirical studies in connection to theory.

These theories tend to explain (a) the formation of relational ties on online communities (b) the group sense of the participants on these communities and (c) the expansion of the life cycle of the members' activity as presented above. Section 2.3 provides background on the theory of social preferences, linked with the findings from Chapter 3 which is related with social preferences.We study their expression of behavior in particular cases such as inequity aversion, altruism and reciprocity. Section 2.4 presents a series of models that are well known in the literature in relation to understanding the contributions on online communities and how these communities evolve with time. Section 2.4 takes this one step further to provide a review of recent models for enhancing contributions on online communities. The chapter concludes with an outlook to the empirical part in relation to the theories and models presented.

2.2 Structural approaches

2.2.1 Social network analysis and online communities

As aforementioned online communities facilitate social interactions that regardless the activity context, require a concrete methodological formalization. This is due to the fact that social interactions can be multiplex (e.g. communication, exchange etc) and such formalization may allow for inter-context studies of the behavior of the involved members [Uzzi, 1999]. One of the most suitable methodologies for observing social activity and modeling the interaction of individuals in a social group is Social Network Analysis (SNA). SNA has been established as a concrete methodology resulting from social psychology and communication studies [Wasserman and Faust, 1994]. The growing amount of SNA research done in organizational context [Borgatti and Foster, 2003] supports a broad positivistic view on the study of technology oriented phenom-

ena by using the formalism provided by SNA and Structural Analysis in general [Zack, 2000].
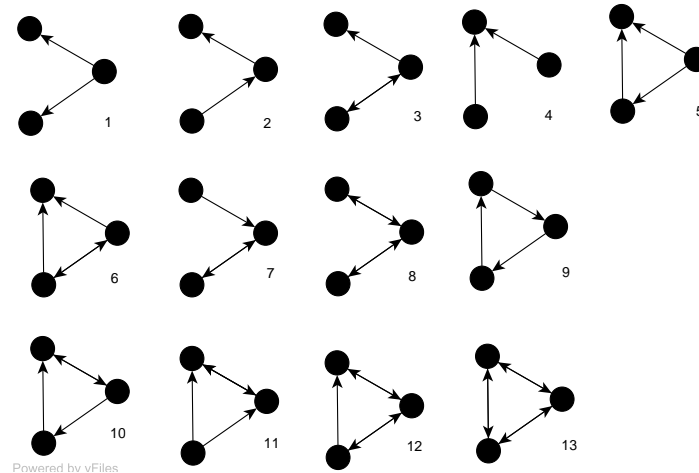
Figure 2.3: The set of different motifs formed in a triadic relation

One of the basic assumptions of social network analysis and in general the part of sociology that relates with structural analysis is that the topology of an individual (actor) on his/her network of contacts/relations has a profound effect to his/her behavior [Scott, 2000, Wasserman and Faust, 1994]. Actors with a better position (at the core) in their network are likely to get access to more resources than actors that stand in the periphery. In order to model social interactions in an online community as a social network, we first need to classify the two major types of network variables namely the structural and the compositional variables. Table 2.1 provides an example operationalization of these variables. Structural variables form the core of the network and contain a dyadic record of social interactions between two actors that belong on the same network. By drawing the set of the structural variables we have the complete network. Nonetheless structural variables describe only the relations and not the individual characteristics of the actor. This role is undertaken by the compositional variables which provide a way for expressing actor related attributes such as demographics etc.

According to Carrington et al. [2005] when conducting Social network analysis and

Focus	**Operationalization**	**Variables**
Focus on the Group	Sets of triads	Compositional Variables
Focus on the Individual	Dyads	Structural Variables

Table 2.1: Operationalization of Structural and Compositional Variables in Social Network Analysis

relational analysis studies in general considers two major assumptions when presenting results:

- That the relationships of the individuals studied, correspond accurately to the real context. Following this direction subconscious or illicit relationships are not represented or either not included as a subject of the sociometric study.

- Group size is the optimal on the sense that the boundary is the optimal and includes those actors that influence directly at least another member of the network or the social group that is examined.

Apart from the variable definition, In order to construct a social network, one has to define the unit of observation from which the structural variables of the social network will be constructed. Due to the fact that SNA tackles with the topological properties of the unit of study we have the flexibility to follow the same set of methods regardless the size or the nature of the unit. For instance we can seek for interaction patterns with the same analytical methods both for individuals and institutions. In relation with the unit of observation, Table 2.1 summarizes the relational quantification of the social network which defines the way the relational ties in the structural variables of the network are formed. Depending on the nature of the network a relational tie can be either directed (directed network) or reciprocal (undirected or symmetric network). The relational quantification is also subjective to the research question that is pursued. For instance if we are interested to examine information flow among individuals in online communities then the obvious way is to model it with as a directed network. However if the research question is broader for instance in the case of membership
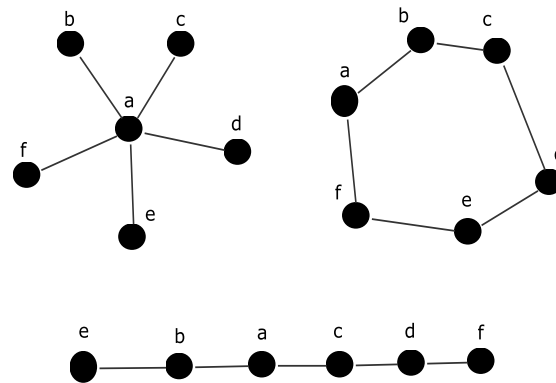
Figure 2.4: Types of connection degree in the network

then an un-directed network is more suitable. Relational quantification also depends on the network data available.

Another approach on the analysis of social networks is the analysis of status [Raub and Weesie, 1990]. Usually in a sociological interpretation the concept of status denotes social power expressed in different contexts such as political or economical. Depending on the nature of the structural variables in the network status or power denotes the case of social support that the compositional variables of the network receive. The most basic theoretical implication of status is the availability of choices the entity receives in the network which gives the entity the advantage of negotiation over the others.

For example let us consider the three different network topologies that are represented in Figure 2.4. It is obvious that the sociometric star (first diagram) is considered the one who has the highest degree (the number of lines adjacent to this particular actor connecting him/her with the other members of the group) thus is more popular. When there is the case of reciprocity in the social connections the node's degree of influence can be used to provide an indication of prestige in the network.

Quantifications of status employ techniques of graph theory such as the centrality index [Freeman, 1979] which have been adjusted to the various representations of ties in a network. In this dissertation the approach of social network analysis techniques

that we employ for the analysis in chapters 4 and 5 is limited to the modeling of dyadic ties and the portion that these ties represent to the overall network connectivity as depicted by the variation measures employed in Chapter 4.

2.2.2 The weak ties hypothesis

Central to Social Network theory is the weak ties hypothesis that was first formalized by Rapoport [1963] and later theorized and discussed by Granovetter [1973] in the context of a job search network in a labor market [Granovetter, 1985, 1978, 1983]. Essential to the weak ties hypothesis is the idea that there might be cases that the structural variables considered do not actually provide a realistic view of the network structure. In particular, there might be contexts where the abstract formalism doesn't consider different forms of strength regarding the relational ties which are formed between members of the network. In fact this formalization doesn't distinguish between the strong and weakest forms of relational ties. Based on the particular strength or weakness of the relational ties there can be assumed cases of influence/obedience or isolation of the individual from the group and this effect to the social interactions that are already formed.
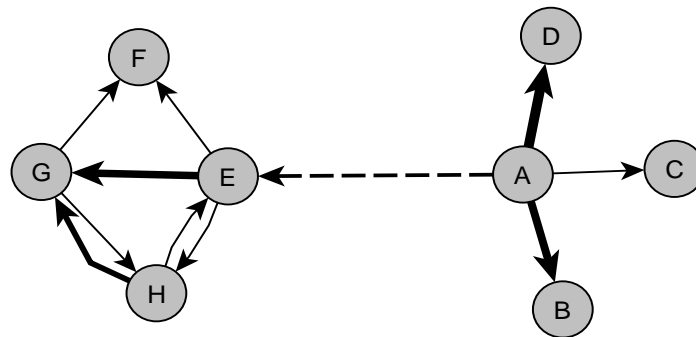
Figure 2.5: The Weak Ties Hypothesis. According to the hypothesis, weak ties (dotted) act as bridges between sets of strong dyadic relations

Nonetheless defining a strong or a weak relation in an online setting is a matter of great complexity. Several sources of interactions could be used to define weakness or

strength of a tie in a dyadic structure. Granovetter (1973, page:1361) in his original formalization argues that "*the strength of a tie is a supposedly linear combination of the amount of time, the emotional intensity, intimacy and the reciprocal services which characterize the tie.*" For example in the case of an email network we could consider the number of emails those two persons have exchanged as an indication of social activity between them, thus being positive correlated with the strength of their social connection as a longitudinal effect. A similar case can be for example in the case of chapter 5 where interpersonal communication is direct (one to one) and repeated interaction might lead to a stronger tie.

According to Granoveter's perspective, weak ties may act as bridges over time, between strong dyadic relations thus positioning an important role in the network structures by bridging strong dyadic neighborhoods of the network. In that case group cohesion (or group efficacy) might be a factor to consider. Weak ties affect the cohesion of the group which in turn affects the activity. It is more likely that a group with strong cohesion will be more active and energetic than a group with a weak one. This is because a weak tie is more likely to expose it's weakness in an isolation from the other entities of the structure. However it often reestablishes itself with another destination. The possibility of this to happen depends on context dependent properties such as affiliation with a third party or a common activity. In that particular case there is the evidence of another structural property that is related to the transfer of network properties from one actor to another.

According to Freeman et al. [1992], transitivity is a basic characteristic of a network structure which describes how resources are transferred from the one part of the dyad to the other. Depending on the setting where the network is studied *transitivity* may be operationalized with the amount of information transferred from the one actor to the other (e.g. in the case of an e-mail network) or the number of goods the one buyer has bought from the other (e.g. in the case of exchange networks). The case of transitive properties of a network becomes interesting to consider when we observe

transitivity among different levels of status or hierarchies (in an organizational context). In research context transitivity is treated as part of the unit of analysis (in that case the dyad). In a more formalized view a transitive relation considers a pair of two connected dyads and their influence to the establishment of a triad (set of three dyads) due to their transitive property. Let us consider a network of three actors *(A, B, C)* with the dyads ($A \longrightarrow B$ and $B \longrightarrow C$) having a transitive property. In that case the chain that is formed between A, B and B,C forces a third dyad to be established between C and A thus resulting in a complete triad or cycle among these actors. For example if A has a good relation with B and this property is transitive on the relation between B and C then the probability that B and C will establish a contact is high. The later is referred to the literature as the triadic closure bias [Watts and Strogatz, 2006].

One particular approach on social network analysis when it comes to the individual level rather than the group is the use of egocentric networks [Marsden, 2002]. Egocentric networks represent an interesting focal point in social network analysis and especially in the analysis of communication patterns [Fisher, 2005] since they provide a direct focal point to the addressee of interaction, without demanding an analysis of the whole structure to induce facts about the structural activity (Sociocentric Approach).Following that approach in an online community and considering the relational ties as the interaction in terms of posting and answers the egocentric network of a user in the online community is modeled after the information flow between a user that has asked a question and the user that has posted an answer to this question. In this dissertation we use the egocentric network of the user in order to model the variables used in the empirical studies presented in the chapters 4 and 5 where we study interactions in a dyadic level.

2.2.3 Collective behavior and social loafing

As aforementioned in the previous chapter, collective behavior is the case where the members of collective (e.g. a group) behave in a uniform manner [Ajzen and Fishbein,

1980]. One of the earliest theoretical studies of collective behavior in an offline setting was presented early in the social psychology literature by LeBon [1897]. According to LeBon's theory in a collective individuals imply to homogeneous behavior through basically three factors: (a) Anonymity (b) Contagion and (c) Suggestability. Although Lebon's model intents to describe the production of antisocial or crowd behavior there are some useful observations that we can use from our own viewpoint on the context of an online community.

The first derives from the case of anonymity that an online setting can assure. Although anonymity is against the factors related with signaling and reputation as we are going to discuss further, it's a decisive factor for individuals behaving in a uniform matter due to the fact that they are not accountable in an offline setting for their action. The contagion contributes to the rapid spreading of these ideas and the way shifts are being made. A case of this example can be the setting of online forums where under specific conditions individuals do not contribute to the public good but direct their efforts to irrelevant activities which are underlined by subconscious social objective (suggestability).

Social Loafing is a central theme on the study of performance of group members in accordance with their motivational incentive. Karau and Williams [2001] define social loafing as the characteristic reduction in motivation and individual effort that takes place where individuals work together with other group members on a collective task. According to theory, very often this individual effort in a group setting is less than in an individual setting. Social loafing is highly dependent on the importance of the group task that the collective has to address. In such cases where individual's effort has a direct effect to the other group members the motivation to contribute becomes high so does the effort that the individual provides during the process of accomplishing the collective task.

2.2.4 Reputation effects

Reputation as a social incentive relates with the status of an individual in the community and the perceived importance that he/she processes by the activity inside the community. At that case an individual would want to sustain his/her reputation and thus behave accordingly. We highlight two types of reputation: group perceived reputation and individual status. Group perceived reputation relates with the reputation/status that the individual possess inside the group as a whole. Individual status relates with the reputation in the individuals ego centric network. That is the network of first and second degree acquaintances in the part of the group that he is active. For instance someone may be active in some topics in the newsgroup and thus reputable on these however he/she might not possess the same status in the whole. Considering reciprocal behavior we can hypothesize that an individual will answer a question in order to increase or maintain his/her status in the community while on the contrary he will answer a question if he considers that it will increase his/her visibility in the network.

The later is very much connected with the concept of signaling first introduced by Spence [1973] on the context of a labor market. Signaling is a way to exchange meaningful information among two parties by having the interested party (agent) to communicate information about itself to the other party (principal). On the context of an online community signaling is taking place between two parties where an agent might be a peripheral member (random visitor) and the principal might possess a higher status. However signaling assumes the establishment of some form of direct contact between the principal and the agent which is not always possible. That's due to the fact that the structure of interactions on most of the types of online communities (e.g. newsgroups) is not maintainable. Since everyone can participate there is no barrier on removing offtopic discussions which as in offline social interaction settings can distract the signaling parties. Another issue with signaling when it comes on online communities is the case of identity [Donath, 2007] where the communication

of information among the two parties might not be feasible due to the high level of cognitive effort required to infer and evaluate this information.

2.3   Social dilemmas and knowledge sharing on online communities

According to Kollock [1998] social dilemmas are triggered in situations where individual rationality directs to collective irrationality. This provides that in a case of a group setting, reasonable behavior directs to a situation that all the members of the group are functioning worse that might have been if they were working alone. If we model social dilemmas from an equilibrium based viewpoint, a social dilemma can be classified as a situation where there exists at least one deficient equilibrium. That provides that in this equilibrium setting an individual is performing worst than it can otherwise. This type of social dilemmas is called N-Person dilemmas[Komorita, 1976]. There are two subclasses of social dilemmas that are of particular interest in the context of understanding an individuals' contribution to an online community.

The first class which is known in the literature as **social trap** [Platt, 1973, Rothstein, 2005] is where the individual is tempted with an immediate benefit that produces a cost shared by all. Such an example is the case of an individual posting a question in an online community, several other community members invest time and effort on trying to find an answer to this question, then the individual comes back and signals to the other members that he/she has found an answer to his/her inquiry but is not sharing the output with the others. The later imposes a cost to those that participated on helping this member to find an answer in terms of opportunity cost and lost benefit (which can be due to individual self-interest in that particular question).

Social dilemmas have an output on externalities which in the concept of an online community is the situation where *"a members' behavior affects the situation of other persons without the explicit agreement of that person or persons"* [Olson, 1971]. The

later incorporates the case of a public good dilemma. In the public good dilemma the key characteristic is that the provided good is non-excludable which can lead to a person that is in a public good dilemma situation, the temptation to consume it without contributing to its creation or maintenance. On that case a key characteristic is the relation between the amount of resources contributed to the production of the public good and the level of the public good provided. An interesting example of the public good dilemma can be found in the case of open source software. An open source software can be characterized as a pure public good since it possesses the nature of being non excludable (everyone can download it and run it, while on the same time they get access to the source code which is key for extending it) while its use is not excludable [Cornes and Sandler, 1996]. Peripheral users will tend to develop more the software in order to be able to gain benefit from it without giving back their modifications back to the community. However if all users' behave like this then it is sure that the software's development will stop and the outcome will be costly for those that use the software to perform necessary tasks. However if we consider the amount of effort and resources required by an open source developer to contribute back to the software project it is obvious that the best strategy on the short run is the deficient equilibrium of non-contributing at all and making full use of the software for the satisfaction of self interest.

However this deficient equilibrium is not evident in most successful online communities. This is due to the ability of the community to provide an increased sense of group identity and self efficacy level for the individual. In that case the individual engages in an imposed precondition once he joins the community where the expected behavior is to comply with the social norm of the group. If the signaling of the individual's reputation to the other members of the group is significant, then the individual will increase his/her commitment to the groups purpose. For example the visibility someone gains in the open source community and the software industry for being a developer of the Linux Kernel is a decisive factor for continuing contributing to this

public good [Moon et al., 2002].

Intervention	Objective
Encourage Communication	Increase Sense of group identity Increase commitment Increase frequency of Interactions Increase identifiability Increase expectations of others' participations
Create Knowledge Sharing Communities or Communities of Practice	Increase Sense of group identity Increase frequency of interactions Increase identifiability
Publicize information about member's contributions	Increase identifiability

Table 2.2: Intervention and objectives related with the social dilemma hypothesis. Adapted from Cabrera and Cabrera [2002]

Cabrera and Cabrera [2002] formalize the case of social dilemmas in online communities considers three levels of intervention (a) On the level of communication (b) On the level of knowledge contribution and (c) on the level of personal incentives. Regarding the level of communication the objective is to increase the sense of group identity. In fact by re-approaching the case of structural capital and collective effort the increase sense of group identity can have a significant effect on the way activity is expressed since the individual have an increased notion of the collective. An aftermath of this direction is an increase to the commitment of the members in order to become more visible in the group. Increased commitment leads also to increase of interactions among the group members resulting to the creation of knowledge flows in the community that permit knowledge sharing.

In order to better enhance knowledge sharing a set of signaling parameters is

needed in order to maintain the knowledge flows constant and increase them. An easy way to provide that is by providing mechanisms that can make the identification of the contributors and their activity easier for the other members of the group. For example a profile page where the activity and commitment of the online community member will be easier to track for other members which in turn has also an effect on the performance of the other group members. Increasing identifiability makes the level of commitment higher to the group in cases where the mean group output is higher than the standard output provided by the individual. Apart from personal profile pages the community mechanism can also publicize information about the members' contribution by using a standard and understandable ranking mechanism (e.g. by aggregating bonus points each time a community member is making a contribution, controlling for the commitment and the longitudinal setting of his/her activity).

2.3.1 The argument of social preferences and the public goods dilemma

As aforementioned Social dilemmas reflect situations where individual rationality leads to cases of irrationality in a collective level. This implies that in the case of a group an individual's reasonable behavior leads to a situation in which all the members, by pursuing their own self interest, lead to the worst possible result. A classic example of such a social dilemma is the public goods game that is studied on Chapter 3 of this dissertation The classical game theoretic setting of this approach considers a group of four individuals which are requested to contribute some of the resources that the possess to a common pool which in return will provide a common utility setting measured by the marginal per capita return (MPCR). For example let us consider the case of a university project which gives bonus points to the students and which requires effort (in hours) to be contributed for its realization. If the project gets a good grade (as a result of the contribution of many hours contributed by its individual) then every individual gets the same grade added as a bonus to the exam grade. The dilemma in this case for each group member is that given that the limitation of the time devoted

for studying, should the individual contribute a lot of his/her time for the academic
project or to devote all or some of its time for its own study in order to perform good
at the exam ?

The standard rational approach to this case is that the individual will use all of
his hours (available for study) to prepare himself for the exam and contribute the
minimal (zero hours) for the group project either by self interest (achieving the highest
grade) or by a tension to free ride (get a pass grade with a minimal effort). However
some of his/her other colleagues might contribute more than the time devoted by this
individual since they want to have a good grade or even secure that they will pass in
case of exam failure. This might lead to a tension inside the group where individuals
who contributed a lot of effort in the group project and didn't do well in the exam
might envy those that contributed the minimal or no effort and received a good grade
(because they devoted more time to study) while on the other hand those that got
a good grade overall and contributed minimal effort in the group project might feel
guilt because of their past behavior. In that case the aforementioned social dilemma
leads to cases where social preferences emerge among group participants that is that
in retrospect their behavior might not solely influenced by material self interest but
affected positively or negatively by the level of contribution of the others.

2.3.2 Social preferences and aversion to inequity

Inequity or inequality aversion addresses the case where individuals resist to inequitable
outcomes and their constant behavior is controlled by what other members of the col-
lective receive [Fehr and Schmidt, 1999]. In fact inequity aversion contradicts the
neoclassical self-interest hypothesis where everyone's actions express their material
self interest as the dominant behavioral pattern. In that case an individual might be
willing to give up some of its material payoff (e.g. a top grade in the previous exam-
ple) in order to get more equitable outcomes for the group participants. In order for
this to happen the general axiom in that case of situation is that in a given social con-

text individuals do care about their own position in relation with the others by some mechanism of social comparison

- *The disadvantageous inequity aversion or envy*: Where individuals dislike the case where other individuals receive more compensation than them.

- *The advantageous inequity aversion or overcompensation*: Where individuals dislike the situation where they feel that they receive more than the other individuals.

Inequity aversion can be observed in any setting that involves exploitation over a common good or a social dilemma related with self and group interest. For example, in the case of an online community an individual might feel envy where another individual receives more relevant answers and in faster time than him/her. While on the other hand an individual might also feel overcompensated if the quality of the service is more than he/she expects to have. Inequity aversion is a useful model to explain whether the behavior of the individual is affected by the general group behavior and thus explain attitudes towards the other members of the groups based on group outcomes. Following the model of the inequity aversion we can reformulate the case of inequity aversion in two examples related with the behavior of a member in an online community. A member is willing to answer a question due to the fact that the support that he/she has received from the community has resulted to an overcompensation effect. On the contrary members are less likely to respond to an answer if they haven't received a response to theirs. In that case we have an exploitation of a disadvantageous inequity aversion where the *envyness* results from the self efficacy of the members' condition. Due to the high cognitive processing involved for the members on realizing their status in the community the Inequality aversion has not immediate effects but evolves over time.

However for inequity aversion to appear there is the precondition of previous activity evident in the community in order to allow for social comparison to take place.

Previous activity is a longitudinal variable that in the case of a community refers to the perceived group activity by the individual at a specific point of time. For some individuals high communication activity is also a factor of lurking due effects attributed to tragedy of the commons [Hardin, 1968]. In that case if an individual participates in a group with a high amount of members then he/she perceives that fact that there will be the belief that there will be other community participants that they will provide a reply to a message posted by someone. Previous activity is likely to influence the contribution of a member in the community due to the group mediation factor. The higher is the communication activity of the group at the time the individual wants to participate then the less is his/her willingness to do so. On the other hand the higher is the communication activity of the group at the time then the individual participates to the threads that are more active.

Altruistic behavior relates with the case where members do actions without expecting any particular reward. This type of behavior is an important issue on online communities since it enhances the amount of social capital contributions to the collective. For example in a community of programmers if someone asks for a solution to a software bug that he is facing then another participant might devote his/her time to run and debug the program and post the answer out of altruistic concern However there might be a case where individuals tend to be reciprocal altruists in order to boost their self interest in a posterior case [Trivers, 1971]. That is in the case that this individual programmer that was benefited in this example will do the same for someone else when the time comes. This is essentially attributed to his/her sense of fairness. Fairness and reciprocity essentially constitute the companion of altruistic behavior. We discuss those concepts on the section that follows.

2.3.3 Fairness and reciprocity

From a sociological perspective, reciprocity is a fundamental structural concept that relates dyadic relations with the behavioral outcomes. It is both a structural and

behavioral topic that is often addressed with offline interactions [Swedberg, 2004]. As can be also observed in offline settings of group or individual behavior, individuals participating in an online community tend to reciprocate the behavior or service they received from another individual during their participation in the online community as a form of reciprocating the service [Fehr et al., 2003]. By receiving a conceivable good service by another individual, an individual feels a debt to that and has a tendency to reciprocate. This is mostly influenced by offline cultural settings [Miller and Bersoff, 1994] and the interpersonal communication factors characteristics that may affect it such as for example the case of the personal contact. In fact personal contact is an influential factor on reciprocation due to its imposition of reciprocity as a socially imposed behavior [Gouldner, 1960].

Reciprocity is an interesting phenomenon from a social perspective in an offline setting but how much different is the study of reciprocity in a virtual setting? What makes the case interesting to study reciprocation on these settings is the degree of anonymity that is provided under pseudonyms. Pseudonyms in fact represent informal contracts whose violation doesn't lead to any consequences for the violating parties since even unacceptable behavior in the eyes of the others doesn't lead to real social retaliations. Bad behavior in online settings is unlikely to lead to a reciprocal bad behavior by the others in an offline social mode. Drawing from that we can hypothesize that an online community member is more likely to answer to a question of another individual if that individual has provided an answer before. On the other hand this particular community member is more likely to not participate if he hasn't had an interaction with that individual before.

2.4 Models for understanding and enhancing activity in online communities

So far we have reviewed models that essentially provide a framework on understanding the activity in the context of an online community. However what matters on an online community is the effort and the output that its members contribute. In this section we are going to review models that can enhance the contributions to an online community. While undoubtedly a particular way of enhancing contributions to online communities could be the case of monetary rewards, the distributed nature and the volume of membership would have made such a solution costly and in fact would have turned the collective into a market where interaction is strictly based on monetary terms and price premiums. Therefore the models that we are examining in this section of the chapter are based on socio-psychological factors.

Koh et al. [2007] propose the model presented in Figure 2.6 as an overview of the factors that stimulate structure which is one of the most important characteristics that affects activity on an online community. The model itself relies in two basic pillars. The factors that drive the online community which have to do with (a) The leader's involvement (b) The level of Offline interaction and (c) the apparent usefulness that the social capital created by the community has to its members. Leaders' involvement and the level of offline interaction affects posting and viewing activity while the usefulness of the social capital affects the viewing activity and in return the attraction of the new members.

To the above there is the controlling nature of the community size which is in direct connection with the critical mass required for the community to operate efficiently. The authors argue that with the improvement of the quality of the IT infrastructure the effect of the virtual community drivers will become more apparent. Although this approach is not a new idea it emphasized the effect that technology has on social activity. Rheingold [2002] takes that case one step further and approaches this case
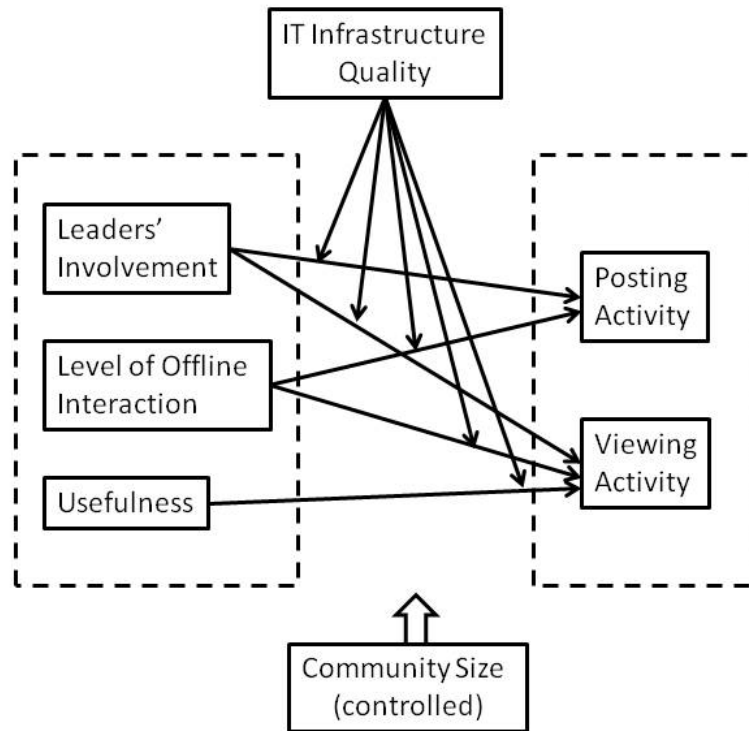
Figure 2.6: Virtual community stimulation structure. Addopted from Koh et al. [2007]

on the potential of mobile communities as new ways of organization where ad-hawk interaction can be facilitated using mobile devices.

In the following sections we summarize models that tackle with the factors that affect contribution on online communities and their effects on community activity and sustainability.

2.4.1 The collective effort model

The collective effort model is a meta-analytic model[3] that highlights the level of connection between individual motivation and group outcomes. The integrative nature of this model makes it more appropriate to use when addressing a social loafing condition than other models which are based mainly on the social impact of the effort, the evaluation potential of the group and the dispensability of the effort. The advantage

---

[3]A model that is build upon the results of several earlier studies

of the Collective effort model is that combine the motivational principles that derive

from *expectancy theory of work motivation* [Pinder, 1992] together with the principles

of *self-evaluation theory* [Ryan and Deci, 2000]. From that perspective the collective

effort model is integrative on the sense that it combines those two theories to address

the case of social loafing.

In particular the original principle of the model is that individual motivation and

effort in collective tasks will stay unaffected as long as a set of contingencies are

satisfied.

Figure 2.7: The collective effort model. Adapted from Karau and Williams [2001]

The collective effort model builds on principles from expectancy theory [Goldman

et al., 1987]. According to these principles, an individual needs to feel that his/her

effort will lead to an individual level of performance. For example in an online com-

munity a participant might start posting and participating in the community activities

if he/she believes that he will gain a higher status or visibility in the community.

An interesting application of the collective effort model has been undertaken by

Ling et al. [2005] where the contributions in terms of individual reviews and ratings

were used by the MovieLens recommender system [Miller et al., 2003] in order to

provide recommendations to other members of the community. In this case It is more

likely that the individual effort will be dropped if with relation with the group outcome is

not significant to the overall group outcome. Unavoidably the collective effort model is a clear case of a social comparison process. It is expected that a similar case of social preferences where the mediated group factor will affect the persuasion of self interests in relation with the group interests [Bolton and Ockenfels, 2000].

2.4.2 The Model of Whittaker et al. [1998].

One of the well known models which address activity in online communities (in that case internet newsgroups) is the one introduced by Whittaker et al. [1998]. Their work introduced a structural model which aimed to provide an overall understanding of newsgroup activity as a function of several properties related both with the newsgroup activities as well as the context and the purpose of the newsgroup discussion. The main latent variable in that model is the *Thread Depth* which characterizes how active the discussion is inside the newsgroup. As can be seen in Figure 8 a discussion thread is formed by several replies on a news item by the other members of the newsgroup. However during the thread discussion members do not only respond to the original poster of the news item but also they start discussing between them as well. The thread depth denotes how many sub-discussions have been formed under the original thread. It can be argued that high thread depth leads to high cognitive effort required by the members to follow the discussion and participate.

A certain amount of criticism can be attributed to this model from our perspective mainly due to the selection of the "Thread Depth" as an expression of activity in a discussion group. Following in fact the original proposition of that study we fail to consider behavioral characteristics apart from the familiarity, which in fact might influence the behavior of the participants in the newsgroup (e.g. provocative content).

Moderation is an important element in online communities since it has to do with the way participants behave and therefore express their activity through postings. This has an effect on the number of posters, the standard posts (FAQs), Cross postings (targeted messages) as well as the message length. Familiarity on the other hand
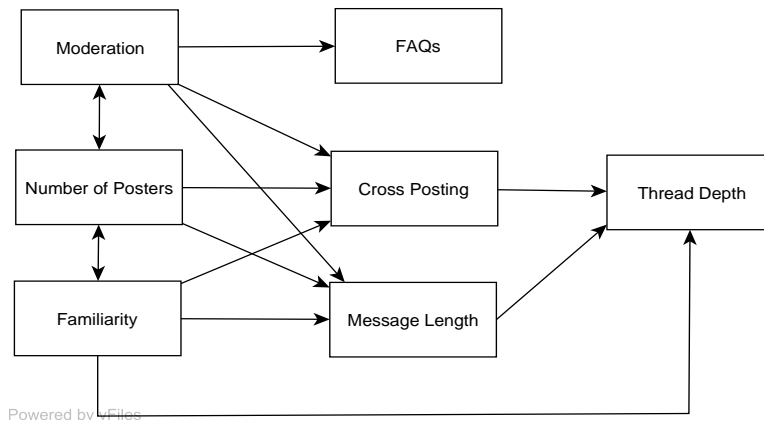
Figure 2.8: A causal model of communication in a newsgroup. Adapted from Whittaker et al. [1998]

considers the alternative case where a positive effect on familiarity will have a positive effect on cross posting and message length essentially due to the fact that familiarity is an interpersonal factor that affects communication characteristics as well [Krauss and Fussell, 1996].

2.4.3 The Model of Jones Ravid and Rafaeli

Jones et al. [2004] adopt a different approach on explaining contributions on online communities by evaluating the level of cognitive effort required by an individual member to participate and interact in an online community. The authors approach the output of the community as a public good where variable levels of cognitive effort is needed by the individuals to invest to (a) Infer and understand the messages posted in the particular group category that they are interested to participate (b) Formulate and post back an answer.

The model is interesting to consider from the fact that it examines community interaction with a less set of rules imposed such as in the case of online newsgroups. Information overload by high activity can lead to a case of lurking [Preece et al., 2004] where individuals simply cannot follow the flow of the activity in the community.

97
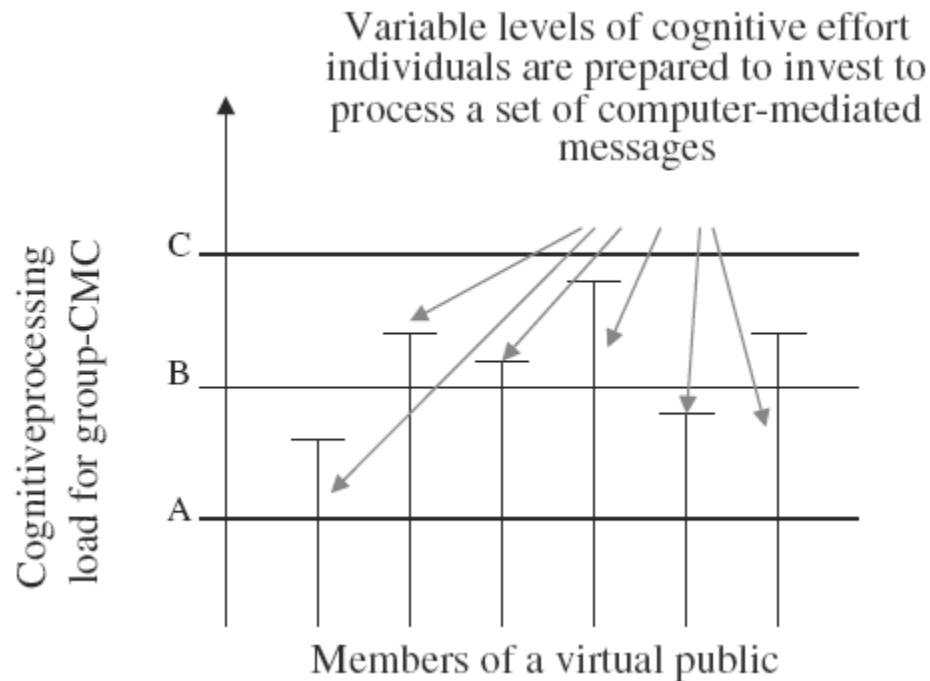
Figure 2.9: Information overload and cognitive effort for the members of an online community

2.4.4   The Model of Butler

The model of Butler [2001] adopts the perspective of the virtual community as a benefit creator for its members or as an aggregator of a benefit provision process that is provided through participation. As can be seen in Figure 10 the model contains three pillars that characterize the activity in a virtual community (it considers a Usenet group as an exemplar case) where individuals can become members or unsubscribe from the group at any time.

Butler's model is composed mainly of three factors: (a) Member attraction and retention (b) Resource Availability and (c) the benefit creation process facilitated by the community.

Member Attraction and Retention encapsulates the community's ability to attract and retain new members in order to keep the level of their interaction constant as well as to increase the size of the members. This in turn has an effect to the resource avail-

ability which can be defined as the volume of social capital available at the specific
time point on the community. Increased resource availability leads to more interaction
between the members in terms of a continuous communication activity, which defines
a benefit creation process.
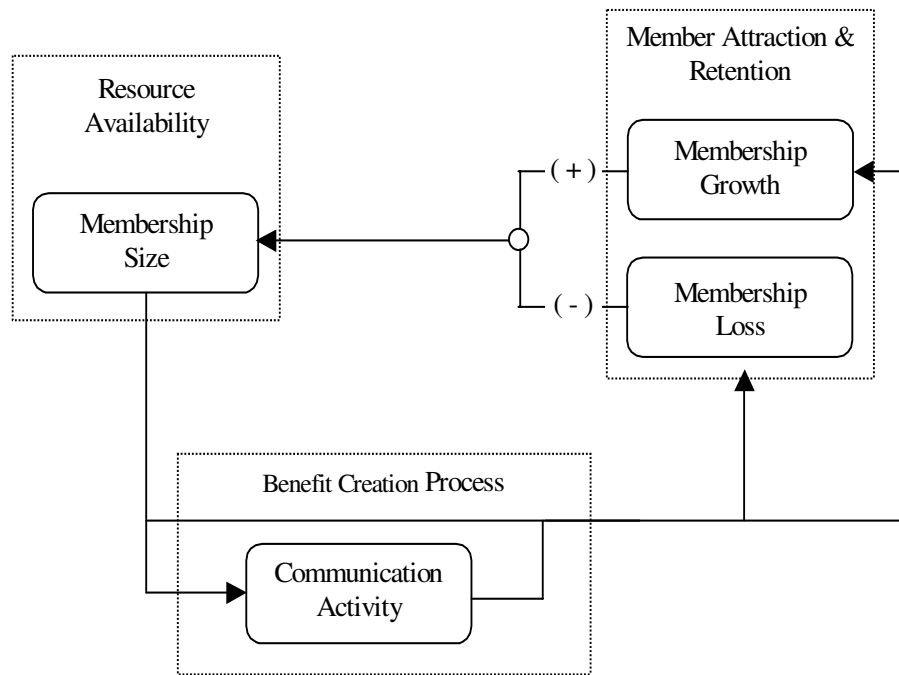
Figure 2.10: Membership Size, Communication Activity and Sustainability in an online Community. Adapted from Butler [2001]

2.4.5 The Model of Wasko and Faraj

Wasko and Faraj [2005] present a social capital based approach on evaluating contributions on an online knowledge community of law practitioners. As can be seen in
Figure 2.11 the model is structured in four factor groupings namely (a) Individual Motivations (b) Structural Capital (c) Cognitive capital and (d) Relational Capital. Individual
motivations consist of the notion of reputation by the community members as well as
a general notion of altruism embedded in each community members' social preferences. Structural capital evaluates the position of the member in the group which if
we consider also the connection with the Karau-Williams model Karau and Williams

[2001] which is a case of group mediation factor, it is expected that it will have some effect on the contribution of the individual members.
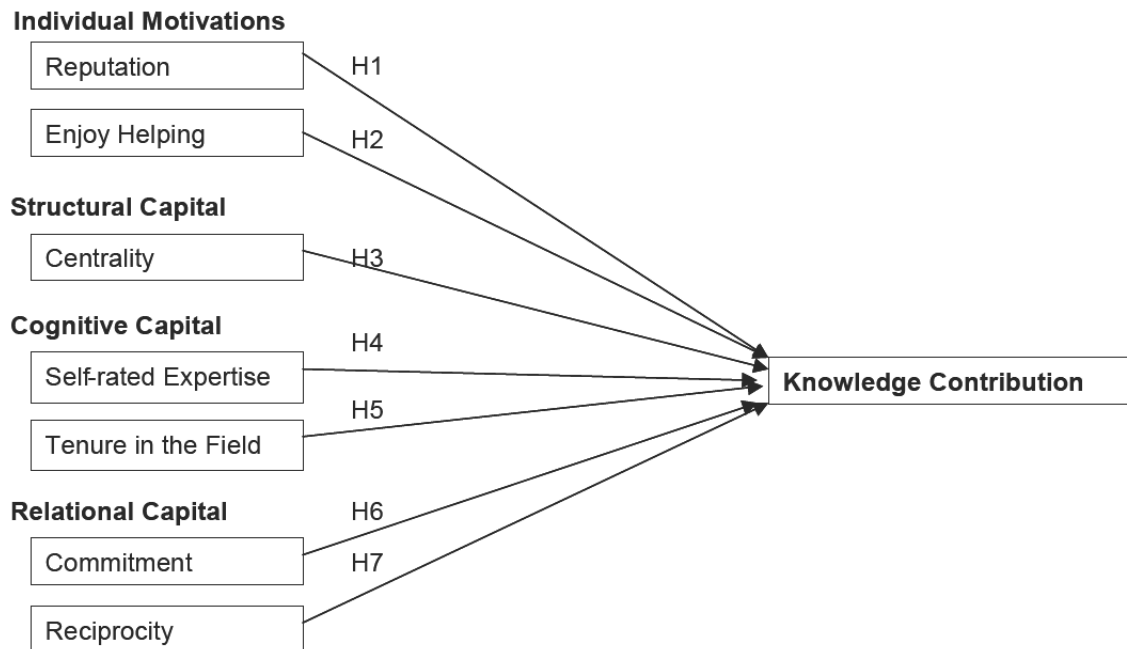
Figure 2.11: The Wasko and Faraj model. Adapted from Wasko and Faraj [2005]

The other two factors encompass both cognitive and relational or structural aspects of the position of the individual member in the online community. Cognitive capital relates with the ability of the member to contribute in terms of expertise and tenure with the field / topic that the community tackles.

2.5 Outlook to the empirical part of this dissertation

The scope of this chapter was twofold. On the one hand, the intention was to introduce the reader to the theoretical concepts underlining the research approach that we adopted on the study of online communities and, in particular, the case of social dilemmas. Social dilemmas on online communities are essentially the reason why understanding and enhancing motivation in these rich social environments is a justified research question to pursue. On the other hand, the chapter aimed to elaborate on

current theoretical development related to the problem of understanding and motivating contributions on online communities such as the models discussed in section 2.4.

We reviewed both structural and behavioral approaches to the study of online communities and provided a modest literature review on the most important models related to cases of contributions to online knowledge communities. In the four empirical studies that follow this chapter we make use of several of the concepts presented here with each of the chapters representing a special research area discussed in this chapter. The case for social dilemmas can be examined further in the chapter that follows where we use a laboratory controlled setting to study behavior in social dilemmas; however, we do not study the inequity aversion part in depth due to the fact that there was no repeated interaction case taking place. This approach along with the case of strong and weak reciprocal ties (as presented in section 2 of this chapter) is set as a departure point in chapters 5 and 6. Both these chapters consider the same setting with a different breed of motivational factors being present (extrinsic vs. intrinsic form). These chapters are in connection with the research objectives B and C and consider as background theoretical ground the theory of social dilemmas combined with the theoretical concepts discussed in section 2.3.

Chapter 6 is a special case of combination of the theories presented in this chapter and the previous. In particular the research approach that we adopt in that particular chapter is a bottom up approach where by examining the actual contributions of the participants on this online social system, we examine the effect that they have on a transaction activity (such as in the cases discussed in chapter 1) where the standard offline contractual commitments do not lead to any consequences for the participants that express their opinion. The rating mechanism discussed in this chapter has a unique characteristic which is the level of usefulness considered by the other participants and, in particular, the justification of this expression in relation to the individual's notion of fairness (as expressed by the review provided for the particular book on that

case).

In retrospect, this chapter is also discussed in the final chapter of this dissertation where we provide a more analytical connection of the theories presented here with the outcome of the subsequent chapters where we also provide (a) how this thesis approaches the research objectives stated in the previous chapter and (b) the relevant contributions that this dissertation provides, both in theory and practice.

Bibliography

I. Ajzen and M. Fishbein. *Understanding attitudes and predicting social behavior*. Prentice-Hall Englewood Cliffs, NJ, 1980.

S. Ba and P. Pavlou. Evidence OF the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *Management Information Systems Quarterly*, 26(3):243–268, 2002.

A. Blanchard and T. Horan. Virtual Communities and Social Capital. *Social Science Computer Review*, 16(3):293, 1998.

G. E. Bolton and A. Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90:166–193, 2000.

S. P. Borgatti and P. C. Foster. The network paradigm in organizational research: A review and typology. *Journal of Management*, 29:991, 2003.

B. S. Butler. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research*, 12:346–362, 2001.

A. Cabrera and E. F. Cabrera. Knowledge-sharing dilemmas. *Organization Studies*, 23: 687–710, 2002.

P. J. Carrington, J. Scott, and S. Wasserman. *Models and Methods in Social Network Analysis*. Cambridge University Press, 2005.

K. S. Cook and R. M. Emerson. Power, equity and commitment in exchange networks. *American Sociological Review*, 43:721–739, 1978.

R. Cornes and T. Sandler. *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge University Press, 1996.

C. Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49, 2003.

J. Donath. Signals in social supernets. *Journal of Computer-Mediated Communication*, 13:231–251, 2007.

E. Fehr and K. M. Schmidt. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114:817–868, 1999.

E. Fehr, U. Fischbacher, and S. GÃd'chter. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13:1–25, 2003.

D. Fisher. Using egocentric networks to understand communication. *IEEE Internet Computing*, pages 20–28, 2005.

L. C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215–239, 1979.

L.C. Freeman, D.R. White, and A.K. Romney. *Research Methods in Social Network Analysis*. Transaction Publishers, 1992.

E. J. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics & Management Strategy*, 10:173–199, 2001.

M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99:7821, 2002.

MS Goldman, SA Brown, and BA Christiansen. Expectancy theory: Thinking about drinking. *Psychological theories of drinking and alcoholism*, pages 181–226, 1987.

A. W. Gouldner. The norm of reciprocity: a preliminary statement. *Sociological Review*, 25:161–178, 1960.

M. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91:481, 1985.

M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233, 1983.

M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83:1420, 1978.

M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78:1360, 1973.

J. Hagel and A. G. Armstrong. Net gain: Expanding markets through virtual communities. *The McKinsey Quarterly*, 1997.

G. Hardin. The tragedy of the commons. *Science*, 162:1243–1248, 1968.

D. Houser and J. Wooders. Reputation in auctions: Theory, and evidence from ebay. *Journal of Economics & Management Strategy*, 15:353–369, 2006.

B. A. Huberman and T. Hogg. Communities of practice: Performance and evolution. *Computational & Mathematical Organization Theory*, 1:73–92, 1995.

Q. Jones, G. Ravid, and S. Rafaeli. Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research*, 15:194–210, 2004.

S. J. Karau and K. D. Williams. Understanding individual motivation in groups: The collective effort model. *Groups at work: Theory and research*, pages 113–141, 2001.

A. J. Kim. *Community Building on the Web: Secret Strategies for Successful Online Communities*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 2000.

B. Klein and K.B. Leffler. The Role of Market Forces in Assuring Contractual Performance. *The Journal of Political Economy*, 89(4):615, 1981.

J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46:604–632, 1999.

M. Koch and W. Wörndl. Community support and identity management. *Proceedings of the seventh conference on European Conference on Computer Supported Cooperative Work*, pages 319–338, 2001.

Joon Koh, Young-Gul Kim, Brian Butler, and Gee-Woo Bock. Encouraging participation in virtual communities. *Communications of the ACM*, 50:68–73, 2007.

P. Kollock. Social dilemmas: The anatomy of cooperation. *Annual Reviews in Sociology*, 24:183–214, 1998.

SS Komorita. A Model of the N-Person Dilemma-Type Game. *Journal of Experimental Social Psychology*, 12(4):357–73, 1976.

N. Korfiatis. Social and economic incentives in online social interactions: A model and typology. *Proceedings of the 30th Information Systems Research Seminar in Scandinavia IRIS*, page 1, 2007.

R. M. Krauss and S. R. Fussell. Social psychological models of interpersonal communication. *Social psychology: Handbook of basic principles*, page 655Ű701, 1996.

J. Lave and E. Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, 1991.

G. LeBon. *The Crowd: A Study of the Popular Mind*. Macmillian, 1897.

K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, X. Li, D. Cosley, D. Frankowski, L. Terveen, and A. M. Rashid. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10, 2005.

P.V. Marsden. Egocentric and sociocentric measures of network centrality. *Social Networks*, 24(4):407–422, 2002.

B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan, and J. Riedl. Movielens unplugged: experiences with an occasionally connected recommender system. *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 263–266, 2003.

J. G. Miller and D. M. Bersoff. Cultural influences on the moral status of reciprocity and the discounting of endogenous motivation. *Personality and Social Psychology Bulletin*, 20:592, 1994.

J.Y. Moon, L. Sproull, K. Aoki, O.K. Darkwa, S. Eskow, J.L. Forbes, and S. Cisler. Essence of Distributed Work: The Case of the Linux Kernel. *Distributed Work*, pages 381–404, 2002.

E. D. Mynatt, A. Adler, M. Ito, and V. L. O'Day. Design for network communities. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217, 1997.

J. Nahapiet and S. Ghoshal. Social capital, intellectual capital, and the organizational advantage. *Sumantra Ghoshal On Management: A Force For Good*, 2005.

P. Nelson. Information and consumer behavior. *Journal of Political Economy*, 78:311, 1970.

M. Olson. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press, 1971.

P. A. Pavlou. Institution-based trust in interorganizational exchange relationships: the role of online b2b marketplaces on trust formation. *Journal of Strategic Information Systems*, 11:215–243, 2002.

C. C. Pinder. Work motivation. *Administrative Science Quarterly*, 37:513–515, 1992.

J. Platt. Social traps. *American Psychologist*, 28(9):641–651, 1973.

J. Preece, B. Nonnecke, and D. Andrews. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20:201–223, 2004.

F. Radicchi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences*, 101:2658–2663, 2004.

A. Rapoport. Mathematical models of social interaction. *Handbook of Mathematical Psychology*, 2:493–579, 1963.

W. Raub and J. Weesie. Reputation and Efficiency in Social Interactions: An Example of Network Effects. *American Journal of Sociology*, 96(3):626, 1990.

P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of the Internet and E-Commerce*, 11:23–25, 2002.

P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay: A controlled experiment. *Experimental Economics*, 9:79–101, 2006.

H. Rheingold. *Smart Mobs: The Next Social Revolution*. Perseus Books Group, 2002.

H. Rheingold. *The Virtual Community: Homesteading on the Electronic Frontier*. MIT Press, 2000.

B. Rothstein. *Social Traps and the Problem of Trust*. Cambridge University Press, 2005.

R. M. Ryan and E. L. Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55:68–78, 2000.

J. Scott. *Social Network Analysis: A Handbook*. Sage, 2000.

R. Shaffer. *Community Economics: Economic Structure and Change in Smaller Communities*. Iowa State Pr, 1989.

M. Spence. Job market signaling. *Quarterly Journal of Economics*, 87:355–374, 1973.

J. C. Spender. Competitive advantage from tacit knowledge? unpacking the concept and its strategic implications. *Organizational Learning and Competitive Advantage*, pages 56–73, 1996.

R. Swedberg. The economic sociology of capitalism: An introduction and agenda. *Economic Sociology of Capitalism*, 2004.

R. L. Trivers. The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46: 35, 1971.

E. Turban, J. Lee, D. King, and H. M. Chung. *Electronic commerce: a managerial perspective*. Prentice Hall, 2000.

B. Uzzi. Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Financing. *AMERICAN SOCIOLOGICAL REVIEW*, 64: 481–505, 1999.

MR Ward and MJ Lee. Internet shopping, consumer search and product branding. *Journal of Product and Brand Management*, 9(1):6–20, 2000.

M. Wasko and S Faraj. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *Management Information Systems Quarterly*, 29:35–57, 2005.

S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *The Structure and Dynamics of Networks*, 2006.

B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythornthwaite. Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Reviews in Sociology*, 22:213–238, 1996.

S. Whittaker, L. Terveen, W. Hill, and L. Cherny. The dynamics of mass interaction. *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 257–264, 1998.

R. L. Williams and J. Cothrel. Four smart ways to run online communities. *Sloan Management Review*, 41:81–92, 2000.

F. Wu and B. A. Huberman. Finding communities in linear time: a physics approach. *The European Physical Journal B-Condensed Matter*, 38:331–338, 2004.

M.H. Zack. Researching organizational systems using social network analysis. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, volume 7, page 7043, 2000.

Part II

Empirical part

Behavioral Characteristics and Cooperation in Online Communities: An Experimental Investigation

Cooperation is an important factor for the sustainability of online communities because it affects the outcome of the interaction between the users. In this chapter the relation between cooperation and online social interaction characteristics is made using the public goods game. We first explain the public goods game and its game theoretical assumptions and then describe the experimental procedure. We use two distinct types of framing (contributing and benefiting), regarding the presence of a subject in an online setting where: (a) the subject contributes to the common good and (b) the subject benefits from it. The framing is distributed in two distinct treatments with an extra treatment acting as a control for offline participation. Empirical results show that the participants in online communities exhibit a high degree of cooperation conditional on the contribution provided by the others.

3.1   Online communities and online cooperation

Undoubtedly, one of the major elements that constitute social capital is the ability of
individuals to cooperate and form structural relations within an organization or a soci-
ety [Putnam, 1995].In the introductory chapter of this dissertation we emphasized the
virtual forms of social capital as a factor that plays an important role to the sustain-
ability and the evolution of social structures on the web such as online communities.
This can be advocated by the fact that networks, norms and trust related issues do, in
fact, interrelate with the theory of social capital and its application to the evaluation of
sustainability and activity of a social structure, either virtual or offline. From this per-
spective of social capital in an online or virtual form, this ability is very much related
to the case of online sociability. In other words, the ability of individuals to form social
interactions in an online setting use the context of an online community in the same
way that an offline community facilitates social interactions. A great deal of research
on online sociability deals with the organization of social structures as a result of fre-
quent computer mediated communication resulting in a form of a network connected
type of communities known in the literature as Online Communities [Wellman et al.,
1996].Online communities pose an interesting field of study on the Internet due to
the fact that they facilitate the formation of ad-hoc social structures which permit the
exchange of information and knowledge over a broad range of topics.

Undoubtedly, the basic compound of these communities is the active participation
of a number of individuals, thus providing information to other individuals who seek
information. Undeniably, there is a need for a critical mass of participants in order for
the online community to sustain and promote the communication activity that takes
place in that setting [Marwell and Oliver, 1993]. Furthermore, for the community to
be sustainable, this critical mass of individuals is expected to co-operate with another
(not necessarily exclusive) set of other individuals who satisfy a particular need by
participating in the online community. From a resource based perspective, an online
community undoubtedly provides a public good which can be consumed by all com-

munity members. Modern economic theory has defined a public good as an economic good that has two major characteristics: (a) non-rivalry: providing that the consumption of this good by a consumer does not prevent another consumer to consume it as well (each consumer has access to the same type of provision of this good) and (b) non-excludability: where consumption does not impose any cost; in other words, the goods are free for everyone to consume [Samuelson, 2000].

Depending on its design, an online community might impose a cost of entry to a participant which, in general, is a cost attributed to social barriers and not necessarily on an actual monetary price (a community functions as a club where the participants who provide the public good might receive some compensation for doing so). In most of the cases, since participation is free, an important factor which affects the value of the information provided in an online community is undoubtedly the cooperation between an often large number of individuals. Further, cooperation plays an important role due to the fact that the production of the public good provided by an online community (in that case information) demands the active participation of two participant states: that of the consumer and that of the producer. These two states are not necessarily excludable. For example, someone who contributes information to an online community can also consume information that becomes available in an online community taking both the role of the information producer and the information consumer. However, studies have shown that there are some online community participants who prefer to consume more information than they actually produce, rather than giving back to the community, such as in social loafing, where participants are inactive and community activity is only dependent on the participation of a very low percentage of the actual participants. From a theoretical viewpoint, such a case is highly related to activity [Schoberth et al., 2003]. The assumption that we follow in this chapter is that low activity in an online community means less contribution to the public good (the overall information) resulting in "free-riding" where participants who do not contribute to the community get benefited in the same degree as those individuals who act more

114

altruistically in relation to the community.

However, what are the characteristics of these contributors/consumers of online communities in socioeconomic terms? Are they actually cooperative persons? Do they have some form of altruism embedded in their behavior? Can we somehow disentangle these two aforementioned states? This study is stimulated by the above research question as departure hypothesis. Our initial goal is to seek whether participation levels in an online community are actually related to the behavior that these participants have shown by participating in an online public goods game carried out with actual monetary rewards and controlled by a random sample drawn from an actual population. Therefore, the context of this chapter is not to study the interaction process in an online community per se, but to elaborate on the social processes that take place during the interaction in an online community and, in particular, on the social dilemmas that arise during the participation of online community members.

Literature has yielded that the public goods game is a standard method to test whether people are actually cooperative in their activities since cooperation is a behavioral characteristic which is embedded in the everyday behavior of an average person [Fehr and Gachter, 2000].]. By controlling for a set of different variables, such as demographics and online communication usage, we are able to define and evaluate our departure hypothesis by using a large pool of subjects that participated in our experiment.

To this end, this chapter is structured as follows. Section 3.3 describes the public goods experiment that we conducted, the procedure behind its dominant strategy and its implications related with cooperation characteristics (such as altruism and cooperativeness). Section 3.4 provides the experimental protocol used, the framing used to disentangle the characteristics above, the methods and the procedure that was followed both for the standard (unconditional) public goods game and a variant public goods game based on interval-based conditional choice. Sections 3.5 and 3.6 report a thorough analysis of the experimental data across the different treatments

and framing used, as well as a discussion of the results. The study concludes in section 3.7 with concluding remarks and issues for further investigation.

3.2 Motivation

Our motivation to study the problem of contributions to online communities from the prism of the public good game was intrigued by the fact that online communities do resemble a social space where the production of a public good takes place. This is evident from the following:

- the digital nature of information as a public good since the production and consumption of information by an individual does not directly affect the utility that another individual will receive by performing the same actions;

- the way we perceive an online community as a space where everyone (depending on the membership policy of the community mechanism) can participate, contribute information and be benefited by the information that is already available;

- the social dilemma that is framed by the temptation of an online community member to free-ride on the effort contributed by the other members; thus, if everyone tries to free-ride, then the public good produced by the community will suffer and gradually dissolve [Kollock, 1999].

Related to the case of the social dilemma in a public goods game (or the public goods dilemma as it is known in the literature) provided in the context of an online community, let us provide some examples for why is important to address the problem of contribution in an online community from the perspective of the public goods game. Let us consider the following scenario. In an online community such as the Yahoo!Answers online communityYahoo!Answers online community[1] a user posts a

---

[1]a more analytical study on the case of contributions in Yahoo!Answers is presented in the next chapter

question related to a topic. Another user responds to this question by providing an answer. The user receives this answer and a piece of information is produced out of this interaction. This information becomes available to every other member of the Yahoo!Answers online community which he/she can use in case there is a similar information need. Here the public goods dilemma can be seen as follows: If a member of the Yahoo!Answers community only posts questions and receives answers in decent time by the other community members without contributing on other open questions (that he/she possesses some expertise to answer), then the information available in the community (which in our case is the public good) will stop evolving and become less useful. On the other hand, if all members give back to the community by taking some time to answer questions in which they have some expertise, then the online community will increase the value of the public good that it provides to its members. But how can we distinguish these cases and assess what makes members of an online community contribute more or equal to that which they take from the community?

Another interesting example, as highlighted by Lerner and Tirole [2002], is the case of open source software. Positioned on successful open source software projects, such as the Apache HTTP server (which is the standard solution for providing access to web pages on the Internet), a very large number of users is benefited by it (e.g., network systems administrators, web developers, etc.), while few people contribute to its development, leading to a case of asymmetry between contributors and users that receive high benefit from it (e.g., large companies providing web hosting services). While contribution is difficult to measure in such cases (e.g., a contribution can be the case of pointing out a software bug or providing some documentation or a patch of source code to enhance a function of the product), the case remains the same. A large number of users are benefited by the effort of a lesser number of contributors who also receive a benefit from the software project that they develop. In such cases if everyone uses the software without contributing, then the software will gradually lose its value and dissolve (no participation will affect the development effort of the

community).

In both of the above described cases the nature of the public good (whether information or software source code) remains the same. It is not rival and not excludable. Therefore, the public goods game can be an ideal instrument to test whether participants in online communities do show some degree of cooperation (related also with intrinsic characteristics such as altruism, as shown in the literature) or are selfish individuals looking towards satisfying their own individual needs. Furthermore, the public goods game in combination with background data (such as socio-economic characteristics) can help to identify whether the communication setting of an online community differentiates its members from rational selfish interests.

However, before proceeding to describe the abstraction followed in this chapter in order to study the process of contribution to an online community using the public goods game, we need to clarify the applicability of this scenario. First, as stated in Chapter 1, we consider the case of an online community where an exchange relation is established between each member with information being the exchange subject. This direction can be seen as an extension of the definition of an online community already discussed in Chapter 1 where we consider an online community as a collective, facilitated through online communication means, of participants who interact over a shared purpose, interest or need. Thus, the shared purpose and interest is the answer to questions or knowledge sharing, and the public good is the overall information that is accumulated through the exchange of communication between the community participants. The more the members of the online community (e.g., a newsgroup or an online forum) participate, the higher will be the outcome of their interaction, and the better will be the quality of the public good formed out of the exchange in the context of the online community.

Having described the motivation and the abstraction prism for using the public goods game as an instrument to study interaction in an online community, we develop how this game was setup in an online environment to evaluate the actions of potential

118

online community members recruited from the standard population.

3.3   Cooperation and the public goods game

In a standard public goods game a subject is asked to make a decision about the contribution of his resources to a common goal where in cases of full cooperation will generate a high return value if the other subjects do also cooperate. From an experimental procedure point of view the protocol requires a set of subjects to be randomly picked and assigned to a fixed size group (to allow for comparison). The participation in the group is done with anonymity as a main issue so the subjects will not be informed at any level about the characteristics of their fellow group members. Then the subjects are endowed with an amount P_e which they have the right to keep or provide (all or part of it) to a pot which represents the effort contributed to the common goal (e.g. the amount of work they contribute to a team project).

After the subjects decide to make their contribution the budget allocated for the common goal is multiplied by a multiplication factor (e.g. doubled, tripled or quadrupled) and divided back to the group members. The multiplication factor represents the actual return that the group had by supporting the common goal. For example in case the endowment of each group is measured in hours then if all the group members devote all their available hours in the project's account then the project succeeds and the return for the group is the equivalent of twice the cost that the hours for each member required.

Table 3.1 provides a numerical illustration of the public goods game configuration that was used in our experiment. For this configuration the group size was four and the multiplication factor of the pot was two. Each subject is assigned in a group together with three other subjects resulting in a group of four subjects assigned to contribute to a common pot which as aforementioned represented a framing of the common budged for a goal. Each subject was endowed with the amount of 50 Danish Crowns.

From this amount the subjects can make a decision of how much they will keep and how much they will provide to the common good. So if a subject decides to contribute an amount P_c to the public good then the budget condition in that case will require $P_c \leq P_e$.

The amount that the subject will have left after its contribution to the common good would be $P_e - P_c$. The maximum contribution that each subject can make to the common pot is equal to its endowment and is $P_c = P_e = 50$ DKK. On the other hand if the subject decides that it doesn't worth contributing to the common good at all, he/she can keep all the money and make a zero contribution to the pot ($P_e - P_c = 50$).

The pot's multiplication factor as aforementioned represents the return yielded by the contribution of all the group members. The numerical example considers 4 subjects initially endowed with an amount of $50DKK$. Each subject decides to make the contributions as follows:

The pot is configured with a multiplication factor which doubles the amount contributed as a return of the group's effort. On that case the amount returned by the pot will be 180. This amount will be distributed equally among the four group members and will be $P_p = 45$. The total earnings of each group member are represented in Table 3.3.

From the results we can clearly observe two cases: the "yellow" subject and the "blue" subject. The first subject decided to contribute nothing to the pot (Pc=0) getting the maximum total amount out of his participation to the group (95 DKK) while the "blue" subject that was the most "altruistic" in the sense of the group and contributed all the amount that it was endowed finally received less that it was endowed in the beginning. Those two cases present the best and worst case of the group behavior in terms of contributions in a public goods game. In fact the zero contribution is the dominant strategy for a participant of a group since "free-riding" where punishment is not imposed is the most profitable strategy. This is in fact the Nash Equilibrium of this game since the dominant strategy yields that for a particular subject there is no other
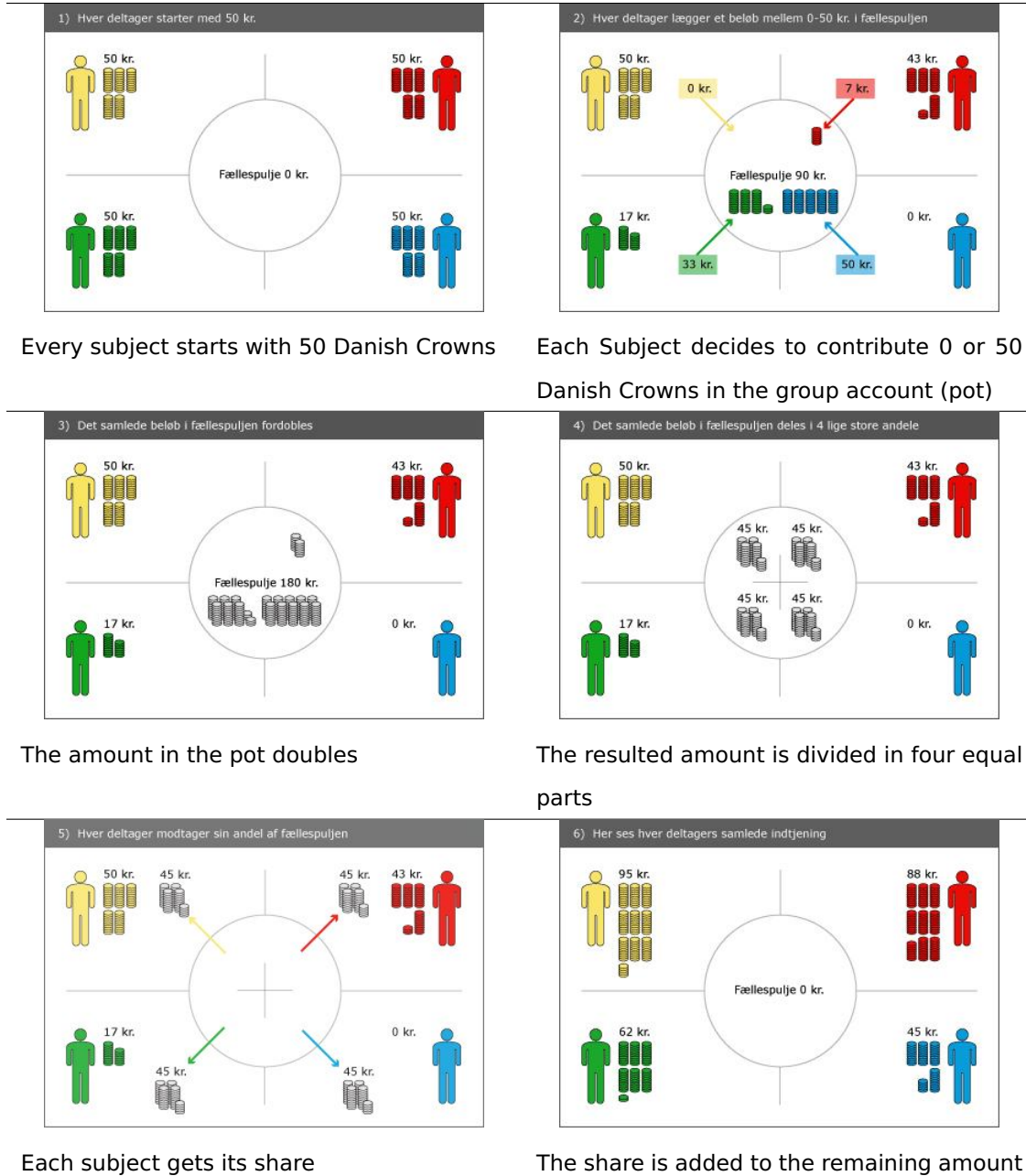
Every subject starts with 50 Danish Crowns

Each Subject decides to contribute 0 or 50 Danish Crowns in the group account (pot)

The amount in the pot doubles

The resulted amount is divided in four equal parts

Each subject gets its share

The share is added to the remaining amount

Table 3.1: A numerical example of the public goods game for a group of four members contributing to a common pot

Subject	Amount Contributed	Amount Kept
Yellow	0	50
Red	7	43
Green	33	17
Blue	50	0
Total Contribution		**90**

Table 3.2: The contributions table for the example presented on 3.1

Subject	Amount Returned from the pot	Amount Kept	Subject Earnings
Yellow	45	50	95
Red	45	43	88
Green	45	17	62
Blue	45	0	45

Table 3.3: The payoff table for the example illustration of 3.4

subject which can result in a better return for the particular subject.

Coming back to the case of information as a public good the above illustration of the public goods game provides one fundamental case. Free riding without sanctioning is the most rewarding strategy for each group member to follow.

However the public goods game has another state which surpasses the returns yielded by the *Nash equilibrium*. In particular in the above example if all the group members were collaborating for the common goal by contributing all the amount that had been initially endowed then the return of their contribution would have been per subject: $P_p = 0.25 * (50 * 4) * 2 = 100$ DKK which results to the exact double of their initial endowment. However this particular state requires that all the subjects will cooperate with the same amount of effort in order to ensure maximum return from the project.

Reshaping that case on the context of an online community, an online community with no free-riders is providing the maximum utility to its members. However excluding free riders in an online community is not an easy task since anonymity prevents permanent sanctioning while preventing a member to access the public good (information) available on the online community might pose barriers to other members as well.

From the above is clear that the best case scenario for a community is to be formed by members who are in general cooperators. However the above configuration exemplifies cooperation in the context that the average contribution of the others is not known prior to the subjects own contribution. It might be that some subjects are *"conditionally cooperative"* resulting that they actually be willing to cooperate if the others do so. If the others are cooperating then an online community member will also consider contributing as well.

In connection with the activity in online communities this is another case of how the evidence of prior activity influences participation in an online community [Johnson, 2001]. Usually in online communities prior activity is a key element to the attraction

Each subject starts with zero Danish crowns

Each participants takes an amount between 0 and 50 crowns from the group account (pot)

The remaining amount in the pot is doubled

The resulted amount is divided in four equal parts

Each participant receives its share

The share is added to the subjects' initial amount
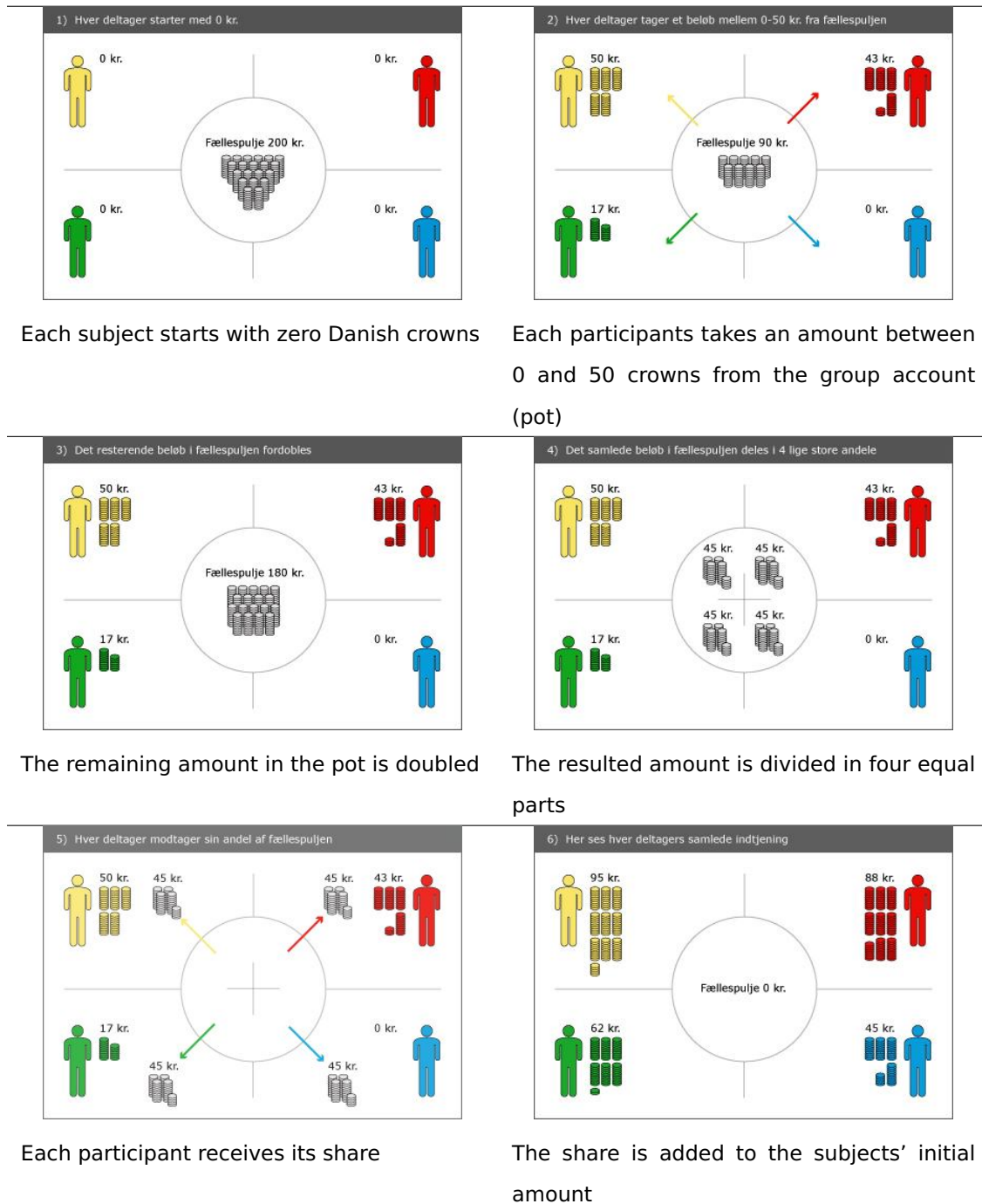
Table 3.4: The same numerical example configured with a TAKE framing for the public goods game

and retention of new members.

In order to control for psychological effects imposed by the method of contribution, a second configuration of the public goods game was available to the subjects as a treatment. In that treatment the subjects were required instead of contributing to a common pot, to take from an already filled one.

On that case instead of contribution the subjects were able to take from the already doubled pot with a limit $P_e = 50$ DKK. The rest were once more doubled and divided equally among the members of the group and distributed. Again the dominant strategy is to take the most out of the pot. If the members were fully cooperating for the common good and deciding not to take out any of the initial amount then this full cooperation would result to the highest available return rate of 100 DKK per subject.

Having summarized our experimental configuration we proceed with the description of the protocol and the assignment of the subjects to treatments representing the different types of framing discussed above.

3.4   Experimental procedure and methods

The experiment was conducted online using a web application that was programmed for that purpose. Figure 3.1 shows the login interface of this application. The selection of the subjects was conducted with the collaboration of the National Statistics Bureau which drew a random sample of the Danish population around the country. Due to the fact that citizens in Denmark need to be associated with a registration number for social security purposes the statistics office was able to obtain the factual mail addresses of the subjects and send a recruitment letter. The recruitment letters were divided in three types: Letter type *A*, Letter type *B* and Letter type *C*. In letter types *A* and *C* it was clearly stated that there will be a monetary reward for their participation while in letter type B it was generally written that their participation will help research in Denmark. The letters contained a personal identification code and the

web address of the web application. The personal identification code was randomly generated and was used as an identifier to associate data for a subject for different parts of the experiment. Each set of letters was assigned in a set of two waves (First wave and Second Wave). The first wave was constructed mainly for participation rate identification purposes while the second wave was the main wave of the experiment.

Although not stated in the letter the subjects were informed by the platform that they could login on a time window which was roughly 7 days from the time they received the letter in their mail box. After that date the subjects were receiving a message that they were not able to login with this specific pincode because it has expired.
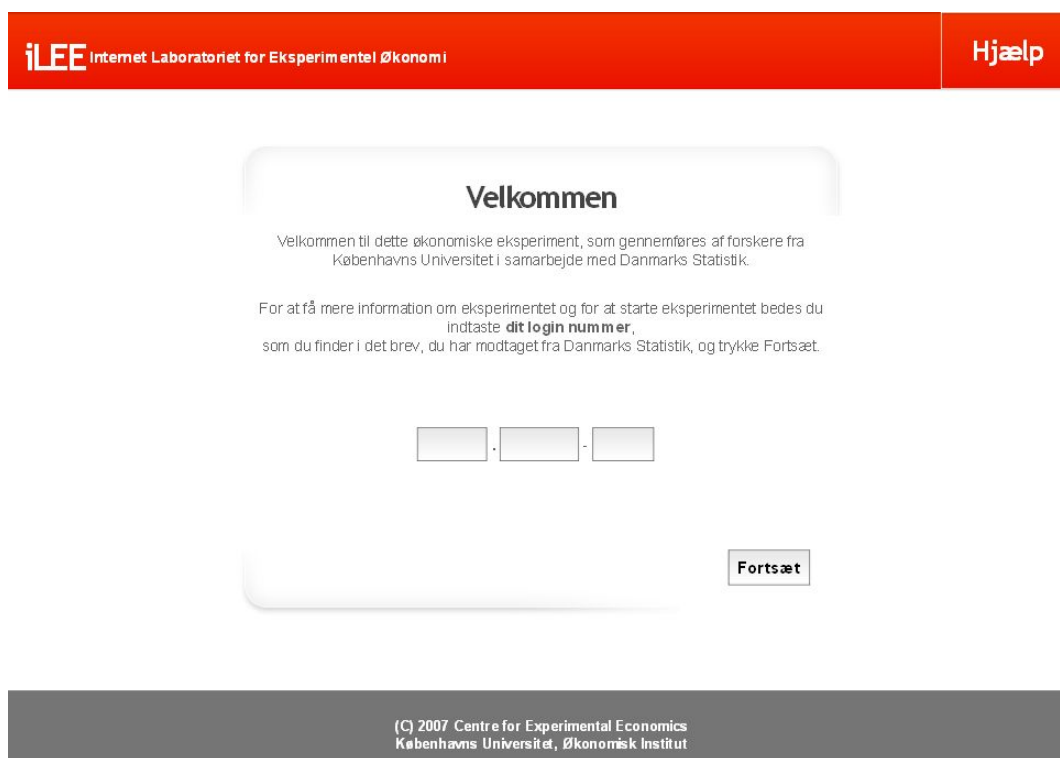
Figure 3.1: The login Interface of the Web application

For the evaluation of different frames the subjects were randomly assigned to one out of a set of three treatments. The two basic treatments were to distinguish the framing in the public goods game as follows

- Treatment 1: Public goods game with give configuration

- Treatment 3: Public goods game with take configuration

Treatment 2 was a public goods experiment with give configuration as well; however it was not giving an actual monetary reward to its participants (economic incentives were absent).

Figure 3.2 depicts the participation rate for the experiment for the two waves since the letters was sent to the subjects. The first wave was an preliminary one to asses the participation level of the invited subjects (2500 invited subjects) and the second wave was the main wave of the experiment (19.500 invited subjects). Both waves have a similar participation rate which averages 15% of the invited subjects. A 16 hours time window was added to control for postage delays. It is clear that there is faster response rate to the subjects for the fourth wave which can be slightly explained by the fact that the time window for participation on the experiment for the third wave included the weekend where participation was expected to be low.

An issue when comparing the two waves might be the demographic or geographical scattering of the recipients of the letters (e.g. lower participation rates would be expected for older people in rural areas). However the participation rates converge at the end of the time window therefore the actual response rate within the time window of the experiment doesn't impose any selection effect on the wave comparison.

3.4.1 Experimental protocol

Figure 3.3 shows the protocol of the experiment. As aforementioned subjects were recruited from the standard population using online invitations. The configuration for the give and take types of framing was a sequence of a set of instructions and decision screens where subjects had to read and then input their decision allonge with other data.

The sequence of the decision states was as follows:

Participation Rate (Wave 1:2500, Wave 2: 19500 invitation letters)
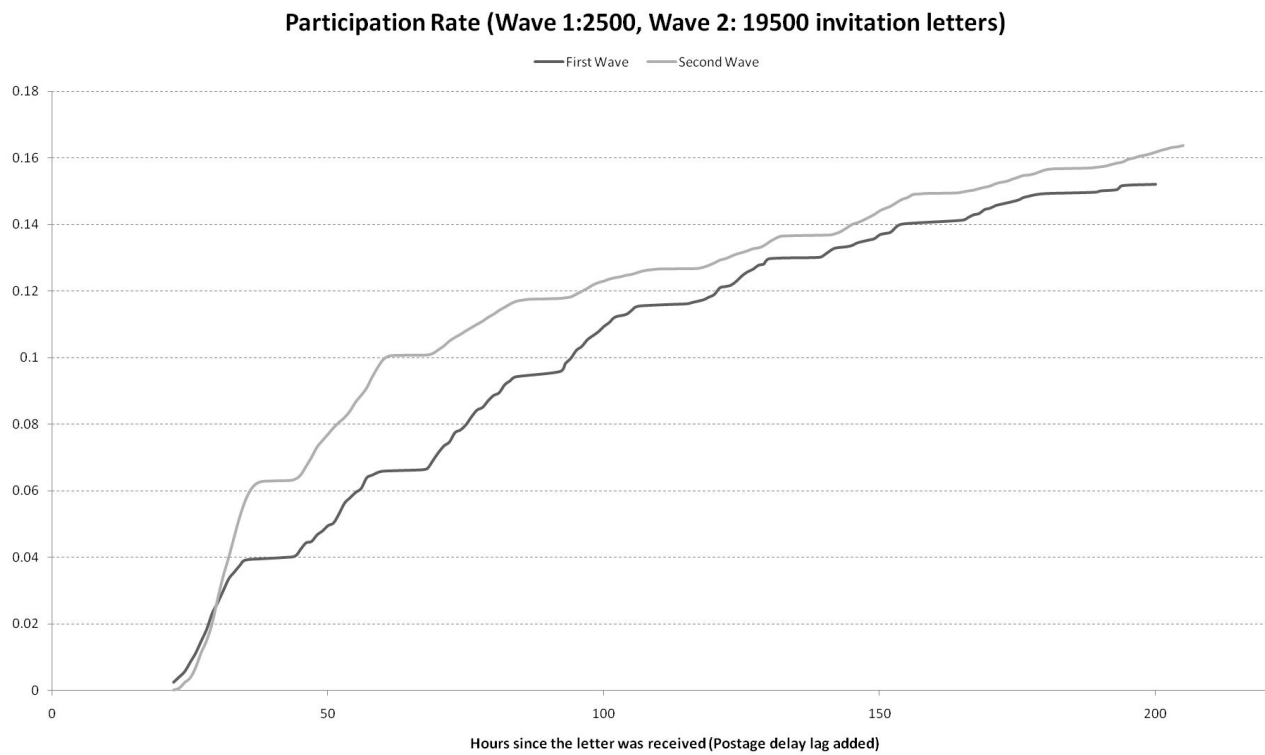
Figure 3.2: The evolution of the participation rate for the two waves that were used in this experiment (Wave 1: green line , Wave 2: red line

1. The introductory and demographics state where subjects were asked to give their demographic details read the instructions of the public goods game and answer to a set of control questions to test whether they understood the questions provided by the public goods game. *(introduction)*

2. The decision about how much of the endowed amount they will contribute to the common goal (unconditional choice)

3. An estimation of how much did they thing that the other members of the group contributed on the common goal.

4. The decision about how much of the endowed amount they will contribute to the common goal if they know how much the other members of the group contributed on average (conditional choice).

5. The online community participation characteristics (online community) where subjects had to answer to a set of questions related with online sociability characteristics such as for example how often do they communicate etc.
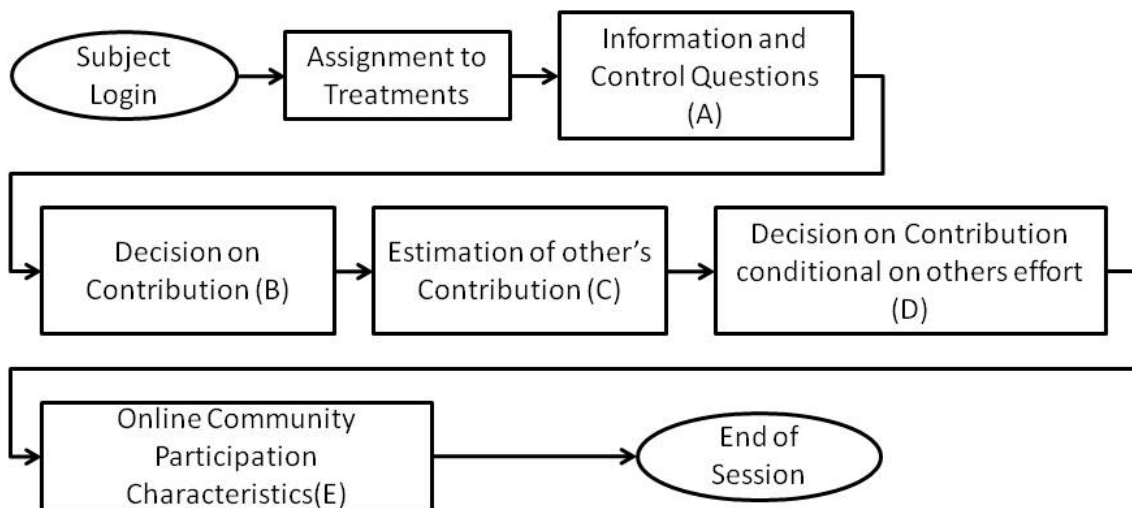
Figure 3.3: The experimental protocol used in this study.

The Take framing had exactly the same configuration with the contribution to be

presented as $P_g - P_c$ so the maximum amount that a subject would take from the pot would be 50 Danish Crowns.

After the experiment was over the subjects were asked to login again to the platform in order to see how much they earned and enter their banking details in order to receive the amount they earned. The next section discusses the assignment of the subjects to treatments.

3.4.2 Assignment to treatments

Treatment assignment was based on two parameters: the Letter type and a randomly generated two point decimal variable (between 0 and 1) generated once a subject accepted the invitation and logged in to the system. Table 3.5 depicts the procedure of the assignment to the treatments.

	Treatments		
Letter Type	Treatment 1 (Give: Economic Incentive Stated)	Treatment 2 (Give: No Economic Incentive Stated)	Treatment 3 (Take: Economic Incentive Stated)
A	RN < 0.66	-	RN $\geq$ 0.66
B	RN < 0.5	RN $\geq$ 0.5	-
C	RN < 0.5	RN $\geq$ 0.5	-

Table 3.5: Randomization procedure for the assignment of subjects to treatments according to the letter type. RN denotes the random variable value generated by the system for the treatment assignment

The give framing was divided in two treatments to observe where an a priori economic incentive might lead to different results in initially participating on the experiment.

3.5 Data analysis and procedures

Having described the procedure of recruiting the subjects and assigning them to treatments we describe the basic demographic characteristics of the subjects and we provide an intra-treatment comparison for the variables that we are interested to analyze.

3.5.1 Distribution to treatments and basic demographics

As aforementioned the data were collected by running the experiment in two separate waves (wave 1, wave 2). Table 3.6 depicts the distribution of the subjects to the waves along with the letter types included.

	Wave	
Letter Type (Completed/Assigned)	**First Wave**	**Second Wave**
A	153/*1503(10.1%)*	1889/*16524(11.4%)*
B	24/*517(4.6%)*	91/*1483(6.1%)*
C	36/*483(7.4%)*	98/*1517(6.4%)*
Total Subjects Completed the Experiment	213*(9.3%)*	2078 *(10.6%)*
Subjects originally assigned	2503	19524

Table 3.6: Distribution of Letter Types by waves.

As can be seen in the table the assignment in separate waves (First Wave, Second Wave) was done highly asymmetrically. That was due to the fact that the first wave acted as an identification wave in order to estimate participation rates for the second wave. This is depicted also on the number of subjects completing the experiment. For the first wave the full response rate was around 8.5% while for the second wave the actual response rate was a little bit higher ~10.6%. As aforementioned this might be attributed to the fact that the first wave includes the weekend of the week 20 (2008),

so physical absence of the subjects from their official residence might have affected the participation rates.

Table 3.7 depicts the distribution of the subjects to the waves and the subsequent treatments. An interesting observation from the participation rate comes from the comparison between treatments 1 and 2 for the letter types B and C where the participation ratio was not significantly different although there was the controlled absence for an economic reward in both letter types for the second treatment.

An important aspect of an experimental procedure is the representation of demographics within and between treatments. Such an important demographic for every experimental procedure, can be the case of gender representation on treatments [Eckel and Grossman, 1998]. As can be seen in Figure 3.4 both genres are represented equally in the two waves of the experiment. In particular for the first wave we had 104 male and 109 female subjects participating while in the second wave we had a slight difference of with 1078 male and 1000 female subjects participating.

The equality between the genre representations can be also attributed to the sample selection procedure which was facilitated by a random sample of the Danish population. Genre representation can be also seen in Figure 3.5 where the distribution by treatment displays also no significant difference between genres. In particular for Treatment 1 we have 770 males (51%) and 728 (48%) females, for Treatment 2 we have 55 males (47%) and 62 (52%) females and for Treatment 3 we have 357 males (52%) and 319 females (47%).

Although genre is one important demographical factor for a population another important factor is the distribution of the age among treatments. We defined four basic groupings for the age variable as: Age group 1 (Age$\leq$30), Age Group 2 (Age>30 and Age $\leq$40), Age group 3 (Age>40 and Age$\leq$50) and Age Group 4 (Age > 50).

Figure 3.6 depicts the assignment of subjects in age groups according to our categorization. As can be seen also in Table 3.8 there is an over-representation of the age group of subjects that are older than 50 years old while on the contrary we have

Letter Type (Completed/Assigned)	Treatment 1 (Give: Economic Incentive Stated)	Treatment 2 (Give: No Economic Incentive Stated)	Treatment 3 (Take: Economic Incentive Stated)
A	1366	-	676
B	47	68	-
C	85	49	-
Total Subjects Completed the Experiment(Treatments)	1498	117	676

Table 3.7: Distribution of Letter Types by framing and treatment
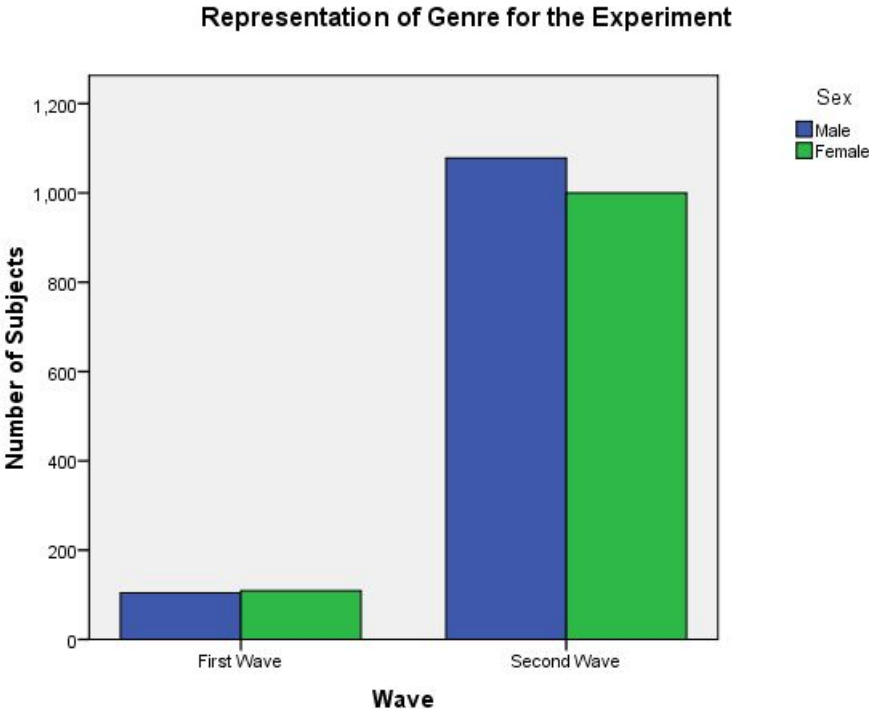
Figure 3.4: Distribution of genres representation in the experiment. The sample selection procedure has resulted into equal representation of both male and female subjects to our dataset
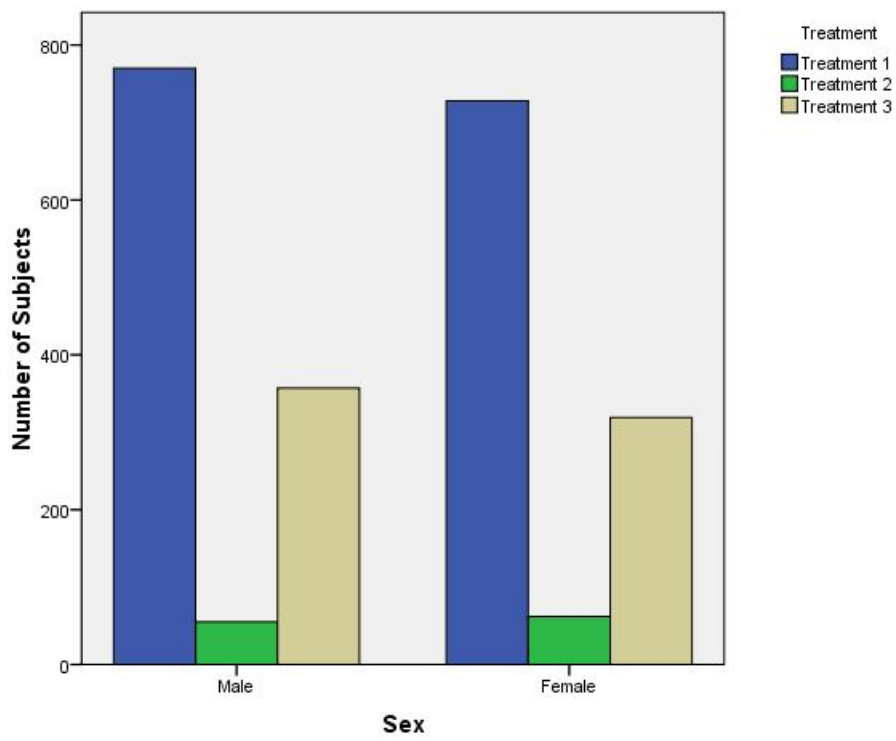
Figure 3.5: Representation of genre among treatments

no representation of younger subjects on the second treatment (Age < 30). For the second treatment in particular we have almost equal representation for subjects that are older than 40 while subjects from 30 to 40 years old are much less represented.
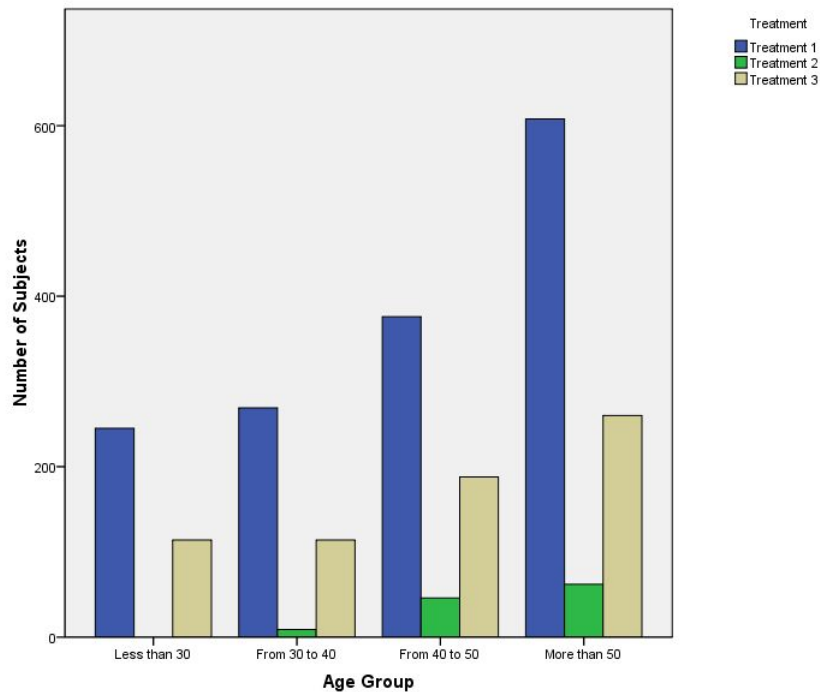
Figure 3.6: Distribution of subjects by treatments and age group

The third informative variable that is important for our analysis is the education level of the subjects. As can be seen in Figure 3.7 there is a high trend towards the participation of well educated subjects. This can be expected from the fact that in order for the subjects to participate in the experiment there was a selection effect on the literacy level due to the requirement to access the experiment interface using a computer with an internet connection. This is also depicted on the participation of highly educated subjects which are over-represented (around 50% of the cases) as can be seen also in Table 3.9.

Having defined the demographic variables (which we consider as informative) we proceed with the analysis of the variables that express online sociability.

Age Groups	Treatment			
	Treatment 1	Treatment 2	Treatment 3	Total
Less than 30	245	0	114	359
From 30 to 40	269	9	114	392
From 40 to 50	376	46	188	610
More than 50	608	62	260	930
Total	1498	117	676	2291

Table 3.8: Age groups by treatment.
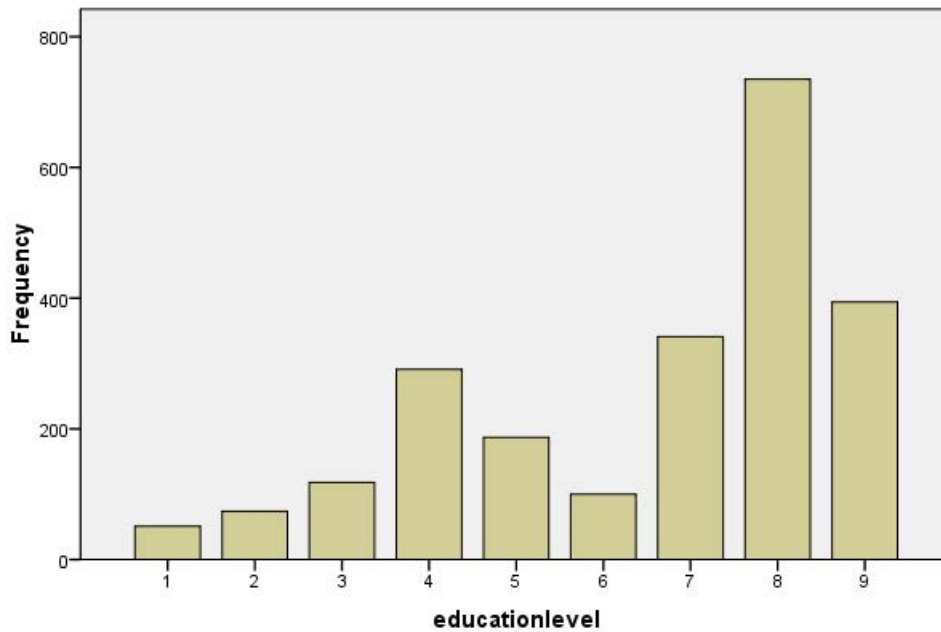
Figure 3.7: Variation of the Education Level among the Subjects for First and Second wave of the experiment

Category	Category Code	Frequency	Percent	Cumulative
Less than mandatory schooling	1	51	2.2	2.2
Mandatory Schooling	2	74	3.2	5.5
Extended Schooling (10th grade)	3	118	5.2	10.6
Practical Education – Technical School	4	291	12.7	23.3
Standard High School	5	187	8.2	31.5
Professional High School (pre-College)	6	100	4.4	35.8
Short University Education (less than three years)	7	341	14.9	50.7
Medium University Education (3-4 years)	8	735	32.1	82.8
Long University Education (more than four years)	9	394	17.2	100
Total		2291	100	

Table 3.9: Representation of the Education Level by the participating subjects for all three treatments (First and Second Wave).

3.5.2 Defining online sociability

As aforementioned the objective of this chapter is to study the case of social dilemmas (as expressed by the contribution in the public good) in the context of an online community. To achieve that we need to control for a set of characteristics that whiteness how active a subject is in the case of participation of an online community. Since the controlled setting of this experiment presents in an abstract level the interaction process in an online setting, this set of characteristics should provide information on efficacy issues such as e.g. how important is for this person to have a membership in the online community and interact with others.Preece et al. [2004] for example argues that efficacy parameters reshape the online behavior of a participant (thus his/her willingness to cooperate in an online setting) while forming a distinct sociability characteristic which represents the degree to which a participant in an online community is active and participating.

In the design of this experiment we considered three main variables for measuring online sociability as a latent construct which can be used in the analysis of the contribution and cooperation levels. The first was the level of participation on online social networking services such as Facebook, MySpace, LinkedIn etc. The second and third variables were the use of chat services and Voice IP services. For the above three constructs we used the times of user of computer per week as well as of how often the subject communicates by email as control constructs.

With the introduction of platforms such as Facebook for example several users consider important to maintain and update their profile in order to signal to their friends about their activities and status. Our motivation for considering participation in online social networks as an efficacy factor is the assumption that since subjects participate on online social networks do have the potential to participate in online communication activities as the ones taking place in an online community. Participation on Online Social Networks was measured using three standard likert-type ranking questions where subjects had to answer the following:

- (a) How much time they spent maintaining their online profile (scale 1 to 6)

- (b) How often they were interacting online with their friends and (scale 1 to 6)

- (c) How important were the connections that were added in their profile in relation with their off-line activity. (scale 1 to 6)

The reason for using the 6 point likert scale rather than the standard 5 point, was to have an easier classification (so the subjects could not select the value in the middle but select an option that was either on the one case or the other. The intuition behind the selection of the above three questions was based on the study by Lampe et al. [2007] in relation with participants' use of profile elements and social connection importance as ways to signal their offline social activities. The theoretical ground for the importance of online profile maintenance in an online social networking website comes from the definition of signaling theory [Donath, 2007]. For example the inclusion of high and up to date volume of information in an online profile provides how important is for an online community member to have the online profile maintained so the participant can indirectly communicate personal values to his/her friends or maintain it so that his/her friends can judge his/her characteristics.

The second and third characteristic that we asked had to do with the quality of interaction with their online social relations and the user perceived importance of them. A study by Cummings et al. [2002] argues that several participants in online social networks consider online relations less valuable than offline relations, however one element that has to be taken in consideration on that case is the perceived importance of these relations. For example the more important the participant perceives his/her relation with another person the more is expected to interact. Wellman et al. [2001] considers this case as important since the more offline relations are sustained in an online form the more important it will become for participants to maintain their online presence.

Although participation in online social networks can be measured in a more pre-

cise manner (e.g. by observing how much time the user has spent on a particular online social network) the elicitation of this indicative group of variables provides independence from the function of the online social network respective with the user's objective. For example the time spent to maintain a professional profile in LinkedIn can be seen for some of the users as an investment to their professional visibility and the ability to tackle new professional opportunities by using their network.

Undoubtedly a limiting factor to this approach is the relation of the online community participation with the amount of time a particular subject uses its computer. As can be seen from the education level of the subjects almost half of them are well educated and the strong correlation between the times they use their computer and the amount of education indicates how important the computer usage for their job function is. Figure 3.8 shows the computer usage among subjects of different age groups.
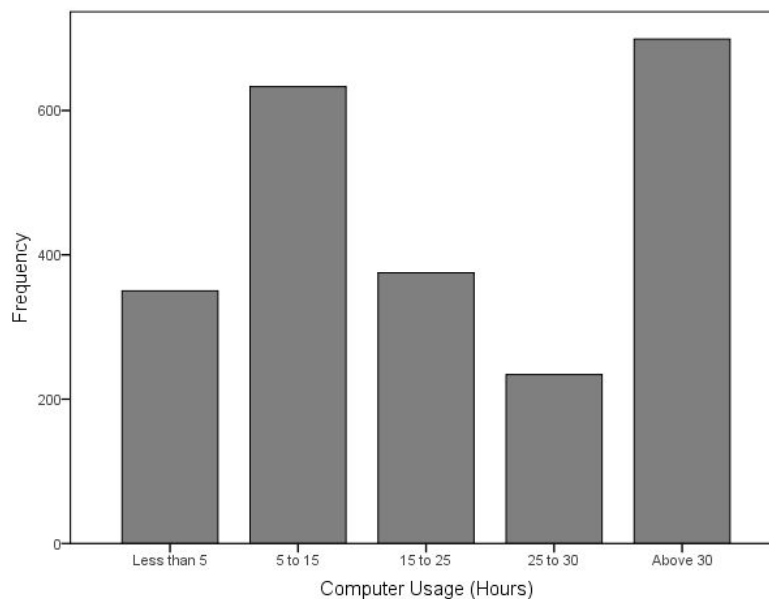
Figure 3.8: Computer usage among subjects of different age groups.

On the other hand online communication does play also an important role to online sociability due to the fact that it provides an exogenous construct for online activity. This was measured by using two indicators. (a) The frequency of times that a subject

141

contacts other using instant messaging services such as for example MSN or Yahoo Messenger. (b) The frequency of times that a subject communicates with other subjects using voice communication provided by Voice IP services (such as Skype, Google Talk).

As also with the case of online social networks, we control the communication activity with another variable that indicates the familiarity with that way of communication. In that case we used the number of times the subject uses its email to contact his/her friends.
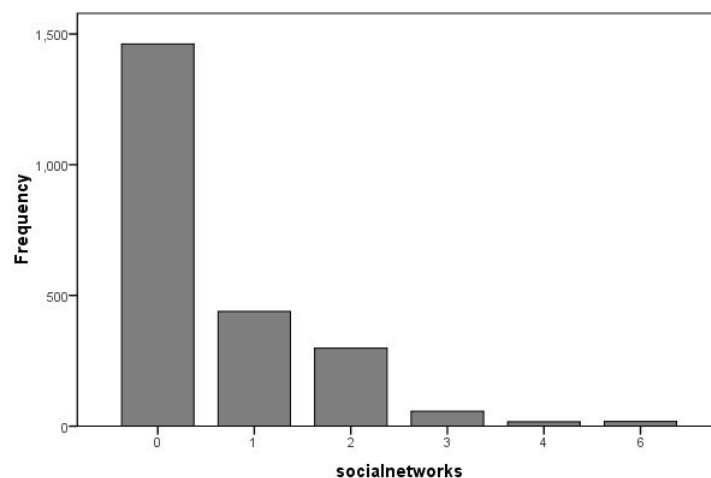
Figure 3.9: Participation and communication on online social networks for all the subjects participating on the experiment.

The use of email as an indication for online sociability has been done in several research studies in the literature where email is treated as an indication of community activity [Kavanaugh et al., 2005]. Early studies of activity in online communities such as those of Wellman et al. [2001] and Whittaker et al. [1998] consider the email communication as an indicative variable of online social activity. With the advent of new ways of communication over the web such as web based fora this trend seems to change [Sproull et al., 2007] since online communication can be formed in the web without the use of email. Furthermore contact by email represents only an asynchronous form of online communication where social interaction doesn't require

participation of both parties at the same time. Therefore in order

However contact by email represents only an asynchronous form of online com-
munication where social interaction doesn't require participation of both parties at
the same time. From that perspective it can be argued that (frequent) email contact
doesn't ensure a strong form of social interaction since there are cases where email
contact is used for reporting activities in e.g. a workplace context and not in essence
social activity[Cummings et al., 2002].
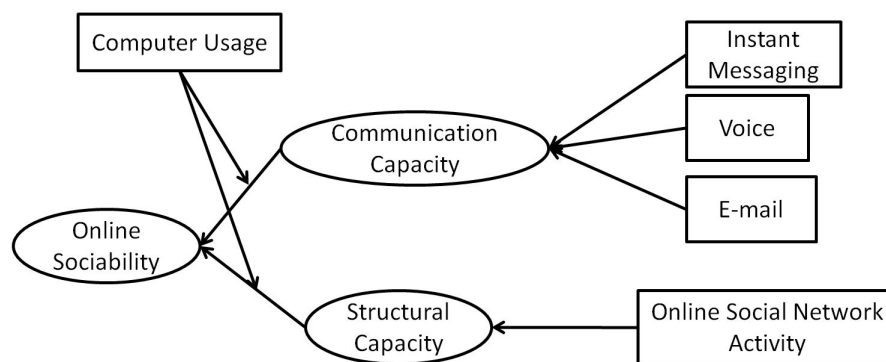
Figure 3.10: Communication Capacity, Structural Capacity and Online Sociability

The measure that we use to track online sociability is a combination of two latent
constructs: *Communication Capacity* and *Structural Capacity*. As can be seen in Fig-
ure 3.10 these two constructs are controlled by the times of computer use by the
subjects. This in facts controls the infrastructure capacity that the individual has in
order to participate in an online cooperation effort as can be seen in the case of an
online community. Communication capacity was measured as a linear factor of two
communication types: synchronous communication (instant Messaging and Voice) and
asynchronous (email). Structural capacity was measured as a linear factor of partic-
ipation in online social network activities (as elicitated from the appropriate control
questions during the experiment).

On top of those two constructs we add the amount of time the subjects spent in
their computer (timesusecomputer) as a control in order to account for cases where
a subject for example does use his/her computer less time than another subject but

from a percentage viewpoint spends more time to communicate online rather than doing other tasks. Furthermore several participants might use a lot their computer at work or to perform functions like shopping online which can add a lot of noise to the defined construct if we don't normalize it.

As aforementioned we treat Online Sociability (OnlineSociability) as a linear combination of the above constructs controlled by times of computer use. The reason for using a linear combination is on the one hand to have the latent variable constructed as simple as possible and on the other hand to make it equally dependent on the sub-construct values (communication capacity and structural capacity). Therefore the linear combination is as follows:

$$OnlineSociability = \frac{1}{timesusecomputer} \left[\frac{chat + voice + socialnetwork}{6} + emailcontact \right]$$

As aforementioned the chat, voice and social network variables were measured as ordinal variables (using a scale from 1 to 6) and the amount of computer usage and frequency of email communication as interval variables(the subjects were asked to state how many hours per week they were using their computer and how frequent they were communicating by email). Therefore in order to normalize the constructs together we divided them with the range of the likert scale (6) in order to able to normalize them with the timesusecomputer variable which controls for how much of the percentage of the overall time the subjects have, they use it for online social activities.

3.5.3 Online sociability and cooperation

Having defined our instrument for measuring online sociability we continue with the relation between online sociability and cooperation on the context of an online community where no-interpersonal bias was present. To avoid group effects to the behavior of the subjects we randomly assigned them in groupings of sizes between 4 and

12 (following a normal distribution).

Contributing to the Community regardless what the others do

On this case we examine the relationship of the unconditional choice to the online sociability characteristics that we defined in the previous section. For that purpose we only used the data from the first two treatments where the subjects were asked two times to make a non conditional choice of how much would they contribute (initial endowment was 50 units).
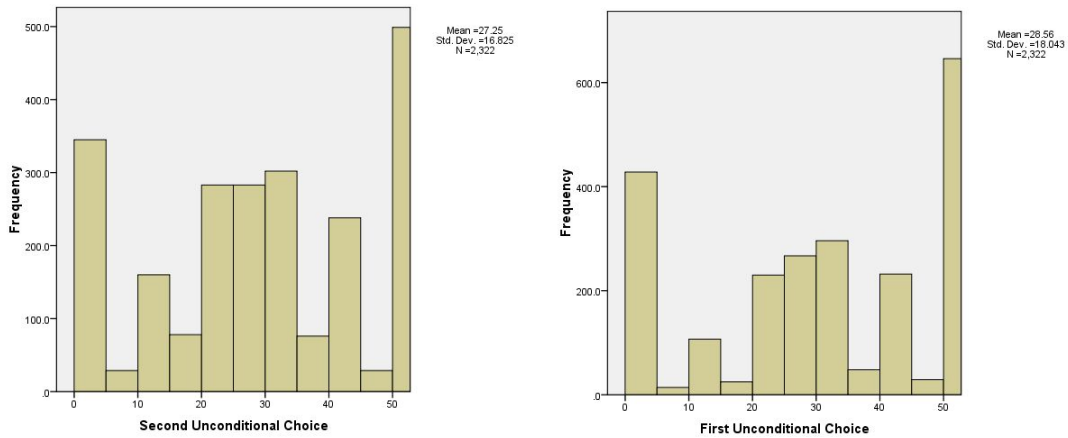
Table 3.10: Contribution Schedules for the first and second unconditional contribution

Figure 3.10 shows the contribution schedule of these subjects for the two uncondi-tional choices. In order to check whether there was a significant difference we had a significant difference for the unconditional choices we ran a two simple double sided t-test using as a split criterion the value of the online sociability index as defined in section 3.5.3

As can be seen in Table 3.11 the difference on the first unconditional choice is very small (almost 1%) while in the second unconditional choice. The difference is bigger (8%) providing that there is a case for a significance for the difference in the second case. By running the t-test on an alpha level of a=0.05 (95% confidence interval) we obtained the results seen in Table 3.12

Group Statistics					
	Online Sociability Value	**Number of Subjects**	**Mean**	**Std. Deviation**	**Std. Error Mean**
First Unconditional Choice	≥ 1	843	28.81	18.209	.627
	< 1	1479	28.43	17.953	.467
Second Unconditional Choice	≥ 1	843	27.77	16.944	.584
	< 1	1479	26.96	16.756	.436

Table 3.11: Distribution of our cases into groups for the unconditional choices for treatments 1 and 2

	t value	df	Sig. (2-tailed)	Mean Difference
First Unconditional Choice	.488	1730.852	.625	.382
Second Unconditional Choice	1.110	1735.120	.267	.808

Table 3.12: Result of the two tailed t-test for comparing the means between the two defined groups

Interestingly enough the results of the t-test give an interesting finding. While the difference between the mean contribution for the first unconditional choice is not significant (with p=.625 we cannot reject the null hypothesis at 95% confidence interval) the second unconditional choice has a significant different mean contribution value between the two groups with those that the online sociability index has a high value, tend to contribute significantly more that those that don't (with p=.267 we can reject the null hypothesis at 95% confidence interval).

In order to measure whether the contribution of the subjects had to do with their personal estimations whether the other members of the randomly created group had actually contributed the same with them, the subjects were asked to state how much they believed that the other members of the group contributed.
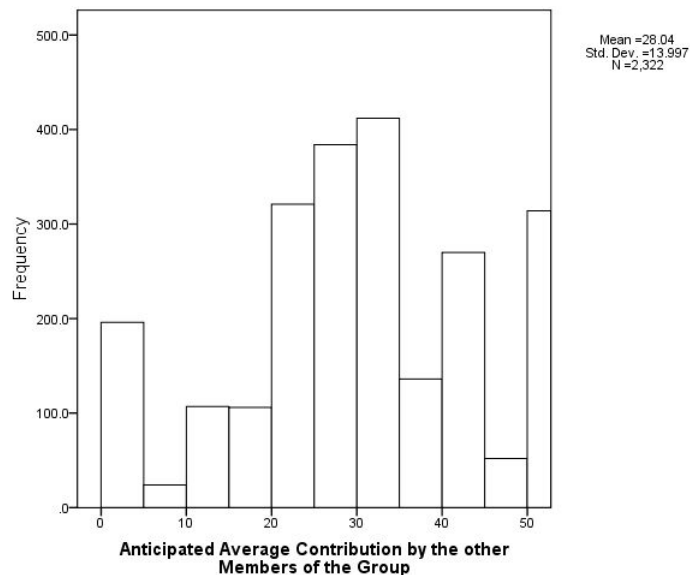
Figure 3.11: Anticipated average contribution stated by the subject for the other members of the group.

Figure 3.11 provides an overview of the anticipated contribution schedules for the other members of the randomly created group by the individual. We were interested to check again whether the anticipated effort had any relation with the online sociability. We ran the same statistical procedure again. The following table displays the split of the sample to the two groups.

147

Group Statistics					
	onlinesociability	N	Mean	Std. Deviation	Std. Error Mean
Anticipated contribution	≥ 1.00	80	29.22	14.419	1.612
	< 1.00	2242	28.00	13.983	.295

Table 3.13: Result of the two tailed t-test for comparing the means between the two defined groups

Once more we have a difference between the anticipated contribution for those that have high online sociability presence and for those that not (4%). To test for significance the t-test gives as a p value of p=.0456 where we can marginally reject the null hypothesis at 95% confidence interval. The case bring as back to the significance we had with the second unconditional choice where we can see that the effect of anticipated interaction in the group might have an effect on the anticipated contribution effort by the group members (1.2 units more).

	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Anticipated contribution	.749	84.388	.456	1.228	1.639

Table 3.14: Grouping on Minimal and maximal contribution level in relation with their online social network activity

Contributing to the Community when the others also contribute

Having obtained the result that the anticipated contribution by the other members of the group slightly affects the contribution of the individual subject we were interested to check whether, stated the average of the contribution of the other members of the group at certain levels of contribution, the contribution of the individual would also be

affected. Figure 3.12 shows the conditional contribution schedules for all the subjects
that have online sociability index greater than zero.
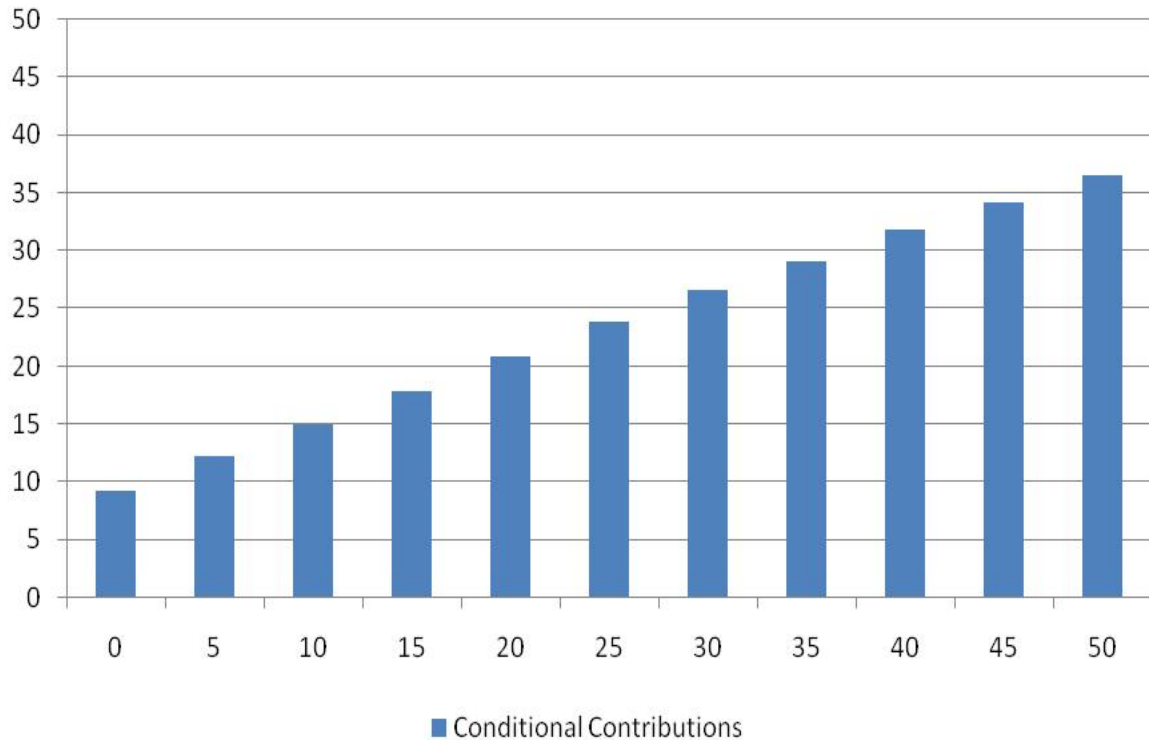
Figure 3.12: Average Conditional Contribution Schedules for the conditional contribution of
the subjects in the experiment for treatment 1 and 2 (Give Framing

From the figure we can observe a clear cooperative contribution behavior where
subjects do respond on the group contribution. Theoretically this can be explained
by the collective effort model where the subjects predominantly judge their individual
contribution to the group outcome [Karau and Williams, 2001]. The case that the rela-
tion of the conditional contributions on a five step interval over the budget endowed to
each subject is perfectly linear as can be seen in the figure also suggests that subject
do respond on their anticipated contribution confirming the tension that we found in
the previous section.

In order to test whether sociability plays a role on the conditional contributions
we followed the same statistical procedure that time on the minimum and maximum
conditional contribution efforts (conditional choice given on a minimal and a maximum

level of contribution). The grouping of the variables is the same as in the case of table 8.

Group Statistics					
	socialnetworks	N	Mean	Std. Deviation	Std. Error Mean
conditionalchoice0	≥ 1	843	8.49	17.096	.589
	< 1	1479	9.65	17.846	.464
conditionalchoice10	≥ 1	843	36.91	19.784	.681
	< 1	1479	36.30	20.056	.522

Table 3.15: Result of the two tailed t-test for comparing the means between the two defined groups.

In that case again the results confirm with a stronger p-value this time that contribution is in fact controlled by the group outcome. Our results have a stronger statistical significance with a p value p=0.126 and p=0.482 holding in both cases for a confidence interval of 95%. The significance obtained from the analysis is also secured by the fact that the F test for equality of variances among the groups also gives a significant value providing that in both groups the spread of the contributions according to the interval scale measuring online sociability also holds for our results.

Taking from the Community: The case of free riding

Considering the case whether participation in a community gives the incentive for free riding we want to see how the values in the contribution schedules might be affected by the framing that they have been positioned. In other words we are interested for treatment effects between the two treatments (Treatment 1 and Treatment 3) that might be attributed in the framing. For this purpose we made a comparison of the contribution schedules for the subjects in treatment 1 and 3.

In order to evaluate this case first we compared their unconditional choices in the

Variable Case	t	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Conditional contribution on Minimal Group Effort (Non- Equal variances assumed)	-1.533	.126	-1.163	.759
Conditional contribution on Maximum Group Effort (Non- Equal variances assumed)	.703	.482	.606	.861

Table 3.16: Result of the two tailed t-test for comparing the means between the two defined groups for the GIVE and TAKE treatment.

Variable	Treatment	N	Mean	Std. Deviation
First Unconditional Choice	Treatment 1	1524	34.92	14.725
	Treatment 3	677	14.34	17.310
Second Unconditional Choice	Treatment 1	1524	31.91	14.812
	Treatment 3	677	17.28	17.357

Table 3.17: Groupings for the two treatments (GIVE and TAKE)

Variable	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
First Unconditional Choice [2]	28.629	2199	.000	20.582	.719
Second Unconditional Choice[3]	20.257	2199	.000	14.632	.722

Table 3.18: Result of the two tailed t-test for comparing the means between the two defined groups for the GIVE and TAKE framing

first and third treatment. As can be seen in the table the difference was statistically significant in both unconditional choices with the average difference of the means to be less in the second unconditional choice. We proceed further to examine whether a treatment effect is also present in the conditional contribution schedules. For this reason we test for treatment effects in the minimum and maximum conditional contribution.

Group Statistics						
	Treatment	N	Mean	Std. Deviation	Std. Error Mean	
Conditional Decision on Minimal Group Effort (Equal variances assumed)	Treatment 1 (GIVE framing)	1524	5.33	12.968	.332	
	Treatment 3 (TAKE framing)	677	18.72	23.046	.886	
Conditional contribution on Maximal Group Effort (Equal variances assumed)	Treatment 1 (GIVE framing)	1524	34.22	20.837	.534	
	Treatment 3 (TAKE framing)	677	42.06	16.488	.634	

Table 3.19: Grouping for the Treatment effects on Minimal and Maximal Group effort

The results are again highly significant with subjects contributing more than they

	t	Df	Sig.(2-tailed)	Mean Difference	Std. Error Difference
Conditional Decision on Minimal Group Effort (Equal variances assumed)	-17.333	2199	.000	-13.390	.772
Conditional contribution on Maximal Group Effort (Equal variances assumed)	-8.657	2199	.000	-7.838	.905

Table 3.20: Two tailed t-test for the comparison between minimal and maximal conditional group contribution effort

get on both contribution levels.

3.6 Discussion and results

The analysis of the results in the previous section has provided some interesting results which can lead to equally interesting conclusions as to the relation between online sociability and cooperation. The results are important if we consider (a) the anonymous setting and the strict adherence to the experimental protocol where interpersonal biases have been omitted, (b) the framing of the participation in the community as a public good where all can contribute and all can benefit from it, and (c) the relation with the participation characteristics and other controls of online behavior.

The analysis that was presented in this chapter has yielded three interesting findings.

- The first finding is the case that participants with a high degree of online socia-

bility seem to be more altruistic than those that do not expose their presence on an equal grade. This might explain some of the characteristics of lurking as a negative factor in online communities [Rafaeli et al., 2004, Preece et al., 2004]. In particular, altruistic behavior has been studied in the context of pure economic behavior [Fehr and Fischbacher, 2003]. In particular the comparison between the two treatments (GIVE and TAKE framing) yielded that the participants tend to contribute more than they actually get. Maximum contribution was significantly different than the maximum benefit (p=0.000 for 95% confidence level). The same result was obtained as well for the minimum contribution and minimum benefit (p=0.000 for 95% confidence level).

- " The second finding relates to the behavior of the subjects with regard to the estimated behavior of the other group members. In fact, the estimated effort level for both categories of framing provides that a reciprocal behavior is founded on the basis of the group returns. As mentioned earlier, this confirms the expected theoretical assumption from social psychology models such as the collective effort model. As the subjects become more aware of the group's outcome and compare it with their own effort/benefit, they adjust their behavior accordingly. A contribution, however, of this study is to actually check whether the denoted online social activity actually explains some of this behavior controlling for the estimated group benefit. As can be seen by the analysis in Section 3.5.3 conditional cooperation on the maximum level is less significant for the conditional cooperation on the minimum level when online sociability is taken into account on the GIVE treatment (p value on minimal cooperation been more significant than that of the higher).This might be attributed to the fact that some of the subjects consider their personal benefits as being less important when they have an increased sense of communication (as it can happen in an online community) where the sense of collective action is better contextualized by the subjects themselves.

- The third finding has to do with the difference in volume between the first and second unconditional choice, both for the GIVE treatment and the TAKE treatment. Although the subjects seem to adhere in a code of group, in the beginning cooperativeness starts to fall after the first unconditional choice and shapes a linear relation in both framings.

All of the three findings hold the initial hypothesis that indeed online sociability is a factor that relates with cooperation. However, cooperation and behavior in public goods cannot be contextualized only under online communication since it is affected by many individual and group level factors [Andreoni et al., 2003].

3.7   Conclusion and further remarks

This chapter elaborated on social dilemmas of cooperation in relation to contributions in online communities, as well as the relationship to socioeconomic variables. In particular, the main objective was to study the relation of the online sociability characteristics of the participants with their intention to contribute in the context of an online community. The findings show that there are symmetric effects between online participation and cooperation. This suggests that the more someone participates in an online community, the more cooperative he/she will become. In terms of the actual implication of this finding in relation to the quality of the public good in an online community, this suggests that more participation can lead to higher contribution rates.

From an online community theory point of view, we can see a connection with the common identity and bond theory advocated by Ren et al. [2007] where participation and attachment in the community is sustained by the constraints and opportunities available. In other words, the more the participants communicate, the more cooperative they will become due to their attachment to the online community and its protocols (bond theory).

A drawback of the experimental procedure that we followed is that there was no opportunity for repeated interaction after the signaling of the effort level from other group members. This might have also given a clearer relation to the conditional co-operation for both minimum and maximum levels. However, the interpretation of the effort levels (how much this member has contributed to the online community) is very much dependent on the other individuals' perception of what constitutes high effort. Furthermore, the information regarding the effort that an individual puts into an online community and the benefit that he/she receives is not observable in all circumstances.

The online sociability measure that we used on this study was framed on explicit characteristics (volume of email, participation in online social networks and communi-cation activities) . However a more clear set of questions towards the subjects related with their intention to participate in an online community could provide a more accu-rate range of online sociability expressions.

Bibliography

J. Andreoni, W. Harbaugh, and L. Vesterlund. The carrot or the stick: Rewards, punishments, and cooperation. *THE AMERICAN ECONOMIC REVIEW*, 2003.

J.N. Cummings, B. Butler, and R. Kraut. The quality of online social relationships. *Communications of the ACM*, 45(7):103–108, 2002.

J. Donath. Signals in Social Supernets. *Journal of Computer-Mediated Communication*, 13(1):231–251, 2007.

C. C. Eckel and P. J. Grossman. Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal*, 108:726–735, 1998.

E. Fehr and U. Fischbacher. The nature of human altruism. *Nature*, 425:785–791, 2003.

E. Fehr and S. Gachter. Cooperation and punishment in public goods experiments. *The American Economic Review*, 90:980–994, 2000.

C. M. Johnson. A survey of current research on online communities of practice. *The Internet and Higher Education*, 4:45–60, 2001.

S. J. Karau and K. D. Williams. Understanding individual motivation in groups: The collective effort model. *Groups at work: Theory and research*, pages 113–141, 2001.

A. Kavanaugh, J.M. Carroll, M.B. Rosson, T.T. Zin, and D.D. Reese. Community Networks: Where Offline Communities Meet Online. *Journal of Computer-Mediated Communication*, 10(4), 2005.

P. Kollock. *The economies of online cooperation*. Routledge, London UK, 1999.

C.A.C. Lampe, N. Ellison, and C. Steinfield. A familiar face (book): profile elements as signals in an online social network. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 435–444. ACM Press New York, NY, USA, 2007.

J. Lerner and J. Tirole. Some Simple Economics of Open Source. *Journal of Industrial Economics*, 50(2):197–234, 2002.

G. Marwell and P. Oliver. *The Critical Mass in Collective Action: A Micro-social Theory*. Cambridge University Press, 1993.

J. Preece, B. Nonnecke, and D. Andrews. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20:201–223, 2004.

R. D. Putnam. Bowling alone: America's declining social capital. *Journal of Democracy*, 6:65–65, 1995.

S. Rafaeli, G. Ravid, and V. Soroka. De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, page 10, 2004.

Y. Ren, R. Kraut, and S. Kiesler. Applying Common Identity and Bond Theory to Design of Online Communities. *Organization Studies*, 28(3):377, 2007.

P.A. Samuelson. The Pure Theory of Public Expenditure. *Readings in Social Welfare: Theory and Policy*, 2000.

T. Schoberth, J. Preece, and A. Heinzl. Online communities: a longitudinal analysis of communication activities. *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, page 10, 2003.

L. Sproull, W. Dutton, and S. Kiesler. Introduction to the Special Issue: Online Communities. *Organization Studies*, 28(3):277, 2007.

B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythornthwaite. Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Reviews in Sociology*, 22:213–238, 1996.

B. Wellman, AQ Haase, J. Witte, and K. Hampton. Does the internet increase, decrease, or supplement social capital?: Social networks, participation, and community commitment. *American Behavioral Scientist*, 45:436, 2001.

S. Whittaker, L. Terveen, W. Hill, and L. Cherny. The dynamics of mass interaction. *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 257–264, 1998.

Effort, Benefits and Commitment on Online Knowledge Communities: An empirical study on Yahoo!Answers

Yahoo!Answers operates a question-answering community of users who post questions and receive answers on various topics. In this chapter we examine the effect of activity on service posture (measured by volume and time) as expressed by user contributed effort and user received benefit in an online community facilitated by members of the Yahoo!Answers community. By programming a web crawler to store a random sample of questions posted over a period of one month, we used a set of time series, random effects and logistic regression models which confirm, to a large extent, our formed hypotheses. Findings indicate that the ratio of contributed effort and received benefit has an effect on service posture (volume and time). In particular, we find that users who contribute more effort in the community than received benefit get a question answered by more users in less time than users who receive more benefit than the effort they contributed to the community.

4.1 Introduction

The previous chapter presented a controlled study of the characteristics of online community participants from the perspective of social dilemmas. The dilemma that was presented to the participants was related to cooperation. If all participants cooperate, then the outcome for the group will be optimal. We have seen from our analysis that continuous communication as expressed by online sociability characteristics (e.g., the participation on online social networks) leads to more cooperation (as measured by the public goods game). However, what would have been the outcome of the public goods game if the participants had been able to observe the previous behavior of each individual member of the group? Would they cooperate in that case or defect, as a way of sanctioning the non-cooperative individuals?

From this perspective, we present the above case in the context of online social mechanism for question answering deployed by Yahoo, known as Yahoo!Answers, which provides a new paradigm on how important information can be retrieved from the web by relying not on information artifacts (e.g., web pages) but on the users. In fact, the use of the Internet, and particularly the World Wide Web, for information seeking has been the innermost driving force behind the development of search engines and information retrieval mechanisms [Baeza-Yates and Ribeiro-Neto, 1999]. Nonetheless, information retrieval mechanisms that rely on user's participation are somewhat difficult to establish due to the well known problem of information scaling exponentially in relation to the number of users that become acquainted with it. Therefore, in traditional knowledge or information retrieval mechanisms, automated methods and algorithms are used for the provision of complex knowledge. In particular, advances on information retrieval have explored the development of automated question answering systems based on similarity measures against a corpus of already defined pairs of questions-answers [Kwok et al., 2001a]. This type of predefined sets of information seeking and matching has been advanced lately in the web in order to provide the users of search engines a usable corpus of well defined knowledge

161

that can be reused to fulfill the needs of users with similar information needs [Dumais et al., 2002].

Nonetheless, apart from the technical value of these systems, what becomes important is the engagement of a large group of users to an online social activity that involves exchange of information through technology supported pathways. This, in fact, constitutes an online community as defined by Preece [2000] where the element of the "shared purpose" is materialized in the context of information exchange. But can information exchange itself provide solid ground for theorizing information sharing websites such as Yahoo!Answers an online community? A great amount of criticism comes from the perspective of content quality, where invaluable content such as advertisements and links floods most of these online information sharing venues. However, if we adopt a social interaction perspective on the way information is retrieved from the web, we tend to unravel an indirect social interaction between information encoded by people (e.g., in the form of web pages) and retrieved/read by other people through a communication channel that is provided by the web and the search engines. In fact, it can be argued that information has a social dimension which results out of interaction between different factors that affect its source and as a result which tends to make information important to the users [Brown and Duguid, 2002]. The newspaper articles, for example, often report factual cases embracing 'whiteness' and authoritative accounts in order to increase the credibility of their sources. However, in the web, information is often provided by social entities (individual authors of web pages) that have no proven credibility since the problem of evaluating the trustworthiness of a piece of information on the web is something that is very challenging due to the very much distributed nature of the web as a communication medium [Chakrabarti, 2003]. To overcome the issue of trustworthiness, several solutions already applicable to modern search engines have been developed [Kleinberg, 1999, Brin and Page, 1998] which encapsulate basic principles from the graph based architecture of the WWW. Applications of these algorithms can be applied, for example, in cases where

users tend to seek information that is relevant to a high number of web pages that are available online where the search engine's results index might lead to a case of information overload [Edmunds and Morris, 2000]. On the other hand, there might be a case where the query supplied by the user is very specific, and not enough information is available online for an interested individual to exploit.

This is where question-answering architectures come to play their role. The fundamental principle of question answering systems is to facilitate retrieval of information around a specific topic which can be seen as a result to a question. However, revisiting the social dimension of the information available on the web and taking into account that for an answer to exist, there needs to be a question, and vice versa, question answering systems tend to provide a solution where modern search engines do not cover the information needs of the users [Kwok et al., 2001b]. Therefore, in order to make valuable web content more widely available and to satisfy special characteristics of users needs, mechanisms and platforms that facilitate question-answers have been developed by the major players in the search engine industry (Google[1] and Yahoo!).

Nonetheless behind the limitations of the retrieval of codified knowledge there is a social process that provides the driving force for social communication and activity over information seeking on the web and not just pure communication purposes. As aforementioned, Communities of Practice [Wenger, 1998, Preece, 2000] represent a case where social interaction is centered on the exchange of ideas with knowledge exchange becoming the final goal [Faraj and Wasko, 2001]. Following that direction, a question still comes to challenge the potential of online communities and the wider adoption of their merits. Why should someone waste time to provide an answer to someone else's question, since from a rational point of view he/she receives no actual compensation for doing that?

This issue is, in fact, the next step of the study discussed in Chapter 3. As aforementioned, the essential social dilemma in an online community is whether to partic-

---

[1]Google's Answer service was shut down on December 2006.

ipate, that is cooperate with other users on the purpose of the community, or defect by becoming either a lurker (observer) or free rider where someone gets benefited without contributing any effort. Although this dilemma is a bipolar case (you either cooperate / participate or not), in reality this can also be observed where a participant gets benefited a lot by not participating equally. As we discussed in the case of the public goods game, someone benefits the most if he/she contributes the least portion of amount that is available to him or her. The optimal is to contribute zero but a player can also contribute an amount that represents 2 or 5 of the total endowed amount. Furthermore, decisions on the public goods game, as it was configured, were one-shot, providing that the profiles of the players were formed immediately. However, when we have a repeated interaction setting, can we guess what the cooperation outcome will be for those that provide less effort and receive a greater benefit?

This chapter moves towards this direction. In particular, we want to examine the effect of social incentives on the activity of a very large online question -answering community that is operated by Yahoo! Answers service in a longitudinal perspective. The standard operational model of this community is that from the beginning of the operations of the service a vast number of individuals has started participating in the service, posting questions and receiving answers by other individuals. This "rich" social world is mapped into direct social interactions by the question of where a service user (an asker) might receive more than one answer by other users, and where at the end he has to evaluate the best as an acceptable answer in order to close the question thread. Furthermore, Yahoo! provides a classification of the best / most active questions for all categories. By programming a web crawler to randomly pick up the best provided questions (in order to avoid not interesting and low quality questions) from the Yahoo!Answers service and inserting them in a relational database, we were able to capture a vast number of questions and answers in our dataset, which we further processed in order to address the following questions:

- Do participants with continuous presence receive many more responses than

participants that just started using the service? The perspective of this research question is related to the longitudinal aspect of interaction. In relation to the results of the previous chapter, we are looking to evaluate whether repeated interaction leads to increased cooperation.

- Do participants who posted an answer to a question receive answers to the question that they posted on the system after the answer they provided? This research question examines the case of the social dilemma related to whether the ratio between the answers provided and questions posted affects the benefit they receive from the community (as measured by the amount of answers received by the participants).

- Do users get an account of the number of times a user who posted questions has provided answers to the community? This research question in essence combines the above two aspects of continuous presence and benefit/effort ratio of the community participants.

To address the above research questions, we ran a panel data analysis based on interaction characteristics applied on the dataset that we collected from Yahoo!Answers online service. By using a set of mixed, random effects, within categories as well as population averaged econometric setting, we were able to model the effect of overall, during time and between categories activity. In order to capture the interaction between two or three individuals over a thread of a question, we modeled time dependent variables which were added in the model.

To this end, the study presented in this chapter is structured as follows. Section 4.2 provides a background on repeated interactions as a focal point for studying online communities and, in particular, the Yahoo!Answers community which is described in Section 4.3. Following that, Section 4.4 presents a description of the methods and the constructs that we used in our analysis along with the indicative results obtained from running our constructs on the collected dataset. Section 4.5 offers a discussion of the

results, and conclusions are given in section 4.6.

4.2 Benefit, effort and commitment on online communities

To address the research questions formed in the previous section of this chapter we theorize our analysis on three specific constructs: Benefit, Effort and Commitment. In the context of the Yahoo!Answers online community we define the concept of "effort" as the amount of answers a participant has *provided* to fellow participants while benefit is defined as the amount of answers a participant has *received* from fellow participants. In order to theorize the longitudinal perspective of the above pair of constructs we use the concept of commitment.

Commitment is a special case of social interaction where the two parties interacting repeatedly over a particular context. The later results to a case of repeated interaction over a setting between two or three individuals. Considering repeated interaction is important when there is a direct involvement over two parties in a particular scenario. In our case for example repeated interaction involves the transition between two opposite states: the state of an asker and the state of an answerer. Taking into account the modus operandi of the service (Figure 4.1) we can observe who was the user who posted the question (service recipient) and who was the user that provided the answer (server). Studies on online communities have diversified over the structural characteristics of interaction such as thread depth [Whittaker et al., 1998], author ranking analysis [Jurczyk and Agichtein, 2007], and visualization based analysis [Fisher and Dourish, 2004]. On the other hand communities diversify over different areas of subjects depending on the characteristics and the expertise of their members [Powazek, 2002] which has a significant effect on the resource output. In particular in cases where expertise between members of a community defers interaction can lead in an expert-apprentice relation which makes the relational ties between the users

stronger [Soden and Halliday, 2000, Oliver et al., 1998]. From that perspective It can be argued that the less broad is the expert-apprentice relation that the community tackles the less is the resource output. However value is provided by a community discussion mechanism (e.g. a newsgroup) where the scope of the information is much more specific than the information posted in general discussion lists.

This particular flow of information between users in online communities of practice results to the case of knowledge flows across and between users interested in a broad range of subjects [Walsham, 2001]. However since each individual user participates on a social interaction environment a set of direct and indirect comparisons takes place when a user asks for information by other users and in that case there is no particular benefit for the user who will make the effort to reply. In offline environments some users might have a fixed social preference towards being altruists [Fehr and Fischbacher, 2003], other users have a somewhat more reciprocal set of inertia that affects their activity as in a standard social setting where contribution is required [Fehr and Gachter, 2000]. In particular the later can be connected with Social Comparison Theory [Festinger, 1954]. According to social comparison theory individuals tend to evaluate their current status against images of others in order to decide upon their own actions. For example in our case an individual that posted a question and hasn't received an answer will have a tension to not provide an answer in case he is capable of since by comparing his case with the other individual his opinion will be towards being more competitive.

4.3 The Yahoo! Answers online service

Having provided our motivation for the constructs used in the empirical analysis of our dataset we proceed to describe the Yahoo!Answers and the mode that it operates. The Yahoo! Answers is an online service which basically structures the content of a web interaction around a controlled discussion in the form of a question. As can be

seen in Figure 4.2 a user posts a question on one of the multiple categories that the service provides. A set of other users reply to the question and the user is notified about the answers he/she has received through the online service. If the user finds an answer that according to his/her needs is a satisfactory answer for the question posted then this particular answer is selected as the best answer and the question is closed. On the other hand If the user receives multiple answers by the other users and a selection of the best answer is not made then the question remains open for other users to post answers until the user chooses the best answer and the question is closed (Figure 4.1). Therefore there is a difference from the time the user receives the best (according to his needs) answer and the first answer to the question he posted. A user can also reply or comment to other questions posted by other users. In order to make the interaction simpler Yahoo provides a set of other functions as well as for example the ability for a user to signal his/her presence with a visual avatar which is linked to a user profile page with the history of the activity of the user in the online service. By doing so, other users can visit the profile page of a particular user and see all the questions that this particular user posted or replied. The profile page acts as an implicit reputation index where users might consider on looking before they actually reply to a particular question by this user.

Yahoo!Answers data have been also used on other studies in the literature such as for example author reputation analysis [Su et al., 2007]. The contribution of this study is that it examines the service operation model on an interaction basis considering the function of the service as a public good where all users are benefited (there is obviously the characteristic of non-excludability which satisfies the condition for a public good).

In Yahoo!Answers interaction is structured in question threads. A thread is an interaction pattern that is formed over a network of interaction between three individuals.

However from the design of the Yahoo!Answers service as can be seen from Figure 4.2 the thread structure is evolving on a linear way which doesn't allow formation

168
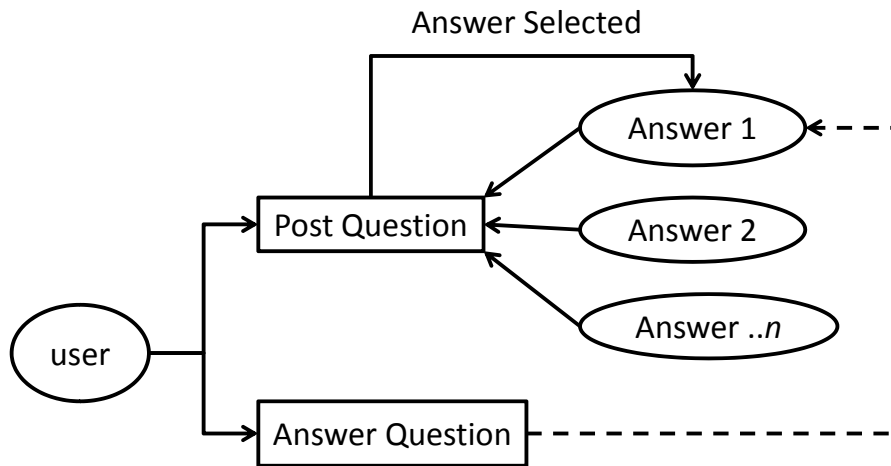
Figure 4.1: The Yahoo!Answers modus operandi

of sub-threads (threads within threads).

Having described the service model we proceed with the description of the dataset variables used in this study.

4.3.1   Dataset variables and description

The data was collected using a random selection of questions scattered around 23 categories. The dataset covers a period of one month. On average there were 6445 questions posted per day, 29767 answers provided daily by other users with users finding valuable 445 of them per date. As can be inferred from the numbers for each question posted on the website there are roughly 5 answers for this question. The acceptance ratio for the answers provided informs that around 1% of total answers per date are accepted by the users who posted questions.

The dataset contains standard variables which are:

- **Questionid:** The id of the question available on the Yahoo!Answers service. It is used for identification purposes.

- **TimePosted:** The timestamp that this particular question was posted in the

Figure 4.2: An example interface from the Yahoo!Answers online service

service.

- **Number of Answers:** The number of answers this question has received by the time it was registered in our dataset.

- **Userid:** The id of the user that posted this particular question

- **Answererid:** The id of the user who provided the selected answer to the particular question

- **Categoryid:** The category id that this question belongs to

- **Chosen Answer Timestamp:** The time that the accepted answer was posted on the service. The time actually represents the actual time that the answer was posted and not the time that this answer was accepted by the user who posted the question.

Table 4.1 provides a summary of the number of questions, number of answer, number of questions closed (an acceptable answer was given). Apart from that we also have the number of users that posted questions on that category and the average time (minutes) for a question to be closed.

Our dataset contains in total 148505 questions with 682725 answers generated and posted by 9421 users. The average time to receive an answer to a question posted on these categories was roughly 160 minutes (2.5 hours) with minimum time to answer be 25 minutes and maximum 787 minutes. As can be seen in Figure 2.1 a user receives several answers for a question and in a second step decides which answer is the most suitable for the question that he/she posted and then the question is closed. The total of the closed questions for all the categories in our dataset was 11448 and since from the operational model we know that for each closed question there is only one accepted answer then out of the 682725 only 11448 or 0.016768 of the answers were accepted. The very low percentage of the answers accepted

171

provides that the quality of the answers judged by those who asked a question varies from category to category.

Nevertheless much of this variation might depend on the concentration of users around the specific category. In order to measure the concentration of a user on a particular topic we used the Herfindahl-Hirschman Index (HHI) [Hirschman, 1964]. HHI is widely used on market analysis to calculate the market share of a particular firm in order to determine whether a market is an oligopoly or not. Since HHI captures concentration, we used the reverse (1-HHI) index as a metric of variation. The advantage of using the reversed HHI as a variation measure is the ability to capture variation within a subset of question/answers taking the total number of them available into account. From the design of the service we know that a user can execute three basic actions in relation with his/her peers: (a) Post a question (b) Provide an answer to a question (c) Select the best suitable answer to a question that he/she posted. We employ the variation index in order to measure variation in two levels: benefit and effort. We measure benefit variation by measuring the variation on the number of questions the particular user got answered. The formula that we employed in order to measure the benefit variation is:

$$BenefitVariation = (1 - HHI) = 1 - ([S_1^2 + S_2^2 + ... + S_n^2]/QuestionsCount)$$

Where S_i is the percentage of the questions [0....1] which are part of the category i and Questions Count is the total number of Questions in the Dataset. With the same procedure we measure the effort variation by employing the number of answers the particular user posted on the service.

$$EffortVariation = (1 - HHI) = 1 - ([S_1^2 + S_2^2 + ... + S_n^2]/AnswersCount)$$

Benefit variation and effort variation in fact explain the two states of a user in the online service. By examining the relation between those two indices we can see for

Category	Code	Number of Questions	Number of Answers	Number of Questions Closed	Avg. Time to answer a question (minutes)	Number of users
Family	1	7881	52470	42	25.2	31
Elections	2	4377	34428	185	33	123
Adolescent	3	4372	37260	144	37.2	123
Celebrities	4	8828	72071	11	37.5	10
Computer Networking	5	8703	19872	50	39	47
Skin Conditions	6	4910	20200	17	40	16
Home Schooling	7	10125	22505	170	43.2	147
Dogs	8	4402	28273	176	47.5	152
Fashion & Accessories	9	11221	68092	202	53.5	171
Other - Science	10	4288	11679	77	54.9	63
Programming & Design	11	824	1709	71	76.2	67
Other - Food & Drink	12	7180	45160	916	88.8	755
Psychology	13	4986	25423	600	102.8	495
Video & Online Games	14	5755	10769	66	103.3	62
Garden & Landscape	15	5132	17248	462	120.8	433
Corporations	16	10418	28955	387	131.8	351
Car Audio	17	10145	31148	583	148.5	523
Performing Arts	18	19905	84257	2522	176.7	1969
Current Events	19	3900	24635	1267	181.1	836
Computer Hardware	20	1320	3228	256	287.3	236
Environment	21	2903	18862	1416	378.3	1176
Geography	22	3129	18250	1147	665.2	1022
Small Businesses	23	3801	6231	681	787.1	613

Table 4.1: Summary Data for the Categories that we used in our dataset

example whether an individual is benefited by a general collective of the service while he/she provides effort for a small fraction of the needs of the online service scattered across the categories.
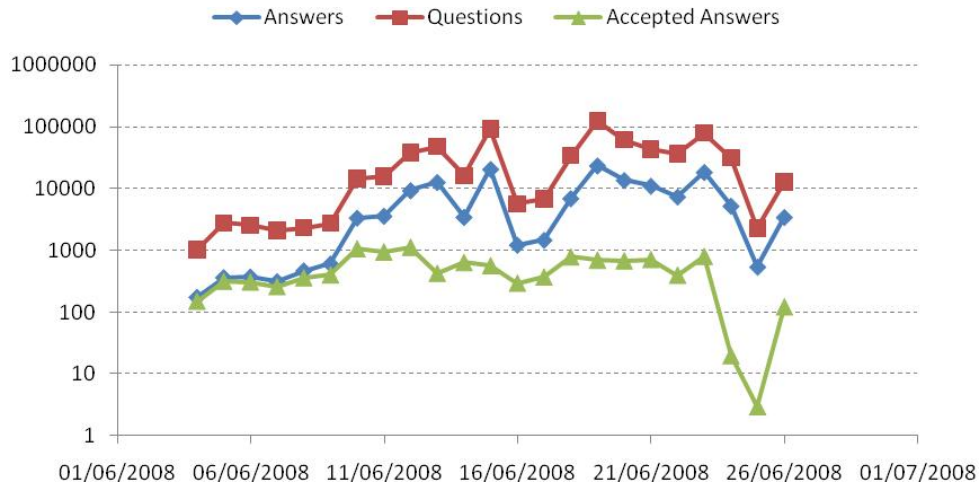
Figure 4.3: Number of Answers and Number of Questions Posted over the one month period on our dataset and the number of accepted answers / closed questions.

On Figure 4.3 we provide the volume traffic on questions posted and answers provided over the period of one month (June 2008) that the dataset was collected. Generally the variation of the number of answers follows the variation of the questions posted providing that the service is in a sustainable level where the amount of questions posted and answers provided slightly evolving over time. The same happens also for the number of questions getting accepted which follows a steady more linear variation over time (if we accept a possible outlier on 25/06).

4.3.2 Panel data variables

Apart from the aforementioned variables and in order to be able to analyze the interaction in the online service, we use a set of three categories of variables which form our theoretical framework. The first category is the overall effort and benefit of the user as measured by counting the total number of answers and questions that were posted by the individual. Then for each particular user we need to be able to track the

effort and benefit history over time. For a user "i" we define two states: u_i^p when the user posts a question and u_i^a when the user answers a question.

Using that formulation and the definition of variation from the previous section, we define the following set of variables:

- $Post_{(u_i^p)}$: The number of times the user i has posted a question

- $Post_{(u_i^a)}$: The number of times the user i has answered a question

- $Post_{(u_i^p)t}$: The number of times the user i has posted a question before the time t which is the system timestamp denoting the time the question has been posted.

- $Post_{(u_i^a)t}$: The number of times the user i has answered a question before the time t which is the system timestamp denoting the time the question has been posted.

- $BenefitVariation_{(u_i^p)}$: The variation of the benefit that the user has received by participating on the service

- $EffortVariation_{(u_i^a)}$: The variation of the effort that the user has contributed by participating on the service

- $AcceptedAnswers_{(u_i^p)}$: The number of questions that have been closed by the user where an accepted answer has been provided by another user.

The $Post_{(u_i^p)t}$ and $Post_{(u_i^a)t}$ represent the longitudinal effort and benefit indices for the particular user "i" while the benefit and effort variation variables measure the distributed effort and benefit provided and received by the particular user. The basic intuition for insisting on introducing variation variables was to avoid the case where repeated interaction among users evolves into a stronger relationship, which is out of the scope of this study. Variability among topics and subjects also further secures the case of anonymity where the true incentive for putting effort on the system takes the behavioral characteristics.

4.3.3 Descriptive statistics

Having defined our dataset variables, we proceed with the calculation of the basic descriptive statistics. For this particular purpose we utilized standard SQL queries using the statistical functions of the relational database system used, over the collected dataset. As can be seen from Table 4.2 on average a user posts roughly two questions and provides four answers to other questions.

Variable Name	Mean	St.Dev	Minimum	Maximum
Time to receive answer(chosen answer)	260	708	0.4	10337
Number of Answers	4.61	5.83	0	338
$Post_{(u_i^p)}$	2.42	3.78	1	69
$Post_{(u_i^a)}$	3.48	10.47	0	419
$Post_{(u_i^p)t}$	0.71	2.27	0	68
$Post_{(u_i^a)t}$	2.37	7.48	0	356
$HHI_{(u_i^p)t}$	0.79	2.88	1	0.997
$HHI_{(u_i^a)t}$	0.68	2.59	0	0.991
$AcceptedAnswers(u_i^p)t)$	2.05	2.48	1	53

Table 4.2: Descriptive Statistics for the variables in our dataset. N = 133694, N-chosen answer: 11448

From the average of about 5 answers that a user receives for each question he/she accepts about one. On the other hand users have some significant background activity on providing answers to other questions. On some extreme cases users who get on average about 5 answers have posted up to 356 answers to a set of other questions. On the other hand the past benefit for a user that posts a question was 1.2 questions with a minimum and maximum range from 0 to 9. Then again this particular user provided on average with a minimum and maximum range from 0 to 98 and a higher standard deviation. Whether an answer was
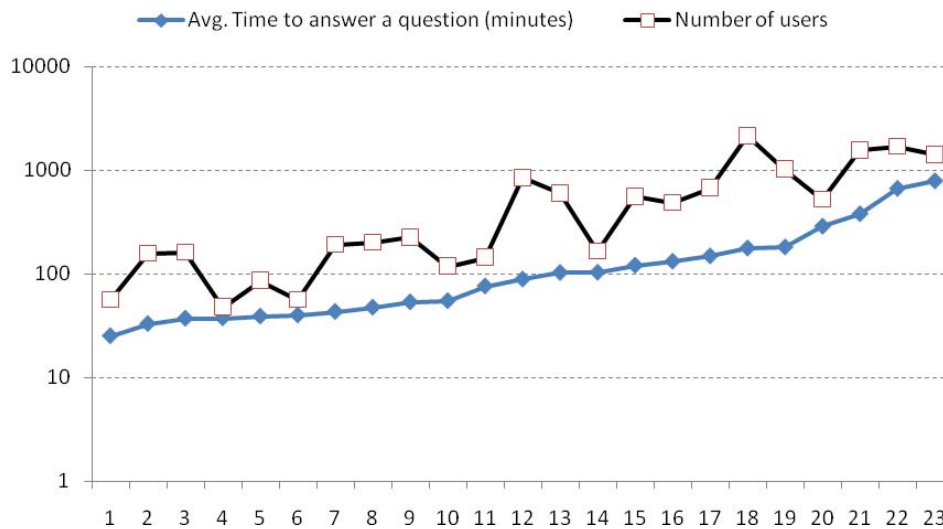
Figure 4.4: Relation between time to answer (minutes) and the Number of Users (Axis y is on logarithmic scale)

In Figure 4.4 we provide a relation between the number of users and the average time to answer (minutes) for the 24 categories in our dataset on a logarithmic scale. Although there is a slight linear variation on the number of users, the time required follows an almost perfect linear variation across categories which initially provide that increase in the number of users for a category requires more time for an answer requires more time for a question to be answered.

As can be seen from Figure 4.5, the relation between the number of answers provided and the fraction of this answers been accepted seems also to develop in a linear way as well over the 23 different categories in the dataset. Interesting is the difference between category 1 and category 23 where the gap of the accepted answers and the total answers provided is smaller. That provides that a high percentage of the population of the posted answers (the effort provided by the users) was evaluated as very good by the other users who posted. On the contrary on category 1 the quality of the answers provided is lower as can be indicated by the number of the accepted answers. Since the graph is displayed on logarithmic scale we can avoid cases where low volume of questions can lead to low volume of answers.

177
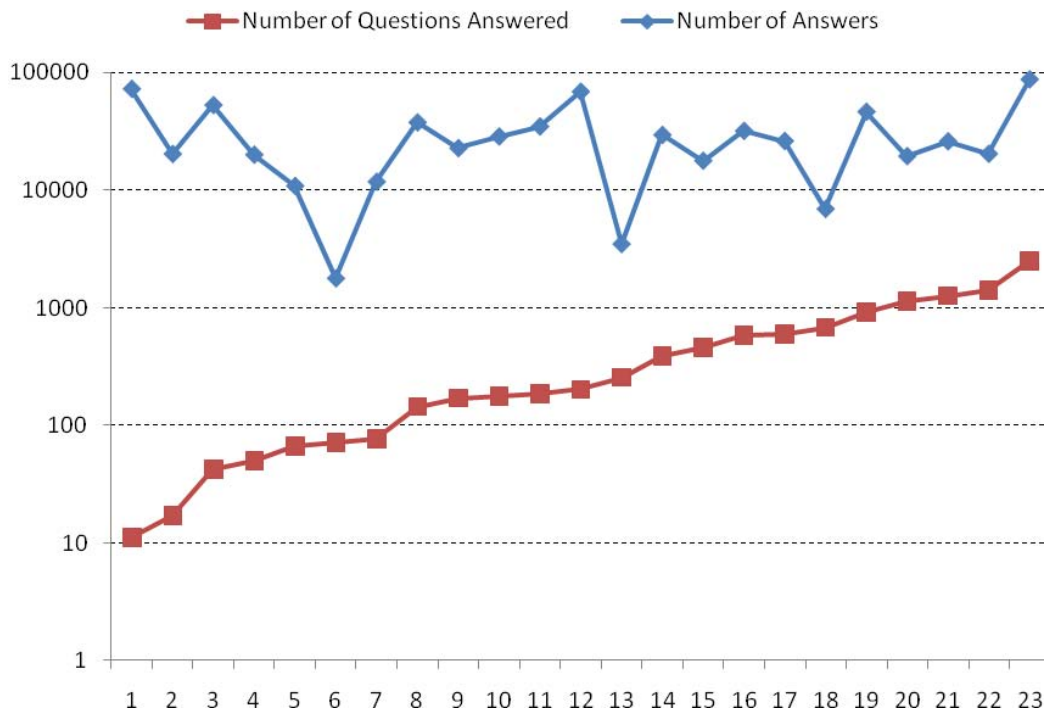
Figure 4.5: Number of Questions Answered (Number of Answers Accepted) with the total number of answers posted on the online service (Axis y is on logarithmic scale) for the period studied.

4.4 Methods and constructs

4.4.1 Constructs summary

Having defined our variables we consider the model provided in Figure 4.6. Server Posture is defined by the volume of answers received and time taken to answer this particular question.

Our model consists of three indicative components namely the Current Status, the reputation characteristics and the commitment to the community in terms of variation. These three components in return have an effect on Server Posture which as aforementioned defined in two terms: number of answers a question has received and (when an answer selected) the time taken for the user to get an answer accepted.

Current Status encompasses the accumulation of the activity that a member has in an online community setting which encompasses both the effort contributed and the

178
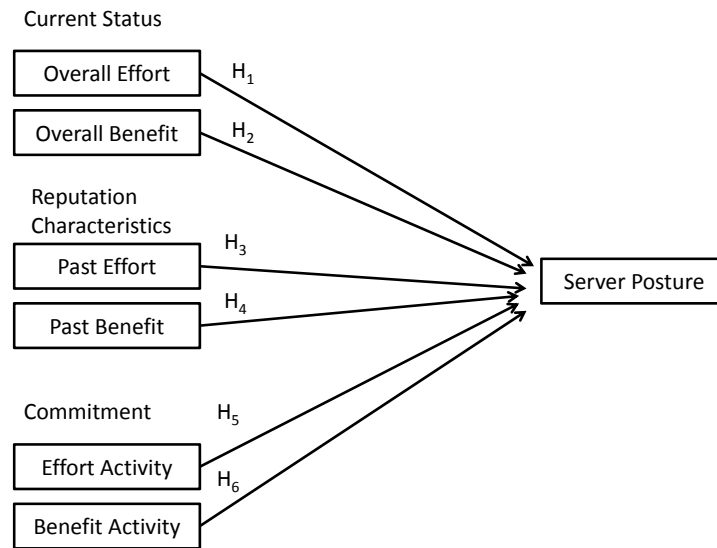
Figure 4.6: Service Benefit, Contribution Effort and Commitment

benefit received. Status is an important element in an online community because it encapsulates the current level of interaction that a user has articulated [de Souza and Preece, 2004, Wang and Fesenmaier, 2003]. Since Yahoo!answers provides a point based status index on the web site a user is able to see how many questions has the user posted on the website and how many answers he/she has received. It is expected that users will react positively on the overall effort that a user has contributed to the community (in terms of posted answers) and negatively to the overall benefit that the user has received from the community. Thus we have the following hypotheses:

- *H1: Overall Effort will have a positive effect on Server Posture*

- *H2: Overall Benefit will have a negative effect on Server Posture*

However activity is a time dependent factor which very much depends on prior history of the user who posted a particular question. Therefore in order to control for past effort and past benefit we use the $Post_{(u_i^p)t}$ and $Post_{(u_i^a)t}$ panel data variables described in section 4.3.2. Thus our additional hypotheses are developed as follows:

- *H3: Past Effort will have a positive effect on Server Posture*

179

- *H4: Past Benefit will have a negative effect on Server Posture*

Additionally a user might post to several categories of questions and his/her activity on one specific category might be a misleading factor for understanding activity as a whole. Since Yahoo!Answers diversifies in several categories and user's interests as well can be only in one or two categories, we measure a user's commitment as an instance of diversification of activities. This in turn can be operationalized in terms of variation by using the $HHI_{(u_i^p)t}$ and $HHI_{(u_i^a)t}$ panel data variables. Therefore we complement our hypotheses as follows:

- *H5: Effort Activity will have positive effect on Server Posture*

- *H6: Benefit Activity will have a negative effect on Server Posture*

As aforementioned the panel variables that we generated refer to the two states on whose relationship we are interested to examine (effort and benefit). In order to examine some possible relation we start with examining autocorrelation coefficients between the variables in our dataset. Table 4.3 shows the autocorrelation matrix for the variables on our dataset. The significance levels were adjusted using the Bonferoni procedure to assert that the significance levels were robust and hold for cases such as random effects.

The coefficients from the autocorrelation provide some interesting information that we are going to use further for our analysis. It is interesting to note that all the factors represented on our model affect negatively the time to get the best answer for a question posted. Since the server posture is defined as the opposite of the time to answer the service we can safely extract that all factors affect positively the server posture in the platform. The higher coefficient was obtained by the variation between the numbers of answers that the particular user, who posted the question, provided to other questions opened in the platform. In particular an increase in the variation of the answers provided by the user by one unit provides a 10% more effective response to the questions that this particular user posted.

The other indicative variable which encapsulates server posture is the number of answers the particular question has received. In particular we can see from the coefficient sign the more a particular user posts questions across categories (Question Variation) the less is the number of answers that he/she receives as a reply. On the contrary the more the user provides answers to other questions both in the particular category and across categories the more likely is to get a higher number of replies/answers to his specific question. The correlation coefficients for the variables that describe the benefit provision to others $Post_{(u_i^a)}$, $Post_{(u_i^a)t}$, $HHI_{(u_i^a)}$
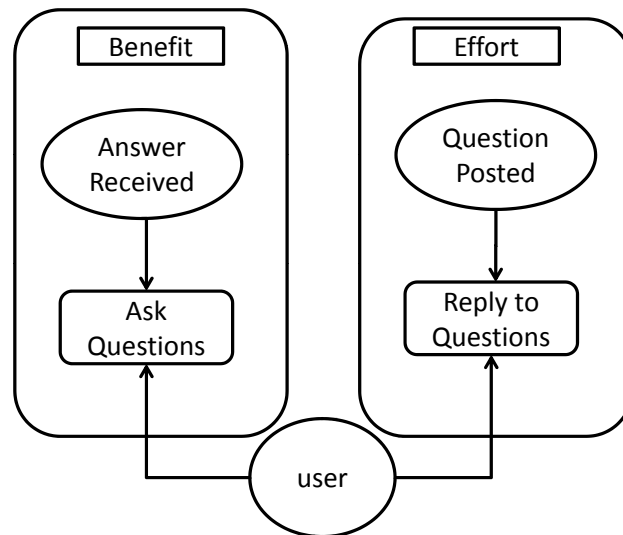
Figure 4.7: Benefit-Effort elicitation for the user states

Figure 4.7 provides an outline of the effort - benefit model that is used on our study. Benefit is measured as the provision of knowledge to the user by other users on the community in the form of answering/replying to the questions that the user posted. On the other hand a user is giving back to the community by replying to the questions that other users posted across the different categories. The amount of the other questions that the user has answered represents the effort that the user contributed to the community. To compute the benefit provision and effort contribution to the community we used a two way approach. The first was to evaluate the total number of answers that the user has provided to other questions posted across categories.

181

	1	2	3	4	5	6	7	8	9	10
1. Time to Answer	1.00									
2. # of Answers	-0.100***	1.00								
3. # of Comments already Given	-0.022***	0.021***	1.00							
4. Post(u_i^p)	-0.04***	0.142***	0.045***	1.00						
5. Post(u_i^a)	-0.055***	0.174***	0.068***	0.467***	1.00					
6. Post$(u_i^p)t$	-0.042***	0.117***	0.022***	0.832***	0.37***	1.00				
7. Post$(u_i^a)t$	-0.058***	0.194***	0.048***	0.379***	0.722***	0.42***	1.00			
8. HHI(u_i^p)	-0.032***	-0.084***	-0.01***	-0.071***	-0.041***	-0.059***	-0.046***	1.00		
9. HHI(u_i^a)	-0.100***	0.164***	0.023***	0.181***	0.257***	0.151***	0.28***	-0.062***	1.00	
10. Accepted Answers(u_i^a)	0.016***	0.116***	-0.013***	0.784***	0.308***	0.685***	0.273***	-0.097***	0.177***	1.00

Table 4.3: Pearson autocorrelation matrix for the variables in our dataset.

The second was to evaluate the total number of questions that the user has posted and in consequence the number of answers that he/she received in reply. To control for cases of time dependent activity when examining each individual case we computed the time based variables to see at the point that the question was posted by a user, how much was the amount of the questions posted and the amount of answers he/she received at that specific time.

4.4.2 Estimation results

A set of time series, random effects and logistic regression models was used in order to estimate the factor that affect the number of answers a user receives after he posts a question to the community. On all three models we include a time period dependent component to control for day specific effects such as weekend effects etc. The random effects model was selected in particular due to the fact that our dataset represents only a fraction of a very large population (data represented only randomly selected categories and activity across one month). To address the case of normality and the robustness issues that we might have with no normal cases in our data we used the log linear model. The results were obtained by using the logit, xtmixed and xtreg procedures in STATA.

Since the xtreg estimation model controls for fixed, between and random effects on a population averaged model the set of observations was reduced to 10.312 while during the estimation of the simple logistic regression the variable describing the number of comments was dropped due to multi-collinearity. In fact the xtreg procedure provides the Population averaged variation of the mixed effects model that we used. Overall the estimations explain roughly 12.4% of the variation represented on the dataset. However we should take into account the large volume of the observations and second the absence of other informative variables such as user participation characteristics, cultural background etc.

Overall the number of questions a user has posted has a significant effect on the

	Logit	Xtmixed	Xtreg
Number of Comments	-	0.950*** (0.155)	0.629*** (0.167)
Post$_{(u_i^p)}$	0.443*** (0.026)	0.127*** (0.009)	0.124*** (0.014)
Post$_{(u_i^a)}$	0.014*** (0.003)	0.021*** (0.002)	0.009** (0.003)
Post$_{(u_i^p)t}$	-0.084*** (0.013)	-0.123*** (0.013)	-0.059* (0.026)
Post$_{(u_i^a)t}$	0.036*** (0.004)	0.090*** (0.003)	0.044*** (0.006)
HHI$_{(u_i^p)}$	-0.039*** (0.003)	-0.137*** (0.005)	-0.108*** (0.022)
HHI$_{(u_i^a)}$	0.064*** (0.004)	0.234*** (0.006)	0.239*** (0.022)
Accepted Answers$_{(u_i^a)}$	-0.386*** (0.026)	0.036*** (0.010)	-0.061 (0.072)
Constant	1.772*** (0.020)	4.191*** (0.034)	5.692***(0.142)

Table 4.4: Model results for the set of estimators used for our analysis. R-squared : 0.124, * p<0.05, ** p<0.01, *** p<0.001,N: 133694, 134127, 10312

number of answers this particular user has received. This is expected due to the fact that the number of questions and the number of answers is a set of complementary variables where for an answer to exist there should be a question first. A similar case is also with the number of comments that a question has attracted. This is due to the fact that users who tend to provide an answer ask for more background information and clarifications on the question that the user posted. Since the number of comments are targeted to the user who posted the question is more likely that after a clarification a subset of the answers represents the reformation of other answers based on the clarifications provided. Therefore the number of comments has the high coefficient which obtained from our regression models.

Returning to the original set of variables that denote benefit and effort we can see that for the mixed effects model are highly significant while when adjusting for population averaged variation their significance somehow varies. The results partially

confirm the hypotheses presented on Figure 4.6. In particular overall effort, past effort and variation on effort have significant positive effect on service posture. On the other hand past received benefit and variation on the received benefit across categories had a negative effect on service posture (as measured on the number of answers a user has received). The exception is the overall benefit where as aforementioned since is a complementary variable with the number of answers it has a positive effect. The signs of the coefficients hold for both the standard random, fixed and between effects estimations as well as for the population averaged variation.

As can be seen in the Table 4.4 the highest coefficient obtained for our model was in both cases of variation (effort and benefit). In particular a roughly 20% increase in variation of the benefit provision will have an increase in one unit of the number of answers given to the questions that the user posted. The question variation has about the half effect (~10.8%) on the number of answers that the user receives with a negative sign providing that to increase the number of answers a user receives by one unit there is a requirement to reduce the variation of the questions posted by 10.8%.
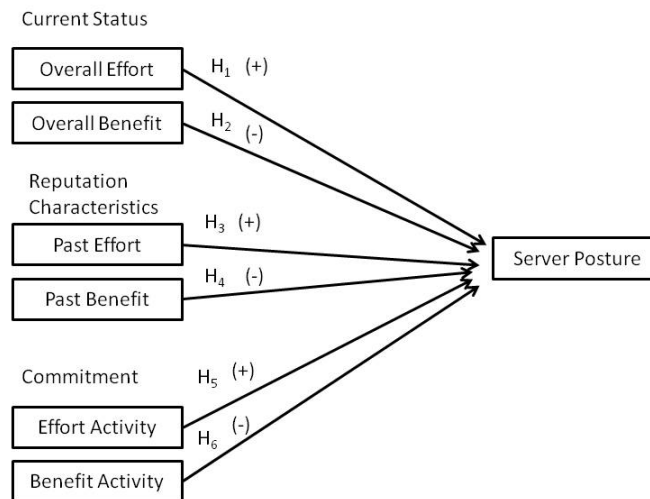
Figure 4.8: The results of the evaluation of our model.

Past activity on the number of questions posted by a user has a less significant effect than the overall volume ($|Post_{(u_i^p)t}| < |Post_{(u_i^p)}|$) of the questions posted. This

is confirmed also by the sign of the high value benefit (as indicated by the number of answers accepted. Past contributed effort has a much higher effect than the overall effort ($|Post_{(u_i^a)t}| < |Post_{(u_i^a)}|$).

4.4.3 Relation between effort and benefit

For evaluating specific cases of the relation between contributed effort and received benefit we used the ratio of the volume of contributed effort and received benefit in order to create a dichotomous variable. In order to be more consistent with the evaluation of service posture we considered two indicative variables for testing between the two groups for each variable namely the number of answers and the time to get an accepted question (defined in minutes). Table 4.5 shows the formation of the groups from our initial dataset. Considering the case of the number of answers we can see that when the effort is higher than the benefit the users receive on average 6.7 questions while on the contrary if the effort is less than the benefit the average number of answers received is 3.746

Group Statistics					
	Effort/ Benefit	Number of users	Mean	Std. Deviation	Std. Error Mean
Number of Answers	≥ 1	26399	6.771	7.323	.450
	< 1	73774	3.746	4.834	.017
Time to get an accepted answer for a question	≥ 1	3520	171.14	471.29	7.943
	< 1	4696	370.22	909.731	13.274

Table 4.5: Summary Group statistics for the groups that were formed

For the time to get an accepted question we see that those that provided more an-

swers than questions to the service received an accepted answer in 171 minutes while those that provided more questions than answers received an accepted answer in 370 minutes on average (more than double the time). In order to check if the difference between the means is statistically significant we run a two tailed t-test within groups. Levene's test for the equality of variance didn't reject the null hypothesis (t=75.316, P=.000) so we proceed with the two tailed t-test across the different groups.

t-test for Equality of Means							
	T	df	Sig. (2-tailed)	Mean Difference	SE	95% Confidence Interval of the Difference	
						Lower	Upper
Number of answers provided	75.316	100171	.000	3.024	.040	2.946	3.103

Table 4.6: Results of the two tailed t-test for the equality of the mean number of answers across the two groups.

The result of the two tailed t-test suggests that the difference between the means is statistically significant (t=-11.646, P=.000)[2] yielding that the number of answers posted was significantly higher for those users that had a high volume of contributed effort to the community compared with those that had a low volume.

The result of the two tailed t-test for the time to get an accepted answer suggests that the difference between the means is statistically significant (t=-11.846, P=.000)[3] yielding that the time to get an accepted answer posted was significantly lower for those users that had a high volume of contributed effort to the community compared with those that had a low volume.

[2]Levene's Test for equality of variances : F= 4410.725 , P=0.000
[3]Levene's Test for equality of variances : F= 368.537 , P=0.000

t-test for Equality of Means							
	T	df	Sig. (2-tailed)	Mean Difference	SE	95% Confidence Interval of the Difference	
						Lower	Upper
Time to get answer	-11.846	8214	.000	-11944.979	1008.322	-13921.546	-9968.41

Table 4.7: Results of the two tailed t-test for the equality of the mean time to get an answer across the two groups.

As can be seen in Figure 4.9 and Figure 4.10, the two factors that we consider to affect service posture as a latent construct change also slightly to the same direction (1: High Benefit , 2: High effort)

4.5 Discussion

The results that we obtained become of particular interest when compared with previous studies [Wasko and Faraj, 2005]. The interesting case with our study context is that activity has a strict supply demand mode of operation as outlined in the introductory section. In contrast to the case with open communities, such as online forums or Internet newsgroups where discussions evolve spontaneously in a linear way, in our context we used the design of the service to be very strict with the formation of activity around a particular thread. In particular, this demand-supply mode that the service operates, provides a clearer way for the formation of interaction, and permits a greater interaction level analysis in this study. Demand reflects demand for information, as depicted by the number of questions posted, while supply depicts the amount of answers provided. Thus, participants play the role of both the customer and the producer, depending on the participation role that they are into, either providing an
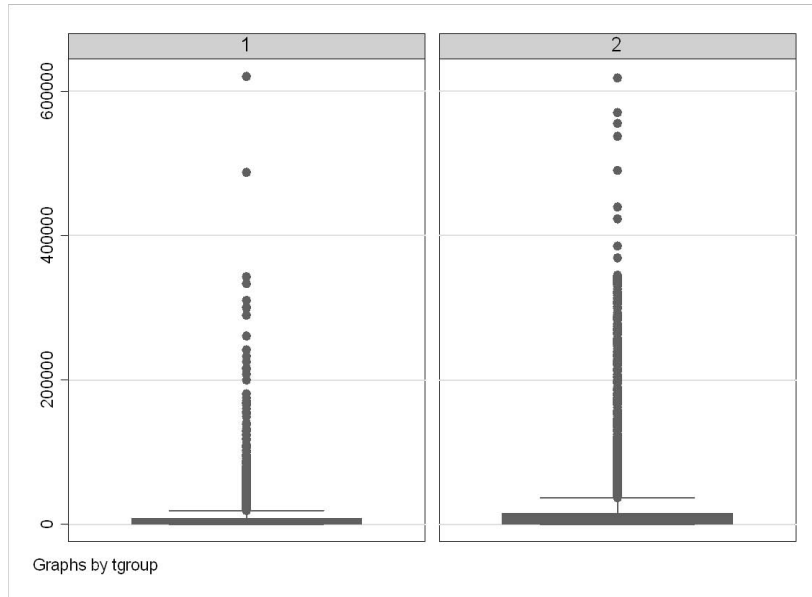
Figure 4.9: Comparison between means for the number of answers across the two groups.
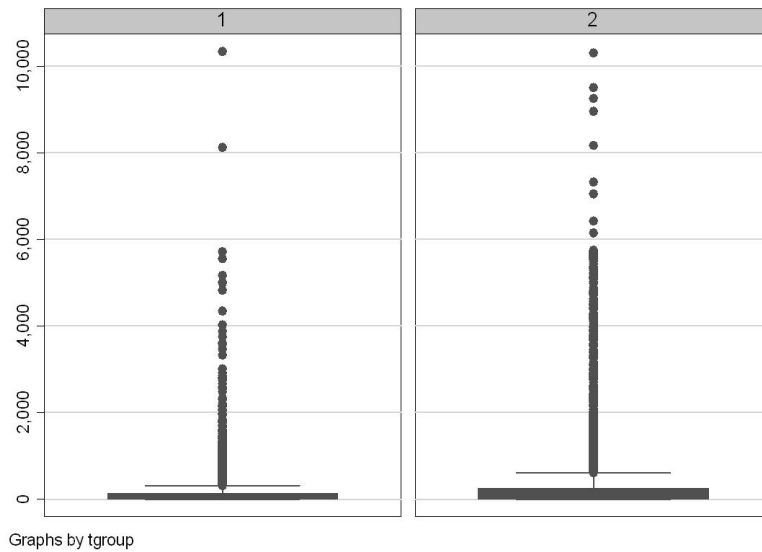
Figure 4.10: Comparison between means for the time to answer (minutes) across the two groups.

answer (supplier) or asking a question (customer).

Revisiting Figure 4.3 we can see that this open setting is also efficient where the number of questions posted and the number of questions closed provides an efficient supply/demand setting. This brings us back to the discussion on social capital creation initiated by Nahapiet and Ghoshal [1998] where an argument is made that organizational controlled settings have advantages compared with open environments, such as markets, due to the fact that they provide a more institutional way of development. While we cannot argue against this, an institutional way of developing social capital is undoubtedly through interaction. If social interaction is facilitated in a controlled procedural way as it happens on Yahoo!Answers, factors such as effort and commitment contributed by the users indeed dramatize a role, as this study confirms.

A second observation from the results obtained from this study is the confirmation to a certain extent of the anticipated reciprocity factor much discussed by Kollock [1999]. As the analysis has shown, when we make the comparison of contributed effort and received benefit and the volume of answers received, users tend to evaluate at some point the way this user has contributed or benefited by the community. Anticipated reciprocity provides that members of the online community expect that they contribute a lot of effort to the community, they can expect reciprocity from another member. However, since a significant number of the users in our dataset (73.2%) contributes less than they received, most of the users tend to evaluate the self-efficacy characteristic indicating that their contribution of valuable information will also affect their contribution characteristics [Bandura, 1977].

Overall, the empirical validation of the differences between the coefficients obtained from our statistical analysis implies that participants assess how much they participate on Yahoo!Answers, based on what they get from others.

4.6   Conclusions

The objective of this study was grounded in two underlying pillars: the contributed effort and the received benefit, in order to analyze contributions in online knowledge communities taking Yahoo!Answers as a case study. The level of service users receive in the future (as captured by volume and waiting time) is indeed affected by their own contribution record. Since the operational model of yahoo!answers differs from that of standard online communities, such as the usenet groups and fora, in the way that interaction happens in a more controlled way, the study contributes an interesting set of results. These results can further our understanding of the motivational factors underlying contributions in online knowledge communities.

Overall, this study provides evidence that participation in the Yahoo!Answers online community is characterized by the same norms that are evident in communities of public goods sharing. Participants who spend less time in providing value to the public good (that is the overall number of answers provided) and therefore contribute less effort in the community, are longitudinally sanctioned by the other participants in the form of receiving an answer in longer time than those that provide higher effort by answering a lot of questions posted by other fellow participants.

One particular limitation in the study presented in this chapter is the abstraction level that we used in order to argue for the relation between benefit and effort. It is argued in the online community literature that interpersonal interaction elements might also play a role in cooperation in online environments [Ardichvili et al., 2003]. Such case can be for example the notions of altruism [Lampel and Bhalla, 2007] and reciprocity [Maloney-Krichmar and Preece, 2005] both in a generalized and dyadic level. Furthermore the analysis presented here assumes the existence of a reputation mechanism tied to a participants' profile available for review by all the other participants. This study has shown that the presence of a reputation system such as activity history for example can be rewarding for those users that contribute a lot to the community while on the other hand can be disadvantageous to those user that seem to receive

more benefit than the contributed effort.

Another important factor that we excluded from our analysis is the knowledge barrier that a user might have in order to provide an answer to the community. However this is dependent on the quality of the question posted and is certainly difficult the measurement of a construct to capture this case as well.

Furthermore one particular element that we should take into account in a more extensive study on the dataset we collected is the case of repeated interaction in terms of structural connections formed by the users in the context of generalized exchange [Ekeh, 1974]. A recent study by Zhang et al. [2008] takes that into account by presenting an information theoretic model of participants' interaction in the context of Yahoo!Answers. However this model is memory agnostic providing that the history of interactions is not measured in terms of previous participation in questions or answer but investigates the knowledge potential of the participants to provide an answer. Furthermore a more extensive study should consider a larger fragment of interactions than those captured by the presented dataset in order to have a more clear view of the activity in a broad set of categories.

Bibliography

A. Ardichvili, V. Page, and T. Wentling. Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *JOURNAL OF KNOWLEDGE MANAGE-MENT*, 7(1):64–77, 2003.

R. Baeza-Yates and B. Ribeiro-Neto. *Modern information retrieval*. Addison-Wesley Harlow, England, 1999.

A. Bandura. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84:191–215, 1977.

S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30:107–117, 1998.

J. S. Brown and P. Duguid. *The Social Life of Information*. Harvard Business School Press, 2002.

S. Chakrabarti. *Mining the Web: Discovering Knowledge from Hypertext Data*. Morgan Kaufmann, 2003.

C. S. de Souza and J. Preece. A framework for analyzing and understanding online communities. *Interacting with Computers*, 16:579–610, 2004.

S. Dumais, M. Banko, E. Brill, J. Lin, and A. Ng. Web question answering: is more always better? *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 291–298, 2002.

A. Edmunds and A. Morris. The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management*, 20:17–28, 2000.

P. P. Ekeh. *Social Exchange Theory: The Two Traditions*. Harvard University Press, 1974.

S. Faraj and M. M. Wasko. The web of knowledge: an investigation of knowledge exchange in networks of practice. *Working Paper. Florida State University, Tallahassee, FL, USA.*, 5:2004, 2001.

E. Fehr and U. Fischbacher. The nature of human altruism. *Nature*, 425:785–791, 2003.

E. Fehr and S. Gachter. Cooperation and punishment in public goods experiments. *The American Economic Review*, 90:980–994, 2000.

L. Festinger. A theory of social comparison processes. *Human Relations*, 7:117, 1954.

D. Fisher and P. Dourish. Social and temporal structures in everyday collaboration. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 551–558, 2004.

A. O. Hirschman. The paternity of an index. *American Economic Review*, 54:761–762, 1964.

Pawel Jurczyk and Eugene Agichtein. Hits on question answer portals: exploration of link analysis for author ranking. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 845–846, Amsterdam, The Netherlands, 2007. ACM.

J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46:604–632, 1999.

P. Kollock. *The economies of online cooperation*. Routledge, London UK, 1999.

C. C. T. Kwok, O. Etzioni, and D. S. Weld. Scaling question answering to the web. *Proceedings of the 10th international conference on World Wide Web*, pages 150–161, 2001a.

C. C. T. Kwok, O. Etzioni, and D. S. Weld. Scaling question answering to the web. *Proceedings of the 10th international conference on World Wide Web*, pages 150–161, 2001b.

J. Lampel and A. Bhalla. The Role of Status Seeking in Online Communities: Giving the Gift of Experience. *Journal of Computer-Mediated Communication*, 12(2):434–455, 2007.

D. Maloney-Krichmar and J. Preece. A multilevel analysis of sociability, usability, and community dynamics in an online health community. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2):201–232, 2005.

J. Nahapiet and S. Ghoshal. Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23:242–266, 1998.

R. Oliver, A. Omari, and J. Herrington. Exploring student interactions in collaborative world wide web computer-based learning environments. *Journal of Educational Multimedia and Hypermedia*, 1998.

D. M. Powazek. *Design for Community: The Art of Connecting Real People in Virtual Places*. New Riders, 2002.

J. Preece. *Online Communities: Designing Usability and Supporting Socialbilty*. John Wiley & Sons, Inc. New York, NY, USA, 2000.

R. Soden and J. Halliday. Rethinking vocational education: a case study in care. *International Journal of Lifelong Education*, 19:172–182, 2000.

Qi Su, Dmitry Pavlov, Jyh-Herng Chow, and Wendell C. Baker. Internet-scale collection of human-reviewed data. In *Proceedings of the 16th international conference on World Wide Web*, pages 231–240, Banff, Alberta, Canada, 2007. ACM.

G. Walsham. Knowledge management: The benefits and limitations of computer systems. *European Management Journal*, 19:599–608, 2001.

Y. Wang and D. R. Fesenmaier. Assessing motivation of contribution in online communities: An empirical investigation of an online travel community. *Electronic Markets*, 13:33–45, 2003.

M. Wasko and S Faraj. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *Management Information Systems Quarterly*, 29:35–57, 2005.

Etienne Wenger. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge UniversityPress, 1998.

S. Whittaker, L. Terveen, W. Hill, and L. Cherny. The dynamics of mass interaction. *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 257–264, 1998.

J. Zhang, L. Adamic, MS Ackerman, and E. Bakshy. Everyone knows something: Examining knowledge sharing on Yahoo Answers. *World Wide Web (WWW'08)*, 2008.

The impact of Extrinsic Rewards on Strategic Interaction in Online Communities: An analysis on Google!Answers

In this chapter we analyze interactions in an online community where contributions are not motivated by intrinsic means but by monetary rewards. The effect of economic incentives, namely price and tips, is investigated in the context of an online information market operated by Google for a period of four years. By collecting data for the whole period of operation and constructing panel data variables for the users' history on the platform, we estimate the effect of price and tips on service promptness as recorded by the time to get an answer and the subjective answer evaluation given by the user. We explain our findings by discussing three complementary theories from a customers'/askers' point of view.

5.1 Introduction and motivation

Thus far, we have argued that online communities resemble rich social environments where participants exchange information either in a direct or indirect mode. In this

chapter, however, we adopt a different perspective on the way information exchange takes place, approaching it from an economic viewpoint. Therefore, the scope of this chapter is twofold: (a) to illustrate and describe a particular case of an online community where social exchange has also an economic nature, and (b) to understand the behavioral patterns in which the participants engage in order to enhance the service that they receive from an online community with the least cost / effort provided in terms of extrinsic means (e.g., payments using monetary rewards).

In order to do this, we use as the subject of the empirical study the case of GoogleAnswers. With greater detail given later, GoogleAnswers represents a unique social environment where social and economic exchange take place with the subject of exchange being information produced by individuals and directed to other individuals who state their demand. However, information is rather an imprecise concept from an economic point of view. Arrow [1996], for example, argues that information as an economic good is rather difficult to theorize as an economic commodity (and thus a subject to an exchange) due to the fact that its value cannot be measured in a standard way compared to other types of commodity.

Re-contextualizing this case of information as a subject of economic exchange on the context of an online community, we argue that the individually perceived value of information is an important issue due to its subjective nature [Rafaeli and Raban, 2003]. For example, how can someone evaluate the economic value of an individual's posting that describes the solution to a specific IT problem if, for example, there is no information about the individual's function (e.g., an IT professional facing the same issue) or whether he/she is just a novice user that is interested in troubleshooting a similar problem. In both cases, in online communities the absence of such hierarchies enhances the motivation for contributing based solely on intrinsic means, such as altruism or the perception of the collective as a public good where everyone should contribute (e.g., as seen in the Yahoo Answers environment in the case of the previous chapter).

Undoubtedly, voluntary contributions in online communities are an important element for evaluating the sustainability and growth of an online community [Butler et al., 2007, Preece, 2000].]. Contributions are, in general, of intrinsic nature [Ardichvili et al., 2003] providing that this pattern of behavior is expressed in contexts where future prospects are highly dependent on the frequency of use [Gu et al., 2007]. Frequency of use, however, results from participation which is also on a voluntarily basis, providing that the members have no obligation to interact unless they have personal interest to do so. Nonetheless, the value of information that is exchanged in an online community might significantly affect the willingness to pay for information for those that expect an answer [Lee and Hatcher, 2001, Varian, 1997]. This can be attributed to the fact that the probability that someone who posts a question will get an answer is dependent on several factors, including expertise, commitment by the other members and activity [Whittaker et al., 1998]. Furthermore issues such as free riding and lurking [Preece et al., 2004, Rafaeli et al., 2004] can make this likelihood smaller due to the fact that there is no contract enforcing mechanism (as an effect of the absence of organizational hierarchies inside the online community).

On the other hand motivation to contribute can also be of extrinsic nature [Teo et al., 1999]. This is because both sides of the users in an online community are expecting an exchange to occur out of the interaction. In this case, the study of interaction is not under the focal point of (intrinsic) social incentives, but is dependent on the monetary rewards that the users are willing to pay. In such cases the online community is transforming itself into a type of information market where suppliers of information (sellers) transact with buyers who receive compensation for posting information which itself is available to everyone else. What makes Google Answers a unique context of study is that apart from using monetary rewards to provide contributions by a controlled pool of researchers (thus securing a high standard of quality) is that the contribution is public and remains accessible to visitors in the community, thus maintaining the characteristic of an online virtual community where information

is a commodity (free access to the content for all the participants). This becomes a social dilemma. Participants know that if they pay to get a specific answer, that this answer will remain public; thus, they might choose to wait in case another user asks (and pays) for the same or similar question. Furthermore, another characteristic that makes this environment unique is the ability of users to reward the providers with a tip on top of the service once they are more than satisfied with the service provided. Google Answers operates in a mode of an information market composed of a controlled pool of suppliers and an open pool of buyers. Information exchange in this market is based on a type of auction model of the standard supply and demand function which incorporates the side of the customers (those who ask for information) and the side of the producers who provide information in the form of answers to a question with a defined price tag. A buyer (asker or customer) posts a question defining a price that he/she is willing to pay for an adequate answer. A researcher (supplier or server) takes over and provides an answer within a time frame of 30 days. If the buyer accepts the answer, the researcher receives the amount stated by the buyer when the question was advertised. If the buyer is more than satisfied with the answer provided, he/she is also able to provide a tip. From the design of the service, the price of a question cannot exceed 200 USD and of the tip 100 USD. The presence of the price and the tip provides a unique opportunity to investigate the impact of an extrinsic form of mo-tivation to the contribution in an online community and, in particular, the impact of monetary rewards for the response time and the quality of contributions by the sellers to the buyers.

As aforementioned, while there is the function of the market, the interaction be-tween the buyers and the sellers creates externalities which are the high quality (as can be seen later) answers of the sellers to the buyers. These externalities, however, have a unique characteristic for the (information) market: (a) they are accessible on-line by everyone, and (b) the recorded history of interactions is accessible to both buyers and sellers. In that sense, if someone was seeking information about a specific

subject and the same question was posted by another buyer, then this buyer would not need to purchase that information because the answer was public and not excludable. We can argue from this case that the externalities produced by the information market, constitute a public good where its contributions are of extrinsic nature.

What makes extrinsic forms of motivation, such as monetary rewards, interesting is the examination of the effect that the volume of the reward will have to the promptness or promptitude of the answerers to the askers in future interactions. This is also the scope of this study. The ability of the buyer to provide an extra reward (in the form of a tip) on the aftermath creates a set of prospects for both buyers and sellers. Sellers will hold the expectancy of this reward as a way of receiving a premium, additional to the already stated compensation, while askers might use the tip to strategically interact with the answerers in order to receive better service in the future.

The main objective of this chapter is to examine the effect of economic incentives on service promptness in the context of an online community, such as the one in GoogleAnswers. Promptness of service is a typical aspect of operation where the use of an incentive mechanism can enhance, due to the fact that it represents a clear effect of the incentives given in performance.

The case of tipping represents a classical paradigm of incentives in the service sector where customers do tip servers in order to reward in addition to the price the service provided [Lynn and Mynier, 1993]. In general, a tip is a reward provided to a service provider by a customer beyond the advertised price of a service in order to demonstrate gratitude related with the background variables of the service, such as behavior of the server, etc. In cases as in the service industry in the US, tipping is an important factor that influences the exchange from service providers to customers. In particular, due to the fact that tips become another source of income added to the wage, servers add a particular emphasis on the tip and very often pay attention to the customers who provided them and recall them in future interactions. Nonetheless, the reason why customers tip is not clear. One explanation that is generally accepted is

that customers tip because they want to reward the service motivated by reciprocity [Lynn and Latane, 1984], or they feel obliged to do it by the imposition of a social norm [Azar, 2007]. Another direction suggests that customers tip due to the fact that they hold future prospects for better service by the service provider.

Furthermore, the study of tipping behavior is an interesting theme from an economics viewpoint due to the fact that the phenomenon itself contradicts the neoclassical rational agent model where the agent is expected to maximize its utility [Azar, 2003]. It is expected that a buyer will optimize his/her utility by following the budget constraint imposed by maximizing his utility at the lowest possible cost. Nonetheless, the case of tipping clearly deviates from this assumption. Through the modern history of the service sector, tipping in principle is considered to be a reward for the perceived quality of the service provided [Azar, 2004a]. Satisfied customers reward quality while those who are not satisfied do not; thus, tipping is used as a form of extra incentive for numerous employees in service oriented tasks (e.g., hospitality) to provide better service to their customers. Depending on the cultural background, tipping behavior varies. In some countries, such as the US, tipping has been embedded in the behavior of the customers until it has become a social norm. This implies that regardless of the satisfaction level of a customer by the server, a tip has to be given usually at a fraction of the paid price.

With respect to the study context of this chapter, askers might be interested also to compare their condition with askers that received utility by imposing the same or smaller amount, thus falling to a case of social preferences, typically on a situation of social comparison [Festinger, 1954]. From this perspective a tip can be seen as an illustrative example of reciprocity [Fehr and Fischbacher, 2003].The customer received good service which was above the standard level and is reciprocating his satisfaction to the server by providing a tip. This type of reciprocation clearly corresponds to satisfaction by the provided service. The more satisfied the customer is, the bigger will be the monetary reward that he will use to favor this level of quality. However, there are

cases where reciprocation is not the objective when there is tipping involved. Studies have shown that customers also use tipping because they have future prospects about the service and thus use it frequently [Lynn and McCall, 2000]. This kind of tipping behavior is more oriented to strategic purposes where customers have future prospects about the perceived quality.

However, distinguishing when someone has tipped because she/he was obliged to do it or perceiving the tip as reciprocation with or without future prospects is something that is very difficult to examine in field settings, due to the difficulty of elicitation of the transaction history. This is due to the fact that elicitation of complex data is demanded as well as a set of ceteris- paribus settings that will not affect the existing behavior of both customers and service providers. Furthermore, there is the need to control for other variables, such as face to face interaction, cultural background of the server, and the customer, etc.

In this study we study the effect of economic incentives (both in explicit and implicit form) to the outcome of interaction in the GoogleAnswers online community. Following the above objective, this chapter is structured as follows. Section 5.2 provides the background theory on tipping and the behavioral patterns associated with the situation where customers do give tips. Section 5.3 provides a description of the service and a review of some descriptive statistics on the tipping behavior of the users based on the data that we collected. Section 5.4 presents a description of the methodology and the constructs that we use in order to examine the effect of economic incentives in the interaction and the quality of exchange in the GoogleAnswers. In particular, Section 5.4.2examines in depth the use of non linear estimation models the factors that affect both the decision to tip and the magnitude of the tip given, as well as its impact on service quality 5.4.3. Furthermore based on the general factors that affect tipping as a reward for extra service we fit the theoretical background discussed in section 5.2 in order to connect our empirical results with the findings provided from the dataset. We discuss our findings and implications in Section 5.5. Section 5.6 offers

conclusions regarding the implications of this study and future research directions, as well as possible limitations.

5.2 Characteristics of tipping as an extrinsic reward

Although tipping might not be considered the perfect instrument to evaluate behavior on an online setting as the one represented in Google Answers, is an important tool to examine whether participants do actually have social preferences (care about the actions of others) and thus change their behavior accordingly. Tipping on that case (irrespective of context, offline or virtual) is an ideal instrument since it violates the assumption of selfish agents by demonstrating the impact of psychological and social parameters as the ones (e.g. communication) discussed on the study of 3.

On the context of this chapter we study tipping as a mechanism that offers to the buyers of the GoogleAnswers service a way to assure the quality of the received service and thus motivate the service participants to hold on this particular behavior. Azar [2003] discusses the implications of tipping as a mechanism that can be extended to other areas of the management science literature including firm contracts for example. One particular case of tipping that we address in this chapter from the perspective of an online community is the case where tipping might be employed as a way for receiving better service in future cases. The expected theoretical result according with the notion of the group efficacy discussed in Chapter 2 is that familiarity will have an impact on the way the exchange is shaped between the server and the customer. But how is tip motivated from customers to servers?

Lynn and Grassman [1990] provide a set of three different approaches to tipping in a restaurant setting.

- Buying social approval with tips

- Buying equitable relationships with tips

- Buying future service with tips

The case of the social approval has to do with the imposition of socially enforced behavior in order to show compliance with the current social norm or to avoid cases of social disapproval [Cialdini and Trost, 1998, Reno et al., 1993]. Azar [2004b] takes this case one step further arguing that tipping represents the case of an "internal norm" where internal sanctions such as the lost of self esteem are more effective that other forms of e.g. external sanctions [Gintis, 2003]. In the context of study of an online community social disapproval is something that can be addressed by a user who changes the username many times in order to avoid such a case. Furthermore coordination of sanctions in the context of an online community is a rather difficult issue due to the fact that is dependent on the cohesion that the online community as a social group possesses.

Buying equitable relationships with tips, builds on the characteristics of equity theory as a factor that evaluates an exchange relation for an (economic) agent with a potential aversion to inequity [Bolton and Ockenfels, 2000, Walster et al., 1978]. For instance if we consider the relationship between a customer and a server as a relationship where both parties want to have equitable outcomes then the tip should be positively correlated with the quality of the service received by the customer. In that case there is an issue with the customer's own perception of the service received. For example in our case a customer may had different (low or high) expectations from the service received by GoogleAnswers' researchers due to the fact that the nature of the service doesn't involve production but as we are going to describe on the next section, processing of information currently available in Google's search index. In that case it is expected that the given tip will have a reciprocal nature building on equitable outcomes, good service will be rewarded with a higher tip than bad service.

A different approach with tipping comes on the case where interaction is not single and un-repeated. In the case of tipping in restaurants for example, customers are coming in repeated cases (e.g. every Sunday) and therefore they build a reputation by their past behavior. In such case there is the involvement of a strategy similar to the

205

infamous *tit for tat* [Axelrod and Hamilton, 1981]. In a continuous case good service will be evaluated with a good tip in the future while a bad service will be evaluated with a bad tip in a future iteration. In our case this suggests that the relationship between service and tipping should be stronger for frequent patrons of the service rather than single case participants. Table 5.1 provides a classification of tipping and its evaluation in a field study

Tipping	Evaluation
Tips as a mean to buy social approval / Imposition of the social norm	Constant across cases
Tips as a subject to equity/ Tipping due to reciprocity	Positive correlation of tip and service quality for single cases
Buying future service with tips / Tipping as a strategic interaction case to obtain better service	Negative correlation of tip and quality for future cases (tip of the previous case should be positively correlated with the quality of the current case). Better service is achieved with gradually lower tips.

Table 5.1: [The three general aspects of tipping] The three general aspects of tipping adopted from [Lynn and Latane, 1984]

To further diversify the above classification, Azar [2007] provides another interesting classification of tipping occasions, classifying tipping cases into six clusters: (1) reward-tipping (2) price-tipping, (3) tipping-in advance, (4) bribery-tipping, (5) holiday-tipping and (6) gift tipping.

In connection with the classification of Lynn and Graham, the case of reward tipping can be classified as a case of reciprocity, since tipping in that case is provided as a reward for a good service. The bribery tipping can also be classified into that category

due to the fact that the server is expected to provide the customer with something in exchange for the tip that received. In contrast with standard tips, bribery tips are expected to have no repeated nature due to the instability of the customer's utility model (it only makes sense to bribe when there is an interest on a particular case with no attempts for reputation building). The case of price tipping and gift tipping can be also classified to the category of social norm. In particular, price tipping is a standard occurrence when the cost of the service is not included in the bill (in that case the customer is expected to provide a tip). Nonetheless a characteristic of this form of tipping is the positive correlation that in most of the cases can be observed between the service price (e.g. the bill in a restaurant setting) and the size of the tip. Similar to that case can be the form of holiday-tipping where for a fixed period the customer tips due to the context of service and the quality of the service received. Such a case can be attributed to temporal fixed effects due to the fact that on the holiday period customers tend to be more generous than in ordinary cases which can be in turn attributed to the positive emotions that a relaxed holiday period might have to a customer [Ruffle, 1999].

Emotions might also play an important role in the case of gift tipping, where the social approval imposed by a social norm allows us to categorize such case of tipping behavior to the category of the social norm due to the attributed fixed effects (customer's might be generous in general). Such cases usually are identified by the significant higher size of the tip. In Table 5.2 we categorize the context specific cases of tipping in relation with the approach in our study. Having summarized these three cases we tend to evaluate whether those three approaches hold for the context that we study.

In order to do that we first analyze the factors that affect tipping in general, then we identify those cases that the three aforementioned aspects hold and finally we evaluate each of these cases separately.

207

Tipping Case (Azar, 2007)	**Tipping Behavior**(Lynn & Grassman, 1990)	**Evaluation Approach in this study**
Gift tipping Holiday-tipping	Tips as a mean to buy social approval / Imposition of the social norm	Social Norm
Price-tipping Bribery-tipping Reward-tipping	Tips as a subject to equity/ Tipping due to reciprocity	Reciprocity
Tipping-in advance Reward-tipping (where strong past interaction is observed)	Buying future service with tips / Tipping as a strategic interaction case to obtain better service	Strategic Interaction

Table 5.2: Cases of tipping in relation with this study

5.3   The Google Answers online community

Google Answers was introduced in April, 2002 as a beta[1] project without much public-
ity by Google itself, passing into full operational mode some months later. In principle
the service targets potential users that are not experienced in web search retrieval
and are willing to pay someone to find information for a monetary reward. The ques-
tions are posted to several categories organized under six large categories concerning
matters related with health, personal relations, business, computers etc.

The service has two user types namely the askers and the researchers. An "asker"
submits a question and an amount that is willing to pay for a satisfactory answer.
The asker also defines a time window that is expecting the question to be replied;
otherwise the default expiry date is 30 after the time the question was posted. The
amount that the asker is willing to pay a researcher ranges between 2 and 200 US$.
Figure 5.1 provides a description of the process flow in the Google Answers website.

Figure 1 provides a graph of the service level provided in Google Answers during all
the period of the operation of the service. We can see that from the beginning of the
service until roughly the end of the first year of operation the activity of the service
increases exponentially however having a significant drop on the beginning of 2003.
This might be due to the maturity of the users visiting the service due to the fact that
internet services receive much advertising in the beginning in order to attract a critical
mass of customers.

The entry fee for the service was 0.50 US$ and was kept even in the case nobody
replies (service entry fee). In fact the entry fee resembles the structure of an adver-
tisement flat fee that anyone who wants to use the service must pay. Once the asker
posts a question, a question is assigned from the beginning to a researcher and is
locked to him/her for a period of 3 or more hours. If he doesn't reply during the spec-
ified time-window then the question opens for all the other researchers participating
in the system. After the answers have been posted the asker decides which ques-

---

[1]The term beta refers to software that although functional is still on testing phase
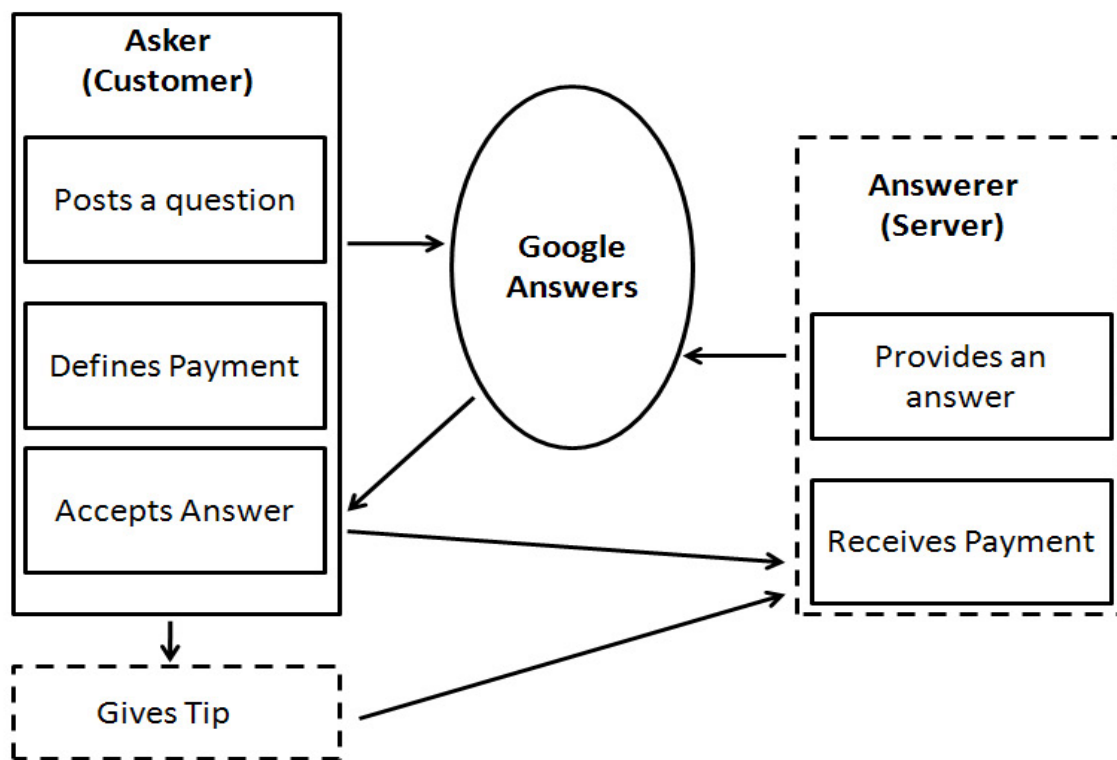
Figure 5.1: The process flow of interaction between askers and answerers in Google Answers
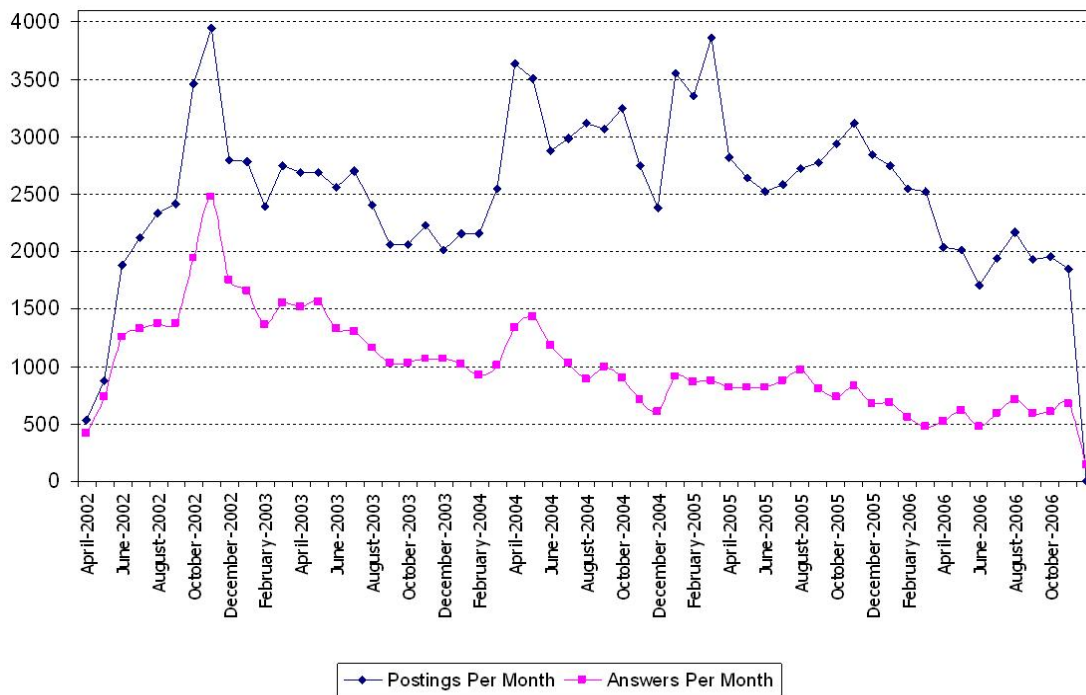
Figure 5.2: Questions Posted per month to the webservice during a four years period (2002-2006)

tion fulfills better his/her requirements and the question is marked as closed/resolved. Upon selecting the satisfactory answer the asker has to pay the answerer and in case he/she is far than satisfied by then given answer he/she can tip the answerer with an amount ranging from US$1 to US$100 from which Google keeps no fee Also the asker evaluates the answerer by using a five point Likert scale (0 to 5). It should be noted that for a researcher the motivation to provide a better service and thus get a type might also be affected by the fact that Google keeps no fees from the tip given by an asker.
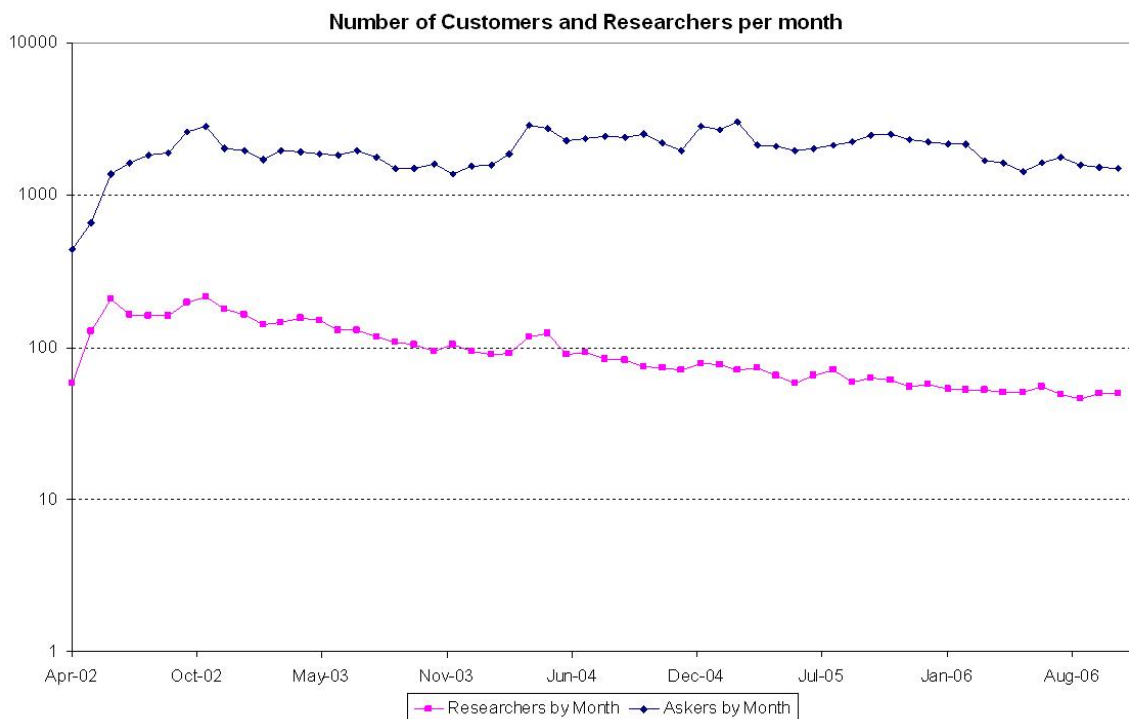
Figure 5.3: Number of unique askers and researchers per month for all categories (Values on Y axis are on logarithmic scale)

While an asker is able to register on the website and post a question after he has submitted his / her service payment information, Google maintains a pool of pre-evaluated researchers who have proven communication skills and are familiar with query techniques for the search engine in order to be able to do better research than an average user of the Google search engine. Figure 5.2 displays the volume of askers

212
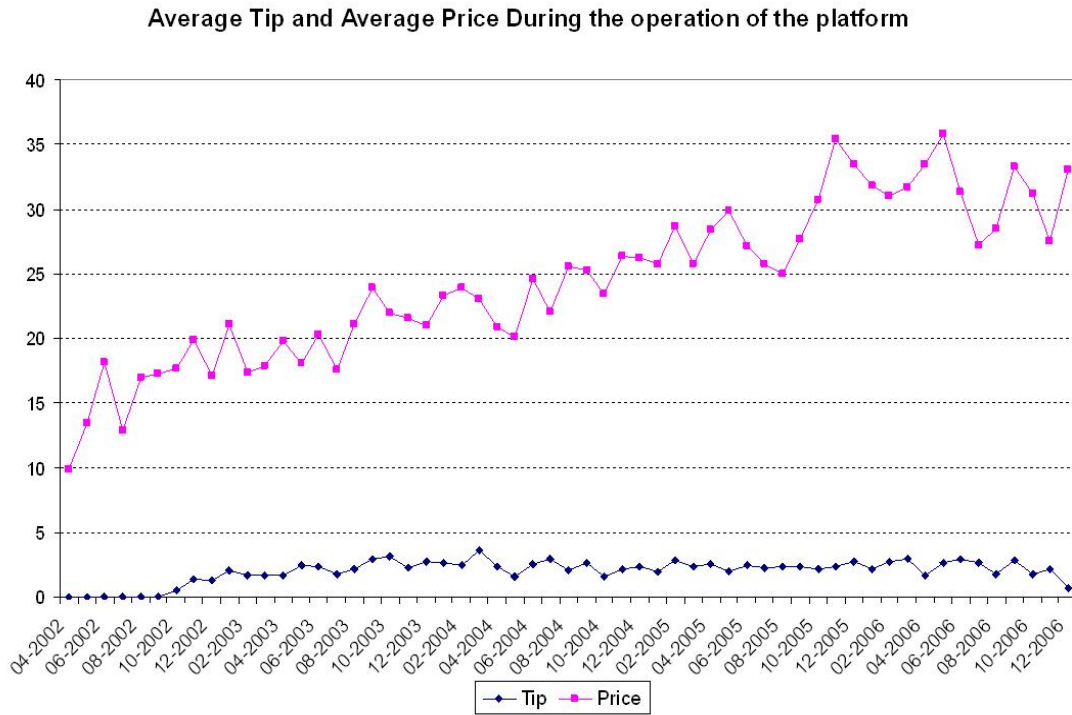
Figure 5.4: Average Tip and Average Price per month during the operation of the platform

(customers) and researchers (servers) participating in the online service per month in a logarithmic scale. While we can see that number of customers participating on the service is maintained quite high during all the operation of the website. On the contrary the number of servers is steadily decreasing resulting in a gap which can clearly explain the gap observed in Figure 1 between the number of postings on Google Answers and the number of questions answered by the researchers.

Having described the general operation mode of the Google Answers service we continue to describe the data collection method and some characteristic of this data in relation with the research objective of this chapter.

5.3.1   Dataset and Variables Description

Our data is based on a complete dataset of the recorded interaction during the Google Answers operation. The dataset was collected using an automated data collection

program that produced sequential web addresses (universal resource locations - URL) for the web pages containing the questions. For example a question page may have a web address of the form:

`http://answers.google.com/answers/threadview?id=318154`

In that case the identifier of the question is 318154. As it is depicted in Figure 5.5 by obtaining the identifier we stored it in a data structure (queue). That was used by the same program to extract the information contained on the web pages and insert them into a relational database for further processing.

Apart from the data available by browsing the archive, we were triggered by the results provided by the websites' search engine (searching for common words such as *and*, was providing a greater number of results than those indicated on the website) we sorted the question id and field the missing sequential ones. By doing that we were able to obtain a question ids that were not retrievable from the website, however could be retrieved by using the search mechanism of the web site.
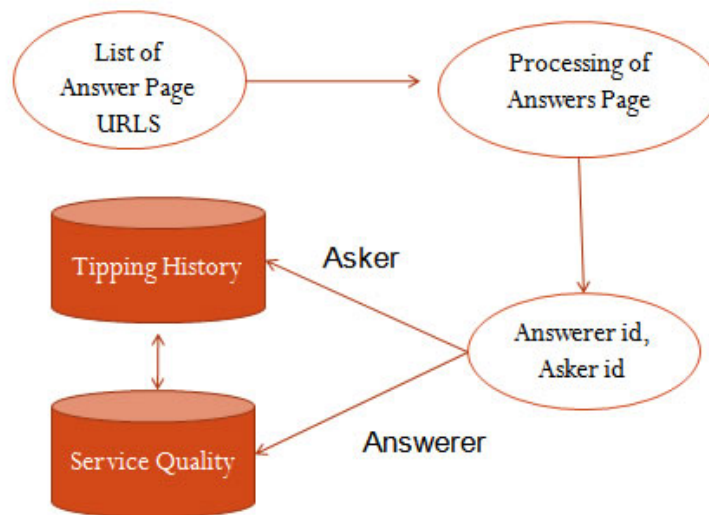
Figure 5.5: Data Collection Method

Our dataset contains all the visible questions (publicly accessible through browsing) on Google Answers and a significant set of hidden (not linked) questions that were available through the search mechanism. The variables in this dataset are as follows:

- **Price:** The amount that the asker is willing to pay to any researcher that will provide an adequate answer. As aforementioned the price cannot be greater than 200 US$. Apart from that the asker has to pay a $0.5 extra for entry fee which is not refundable.

- **Tip:** The tip that was given to the researcher after he gave an acceptable answer to the posted question. The tip cannot be more than $100.

- **Time to answer:** The time between the date that the question was posted and the answer date. On the Google answers website, the time is represented using two different time zones, Pacific Standard Time (PST) and Pacific Daylight Time (PDT). We normalized the time zones to coordinated universal time (UTC) by using the formulas: PST = UTC-8 and PDT = UTC-7. The time to answer is measured in days which mean that if an answer was given less than 24 hours than the time the question posted then the time to answer is zero.

- **Answer evaluation:** After an asker accepts an answer he provides an evaluation of the answer by using a 1 to 5 point likert scale. This variable depicts the *subjective* evaluation of the answer by the asker. Answer evaluation is optional since someone can provide a payment without evaluating the answer.

- **Clarifications Given:** If a researcher is not sure about the scope of the question then he/she may ask the customer/asker for a clarification. This variable depicts the number of clarification requests that a researcher has asked from an asker.

- **Number of references:** The number of external references (hyperlinks) that the researcher has provided to his answer. This is sometimes implied if the researcher wants to support his/her answer by providing links to credible sources (citations to web pages etc). However it might be also problematic if the answer is rather not clear but based on a set of references, directing the asker to seek further the answer in those references.

215

- **Answer length:** The number of characters that the accepted answer contains. This was calculated by counting the number of characters on the answer text[2].

Table 5.4 summarizes the distribution of **answered** questions to the categories available in our dataset. The average price per question answered is around US$ 22. Roughly an amount of 20% of the questions posted in all categories is tipped by the askers with an average tip of US$2 with the highest average tip given in the category health. This probably could be explained by how important the askers value information regarding health issues.

5.3.2 Descriptive Statistics

The dataset that we collected contains 140.605 questions that were posted over a period of 56 months (roughly 4 years) from the website. The definition of the variables required excluding some of the questions. In particular we excluded the expired or non-answered question from this dataset; we compiled a new dataset with 56.976 answered questions. Table 5.3 provides the descriptive statistics for the variables in our dataset.

As can be observed by the table the average price for a question posted in the service was about 23 US$. If the asker was in general satisfied by the answer he was rewarding the researcher with a tip which the average price is about 2 US$ with a range of 2 to 200 which is the minimum and maximum value resulting from the definition of the service.

The average time (in days) that was required to answer a question by a researcher was 1.73 or about 40 hours. The reputation of researchers (as calculated by averaging the answer evaluations they have received) is quite high and is 4.5 out of 5. This can be expected due to the fact that Google is evaluating the researchers before giving them permission to participate in the platform, so there is a high selection effect on this variable. The number of clarifications given is low which means that in general

---

[2]This was done by using the CHAR_COUNT command from MySQL.

researchers do not "bother" the askers to clarify their question. There might be also some cases where the researchers ask the askers for more details that may come up during the research.

Variable	Mean	Std. Dev.	Min.	Max.	N
Tip	1.895	7.7	0	100	56976
Number of references	7.347	12.219	0	582	56976
Time to answer	1.735	4.717	0	30	56976
Price	22.789	36.097	2	200	56976
Clarifications given	0.166	0.674	0	18	56976
[Frequent Patron]$_{i(t)}$	4.815	15.589	0	127	56976
[Previous Met]$_{ij(t)}$	0.365	2.162	0	54	56976
[Tipped Before] $_{i(t)}$	1.793	9.531	0	127	56976
[Tipped Before] $_{ij(t)}$	0.151	1.215	0	46	56976
[Reputation] $_{j(t)}$	4.566	0.268	1	5	37941
[Avg.Tip] $_{i(t)}$	0.935	4.486	0	100	56976
Answer Length (Characters)	4561.727	5708.668	0	65535	56976

Table 5.3: Descriptive statistics. N denotes the number of observations for the specified variable

The time dependent variables also provide an interesting insight. While posting a question an asker has used the service on average five times before. This indicates a high number of returning customers who are continuously using the service.

As can be seen in Table 5.4 the highest amount paid on average for a question is 35.85 in the category: Business and Money and the lowest on the category Family and Home. This can provide an indication that the perceived value for information is greater (as seen by the demand) in the category Business since buyers are willing to pay more for answers on this category than answers for the category family.

Category Name	Total Questions	Tipped	%Tipped	Avg. Price	Avg. Tip
Arts and Entertainment	6303	1606	0.255	14.70	1.55
Business and Money	9592	1715	0.179	35.84	2.22
Computers	8983	1810	0.201	20.51	1.88
Family and Home	2181	448	0.205	17.89	1.68
Health	4346	872	0.201	28.39	2.50
Reference, Education and News	6553	1438	0.219	21.05	1.91
Relationships and Society	2573	677	0.263	22.18	2.14
Science	3922	788	0.201	19.96	1.74
Sports and Recreation	1876	390	0.208	18.38	1.33
Miscellaneous	10674	2116	0.198	19.47	1.68
Summary	56967	11860	0.213	21.837	1.863

Table 5.4: Distribution of Questions posted to the various categories on the Google Answers website

On the other hand we can observer that the percentage of the questions tipped remains relatively constant through the categories. In fact about 20% of the questions tipped were in the category "Relationships and Society" where most of the questions were concerning personal advice which indicated also a substance of a more personal approach between buyers and sellers.

5.4 Methods and constructs

Having described our dataset we continue to the definition of the constructs that we are going to use in this study. As aforementioned in the introduction the original intention is to examine the effect of tip/price to the quality using the three classes of tipping behavior discussed in Section (2).

5.4.1 Definition of constructs

Measuring Past Interaction

In relation with our research question we constructed a set of time dependent variables from the initial dataset. In order to generate the variables we run a set of SQL queries to obtain some characteristics for both askers and researchers.

Figure 5.6 describes the procedure for the generation of the time dependent variables that affect tipping. In particular for an asker i and a researcher j at a time point t we define:

- *[Frequent Patron]*$_{i(t)}$: the number of times the asker i has used the service and has taken an answer until the time point t.

- *[Previous Met]*$_{ij(t)}$: The number of times the asker i has received an acceptable answer by the researcher j until the time point t.

- *[Tipped Before]* $_{i(t)}$: The number of times the asker i has tipped any researcher until the time point t.
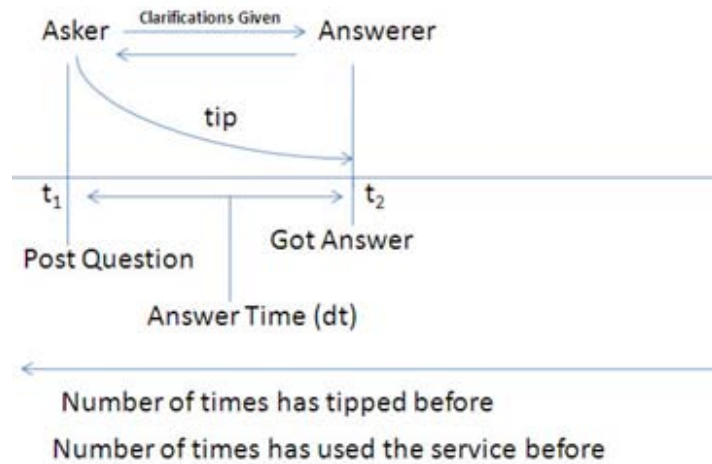
Figure 5.6: Generation of Panel Data variables

- *[Tipped Before]$_{ij(t)}$* : The number of times the asker i has tipped researcher j until the time point t.

- *[Avg. Tip]$_{ij(t)}$*: The average tip that the asker i has given to researcher j until the time point t.

- *[Reputation]$_{j(t)}$*: The average reputation of the researcher $j$which is known until the time point t.

By defining this set of longitudinal variables we are able to analyze past behavior for a future transaction. For instance if an asker has tipped many times in the past then the researchers can have this knowledge by observing his/her history using the system or discussing with other researchers. In fact returning to the initial assumptions we hypothesize that the researchers are fully informed about the past behavior of the asker regarding use of the service and tipping and in some cases adjust their behavior accordingly.

Measuring Service Quality

In order to measure service quality in GoogleAnswers we created an index that con-
sists of a linear relationship between the following two variables: answer evaluation
and the time to answer. The reason that we insisted on the inclusion of the time re-
quired to answer a question instead of only including with the actual rating provided
by the customer its because of the proven inability of customers in general to ratio-
nalize the level of service received and the limited variance that the likert type scale
of answer evaluation has [Nisbett and Wilson, 1977]. Furthermore service quality as
perceived only from the asker is also problematic from a validity perspective since
there is no way to control what was the actual expectation by the asker while using
the service. Low expectations might have transformed a mediocre answer to a good
service therefore giving the reasoning for an asker to reward a researcher with a tip.

Furthermore for developing this index we assume that time is the main indicator of
service quality since other interpersonal characteristics of the interaction between the
researcher and the asker (e.g. the presence and behavior characteristics) are absent
in a virtual environment such as GoogleAnswers.

We define this index as follows:

For an asker i we define the service quality index as

$$QI = answer\_evaluation + \frac{1}{1 + time\_to\_answer}$$

Where the variable answer_evaluation belongs to the discrete set {1,5} and time_to_answer
is measured in hours.

In case the answer was given after 24 hours (Time to answer = 0) after the question
was posted, then the subjective evaluation of the researchers by the askers (depicted
by the answer evaluation) defines solely the quality of the service provided by the
researcher. For example if an answer has received the maximum evaluation (5) and
was answered the same date (time to answer=0) the QI will be QI=5+ (1/1+0) =6

221

which is the maximum value. On the contrary if an answer was given after 30 days, just before the question expired and received an absolute not satisfactory evaluation then the QI will have a value of one.

Service quality is an important parameter for identifying the effect of tipping in this particular interaction setting. In the literature previous studies have addressed the importance of service quality as the main reciprocal parameter for tipping. Lynn and McCall [2000] provided a meta analysis of papers studying the phenomenon of tipping concluding that service quality is an important factor for giving a tip. In our argumentation we expect that for one time users of the service that received a fast and good answer tip and service quality will be positively correlated. While on the other hand if the tip is imposed by a social norm behavior we expect that the significance of the service quality on the size of the tip will be of a lower magnitude compared to the case of service reciprocation.

5.4.2 What drives tipping in general in Google Answers?

In order to get an insight of the factors that affect the tipping behavior of an asker using the platform, we conducted an econometric estimation of the significance of the variables. In that case we consider the tip as a dependent variable depicted by the equation:

$$Tip = \alpha + \beta X + \varepsilon$$

Where α is the constant, β is the vector of the coefficients of the independent variables and ε the error term of the estimation. The vector X contains the independent variables of the estimation which in that case are: the price, the number of references given in the answer, the time to answer the question after it was posted, the evaluation provided by the asker and the length of the answer text provided by counting the number of characters in the text. In that group we also add the time dependent variables provided in section 3.

Furthermore to provide some categorical characteristics to the dependent variable we generated categorical (dummy) variables for the category that the question belongs (one if it is in that category; zero otherwise) as well a general dummy variable that indicates if the tip that was given at that point was greater than the average tips provided at that time by that particular individual. Generating this dummy variable we are able to distinguish whether someone is a generous user or not.

Considering the limitation of the range of the *tip* as the dependent variable ($0 \leq tip \leq 100$) applying a standard linear estimator is not advisable due to the likelihood that the coefficients will take into account cases where tip is greater than 100 and less than 0. Therefore we used three different regression models to estimate the effect of the independent variables namely: (a) A type I Tobit regression model (TOBIT), (b) the probit model (PROBIT) which is a generalized linear model estimator that uses the inverse cumulative distribution function and finally (c) a truncated regression (TRUNCREG) models. The PROBIT model is particularly informative due to the fact the provides an estimator reflecting the binary decision of whether to tip or not while the TOBIT model provides maximum likelihood estimates taking into account the left and right limits of the dependent variable [Amemiya, 1984] which in that case is 0 and 100.

The results of the estimations are presented on Figure 5.5. We used three different non linear regression models, the TOBIT, Probit and Truncated logistic regression models. We represent the categories by dummy variables providing when the question was posted to that general category (1) or not[3] (0).

As can be observed by the (pseudo)R-squared significance, our TOBIT estimation is quite robust predicting 40% of the cases in our dataset. Regarding the regressions estimators, for the sake of inclusion of all the variables the dataset was truncated to 23996 observations due to the fact that: (a) The answer evaluation is optional and many of the askers were not evaluating the answer they were receiving, (b) The

---

[3]As aforementioned there are categories that belong to greater general categories. For instance the subcategory *"Medicine"* belongs to the general category *"Health"*. As such, questions posted in that subcategory were aggregated to the parent general category.

history of the researcher reputation was not available to some cases due to (a), and (c) Dropping of some variables by the regression model (Case of Truncation).

From the estimation results we can extract some interesting findings first of all among the different categories that an answer is tipped. The category that has the highest coefficient and is high significant is the category related with computers and IT (b=0.580**,p<0.01) which is somehow expected since most of the customers of the service are familiar with computer technology and on the other side the high amount of information available online. Other categories that are also significant are the Arts and Entertainment (b=0.580, p<0.05) and the Reference Education and News category (b=0.567, p<0.05).

In all three estimation models the price is highly significant (p<0.001) with however not being the decisive factor for someone to provide a tip as can be seen from the sign of the coefficient on the PROBIT model. On the TRUNCREG estimation which censors the dataset to estimations of tip between 0 and 100 the price receives the highest coefficient (b=0.678***, p<0.001) and is positive providing that the highest is the price the highest the tip will be. That can be also connected with the average ratio of tip to price for the answers tipped in all categories and depicted in Table 5.4. Revisiting the theory binding, the positive relation between price and tip provides ground to the argument that there are significant cases where tipping is given as a result of a social imposed norm on offline settings. This result is quite interesting from the view that although anonymity is secured in an online environment there is still the behavior of a social norm attached to tip giving.

The time required by a researcher to provide an answer for a question seems to have a significant effect (Tobit Estimation) as well on the size of the tip provided. In fact by using the estimation coefficient from the Tobit model (b=-0.114, ***p<0.001) we can see that one day delay to answer can have a 10% negative effect on the size of the tip.

	1	2	3	4	5	6	7	8	9	10	11	12
1.Tip	1.000											
2.Num.of refs.	0.112*	1.000										
3.Time ans.	0.026*	0.043*	1.000									
4.Price	0.241*	0.329*	0.151*	1.000								
5.Clar.given	0.008*	0.053*	0.157*	0.149*	1.000							
6.[Freq.Patron]$_{i(t)}$	0.006*	-0.007*	-0.025*	-0.028*	0.022*	1.000						
7.[Prev.Met]$_{ij(t)}$	0.067*	-0.008*	-0.011*	-0.004*	0.044*	0.538*	1.000					
8.[Tipp.Bef.]$_{i(t)}$	0.09*	-0.013*	-0.015*	-0.032*	0.026*	0.767*	0.438	1.000				
9.[Tipp.Bef.]$_{ij(t)}$	0.104*	-0.005*	-0.006*	-0.003	0.043*	0.455*	0.775*	0.587*	1.000			
10.[Rep.]$_{j(t)}$	0.037*	-0.026*	-0.059*	-0.060*	0.047*	0.063*	0.090*	0.062*	0.078*	1.000		
11.[Avg.Tip]$_{i(t)}$	0.353*	0.052*	0.009*	0.142*	0.050*	0.129*	0.105*	0.168*	0.152*	0.030*	1.000	
12.Ans.Length	0.189*	0.626*	0.063*	0.442*	0.065*	0.001	-0.001	-0.005*	0.007*	-0.021*	0.096*	1.000

Table 5.5: Pair wise correlation Matrix for all the items in our dataset. ($*P < 0.01$)

	Estimator		
	TOBIT	**PROBIT**	**TRUNCREG**
Dependent Variable tip			
Price	0.055***(-0.004)	-0.001***(0)	0.678***(-0.067)
Time to answer	-0.114***(-0.033)	-0.009***(-0.002)	1.626***(-0.291)
Number of references	-0.006(-0.012)	0(-0.001)	-0.051(-0.055)
Clarifications given	1.239***(-0.195)	0.064***(-0.011)	3.776***(-0.993)
[Frequent Patron] i(t)	-0.037* (-0.016)	-0.005***(-0.001)	-0.023(-0.154)
[Previous Met] ij(t)	-0.549***(-0.133)	-0.038***(-0.008)	-1.437(-1.238)
[Tipped Before] i(t)	0.237***(-0.023)	0.027***(-0.002)	-0.03(-0.199)
[Tipped Before] ij(t)	1.172***(-0.214)	0.119***(-0.016)	2.76(-1.682)
[Reputation] j(t)	12.898***(-0.688)	0.732***(-0.036)	-8.268(-6.25)
[Avg.Tip] i(t)	1.062***(-0.024)	0.047***(-0.002)	1.762***(-0.186)
Answer Length	0.000***(0)	0.000***(0)	0.001***(0)
Categories			
Arts & Entertainment	0.711 (-0.814)	0.05 (-0.045)	-7.681 (-7.671)
Business & Money	-2.282** (-0.786)	-0.150*** (-0.043)	13.687 (-7.158)
Computers	0.321 (-0.784)	-0.028 (-0.043)	19.078** (-7.376)
Family & Home	-1.216 (-0.991)	-0.086 (-0.054)	(dropped)
Health	-2.012*(-0.875)	-0.149**(-0.048)	12.133 (-7.657)
Reference, Education & News	-0.104 (-0.817)	-0.022 (-0.045)	4.331 (-7.40)
Relationships & Society	(dropped)	(dropped)	8.776 (-8.586)
Science	-1.51(-0.957)	-0.126* (-0.052)	8.316(-8.846)
Sports & Recreation	-2.535*(-1.204)	-0.137*(-0.065)	-17.91(-12.416)
Miscellaneous	-1.656*(-0.772)	-0.120**(-0.042)	7.99(-7.134
Constant	-5.093***(0.693)	-79.579***(3.282)	-4.224***(0.171)

Table 5.6: Estimation Results for the factors that affect tipping. R-squared (pseudo):0.394
($*p < 0.05, **p < 0.01, ***p < 0.001$)

Another interesting observation can be derived from the interpretation of the co-efficients connected with the other descriptive variables attached to the service. In particular the number of given clarifications which in essence captures interaction be-tween the asker and the researcher is significant both on the decision to tip (PROBIT) and the magnitude of the TIP (TOBIT, TRUNCREG). Requests for clarification between a researcher and an asker usually result to a better answer delivered therefore it is expected that the more the clarification are the better the final answer will be. On the other hand, the answer length as a descriptive variable seems to have no influence at all both on the decision to tip and its size. Hyperlink references added to the text are also another factor that doesn't affect these decisions.

The time dependent variables described in the previous subsection are according to the estimation models significant, providing proof that the decision to generate such variables adds extra validity to our model. The number of times an asker used the service before asking a particular question at time t seems has a significant negative effect in some cases (PROBIT Model) which is also another interesting finding con-sidering that a frequent user of the GoogleAnswers service has already established a history of using the website. On the other hand the reputation of an individual re-searcher (Reputation j(t)) has a highly significant effect both on the decision tip and the size of the tip given. This provides a ground for the consideration of tip as a re-ciprocal case since the reputation index of a researcher is constructed by the ratings that the askers will provide after the acceptance of a given answer. Researchers with a good reputation at the time point t are supposed to have provided good service and the supposedly provided service is a confirmation of their current reputation index. In fact researchers have a high incentive to provide a good service not only for obtaining a tip but also for maintaining or improving their current reputation index. A negative factor however in that case is that the service design allows only for a discrete repu-tation index based on a five point likert scale. However this can be overcome by the fact that every researcher participating in the service has his/her pseudonym linked to

a page presenting its individual rating history containing also comments by the asker who evaluated his/her service.

The previous interaction between askers and researchers has also a significant effect both on the decision to tip and the size of the tip. The most interesting finding in that case is that previous interaction affects this case negatively. The coefficients provide that a customer who has interacted with a particular server (the GoogleAnswers' researcher) will give lower and lower tips the more they interact together. On the contrary we can derive that in the case that customers haven't interacted with a particular server before are expected to tip in order to build a reputation for themselves.

A particular state of interaction can be attributed in case that there was a tip given in the past and the customer is posting a new question at the time point t. We assume that from past experience, the researchers are informed about the size of the tip that a customer gives, similar to the same case that happens also in restaurant settings [Lynn and McCall, 2000]. The coefficients of the models indicate that this is a significant factor both for the decision (PROBIT) as well as for the size (0.237 in TOBIT). The case of prior tipping of a particular asker to a particular researcher ($TippedBefore_{ij(t)}$) has also the same effects and strong significance since the customer is probably reciprocating the good service that the server provided.

Tipping decisions also seem to differentiate across categories with the categories to Business and Money and with Health Matters influencing the tipping behavior. This can be attributed to the attitudes of the participants posting questions to these categories and heterogeneity related with the perceived value of information regarding these matters.

One particular attention needs to be added at that point in relation with the significance of the coefficients that we obtained, one particular attention needs to be added at that point in relation with the significance of the coefficients that we obtained from the analysis above. When it comes to measurement of the significance level of a relation an attention should be given on the sample size related with the significance level

as to when the p value should be important or not. That is because the p value is in absolute relation with the variance of the sample. In our case the sample represents over 90% of the population so it is expected that the variance will be lower (as it is implied by the standard definition of the variance in relation with the population or sample size). In that case with smaller variance it is normal that most (if not all) of the coefficients in a relation will become significant. Therefore we should mostly focus our discussion not on the significance of the coefficients but instead on the implications of their size with the perspective of the research question pursued from the analysis above. When it comes to measurement of the significance level of a relation an attention should be given on the sample size related with the significance level as to when the p value should be important or not. That is because the p value is in absolute relation with the variance of the sample. In our case the sample represents over 90% of the population so it is expected that the variance will be lower (as it is implied by the standard definition of the variance in relation with the population or sample size). In that case with smaller variance it is normal that most (if not all) of the coefficients in a relation will become significant. Therefore we should mostly focus our discussion not on the significance of the coefficients but instead on the implications of their size with the perspective of the research question.

So far we have had a general insight for what drives tipping in general in the GoogleAnswers service. However we haven't provided a particular insight on how service quality affects tipping and the decision of a customer (asker) to give a tip to the server (researcher). We continue our analysis in the following section where we consider tip as an independent variable which affects the service provided in order to see whether tip affects service and the decision to give a tip is dependent on the service provided by a server to the customer.

5.4.3 The effect of tipping to service quality

Having defined the service construct we are going to estimate the effect of tip on the service quality as depicted by the time dependent variables. The general estimation model that we employ is of the form:

$$QI = a + bx + \varepsilon$$

Where the coefficient vector b consists of the time dependent variables. For this estimation we used the standard TOBIT regression truncating it at zero (since the construct definition doesn't allow the quality index to be negative).

TOBIT	
Dependent Variable (QI)	
Variable	**Coefficient (Standard Error)**
[Tipped Before] i(t)	0.008*** (-0.001)
[Tipped Before] ij(t)	-0.034** (-0.013)
[Avg.Tip] i(t)	0.008*** (-0.002)
[Previous Met] ij(t)	0.025** (-0.008)
Constant	5.050*** (-0.012)

Table 5.7: Regression of the Quality index against the variables that characterize tipping history.($*p < 0.05$, $**p < 0.01$, $***p < 0.001$, Pseudo R: 0.056, N = 13030)

Table 5.7 provides the estimation results for the Quality index as a dependent variable regressed with the previous tipping history. For the quantification of the quality index we use the measure described in subsection 5.4.1. The results clearly indicate that the tipping history is significant for the quality of the service. The coefficient sign for all the time dependent variables used in the TOBIT model have a positive sign apart from the previous tipping relation between the customer and the server apart

from the current tipping history of the customer (at the time point that a new question is asked by this particular customer). This provides the main motivation for examining in depth the behavior of an asker in that setting since it provides evidence that customers with frequent use of the service do tip less than those with less history in the service.

In order to analyze this further and in relation with the theoretical ground provided in Section (2) we analyze the individual behavior on tipping in relation with service quality in the following three cases (Reciprocity, Social Norm, Strategic Interaction). In order to provide such a case first we calculate the proportion of tip to the quality index as:

$$TipQI = \frac{Tip}{QI}$$
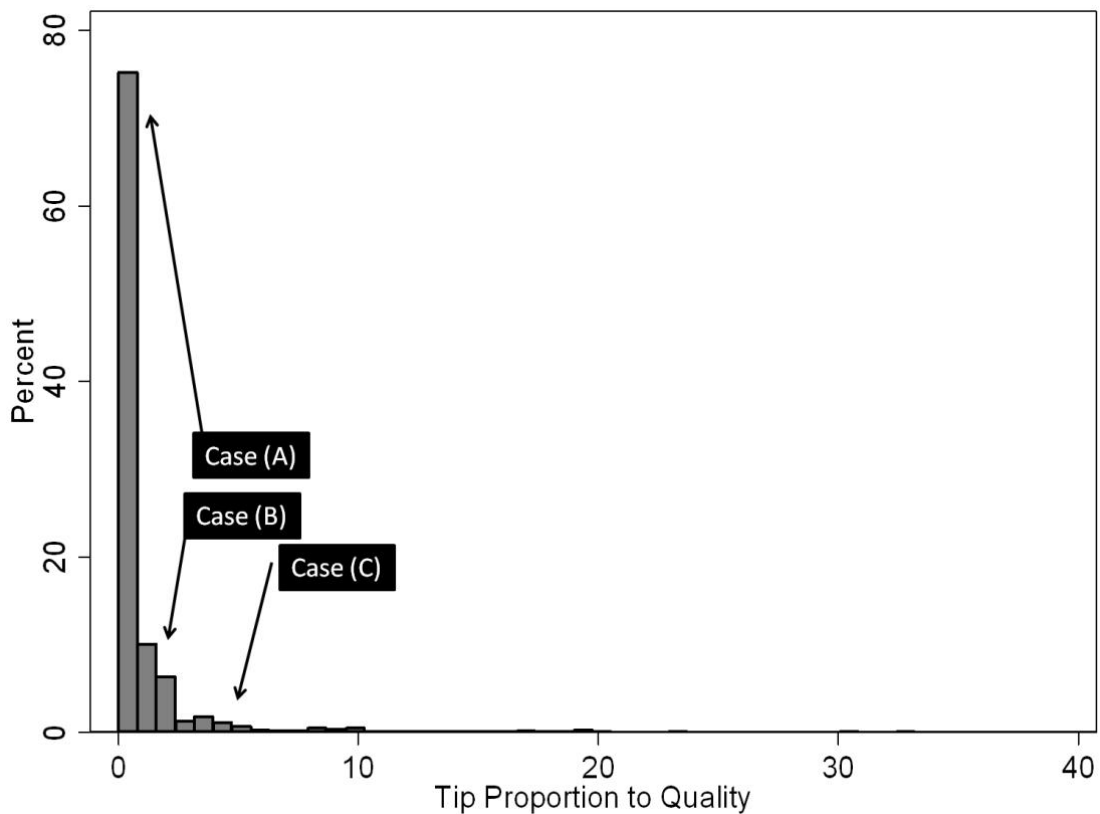
Figure 5.7: Distribution of Tip relative to the quality index

As can be observed in Figure 5.7 we graph the tip proportion to quality (tipqi index) and based on the distribution we divide the dataset of the observation in the following three groups: (a) Very High Tip and proportionally high service (b) Tip and Service relatively constant and (c) Low tip for high service. Case (A) represents the case for reciprocity, case (B) the case for social norm and case (C) the case for strategic interaction.

The case of reciprocity

In order to evaluate the case of reciprocity between the service provided and the tip we examine the relation between the variables of Quality Index and the variables that reflect tipping history. We can hypothesize that (whether given) a high answer evaluation depicts satisfaction of the service and therefore askers reward the researchers by providing a tip. In fact positive reciprocal connection between the answer evaluation and the tip represents the case where askers reward the "treatment" provided by the researchers. The ideal case can be depicted in Figure 5.8. As the service quality increases the customer becomes more satisfied and rewards with a higher tip.

However there might be the case that some customers react differently to higher service quality and reward it with fewer tips (Figure 5.10). This for example might be in the case they have provided a high price for the question they ask therefore they feel that they are generous by providing this reward only. In our case the significance of the price in the quality perceived by the customers can be evaluated by the prediction results provided above.

However the estimates provided by the coefficients do not provide substantial evidence on that (it might be that while the tip increases on increased service quality the contrary case where tip decreases with lower service is not assured. Therefore we are going to use a different method of estimation based on marginal effects of the
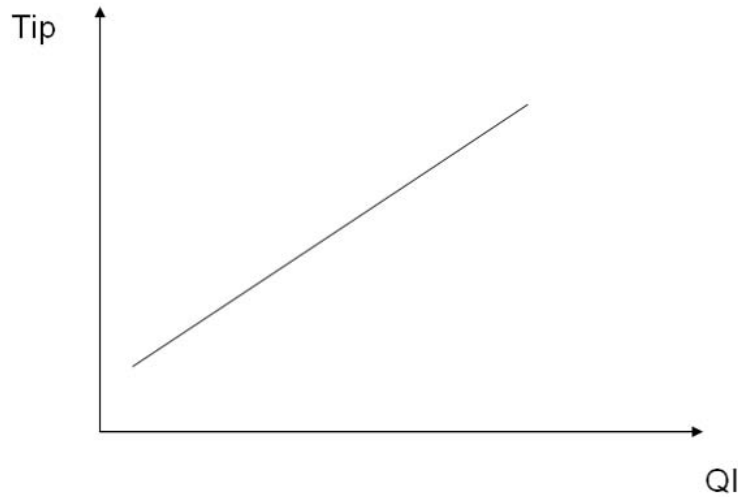
Figure 5.8: An ideal case of reciprocation between tip and quality provided

coefficient estimates. For the Case A we consider the following two estimators:

$$Tip = a_1 + b_1 QI + e_1$$

$$QI = a_2 + b_2 TIP + e_2$$

We ran the regressions and then we calculate the marginal effects of the regression estimators (dy/dx).

Dependent Variable	Variable	Dy/dx	P> |z|	Significant
Tip	Qi	0.6178	0.0190	YES
QI	Tip	.084377	0.0190	YES

Table 5.8: Marginal effects for tip and quality in the reciprocation case (N=8538)

Reciprocation of quality to the tip is quite strong (in order of magnitude) and significant in the 95% confidence interval while reciprocation of tip to quality is on a much lower magnitude.

The case of social norm

The case of tipping as a social norm provides that customers will tip unconditionally of the service quality perceived. This is the case where customers provide a constant amount of tip every time they receive an answer regardless the quality.

Figure 5.9: An ideal case of tipping reward as a social norm

Figure 5.9 depicts an ideal case of tipping behavior in relation with the service quality. Low and high service quality are rewarded with the same tip. In general a fixed response to any level of service quality denotes that the customer is generous in general. In that case we expect that an increase to a standard tip has no *effect to the service quality and the increase of the quality is not related with the increase on the tip*

Dependent Variable	Variable	Dy/dx	P> |z|	Significant
Tip	Qi	0.00009	0.890	No
QI	Tip	.2212789	0.890	No

Table 5.9: Marginal Effects for tip and quality on the social norm case (N=2544)

The TOBIT marginal effects indicate no significance at all (P> \|z\|) both from the

perspective that a marginal increase on the tip has an effect on quality as well as a marginal increase on the quality by a server has an effect on a future tip.

The case of strategic interaction

As aforementioned we consider the case of strategic interaction in the way that a customer can achieve high quality by giving lower and lower tips at a time. However we need to control for the magnitude of these tips (as well as the price) and to examine when past tips have an effect on that particular behavior.
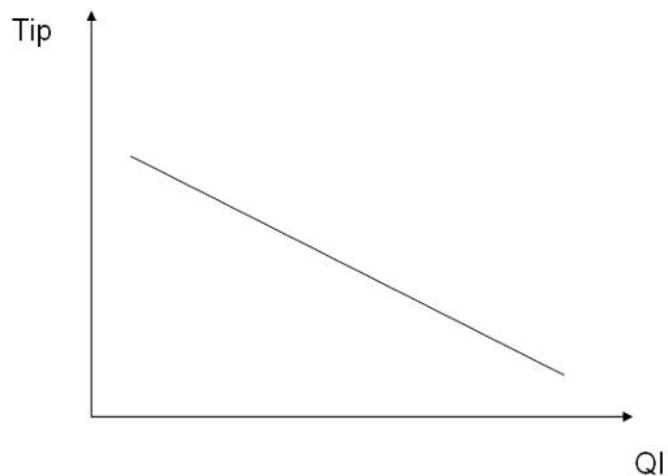
Figure 5.10: An ideal case of tipping used as an incentive mechanism for strategic reward where lower tips result to higher quality

Essentially this will provide ground for examining how an individual can maximize his/her benefit from the service use by essentially giving less effort. Figure 9 depicts such a case where the tip decreases over higher quality thus providing that the asker will receive higher service with less reward on the tip.

The marginal effects estimation presented on Table 5.10 provides an interesting insight on the relation between the variables *QI* and *TIP* and vice versa. In the estimator we control for the price in order to avoid cases of tipping becoming influenced by the price tag attached to the particular answer. In particular we have a stronger significant marginal effect (as compared with the reciprocity) both from a significance and

Dependent Variable	Variable(Controlling for frequency of use)	dy/dx	$P > |z|$	Significant
Tip	Qi	.454	0.000	YES
QI	Tip	**-.0137951**	0.000	YES

Table 5.10: Marginal Effects the TIP and Quality on the strategic interaction case (N=892)

magnitude perspective. In particular increasing TIP by one unit this will increase quality by 0.45 units which is a result of the intrinsic nature of the incentive provided. The other important result that can be extracted from the estimator model is the case that increase in the past quality will result to a slightly lower tip which is highly significant for the post effects estimator.

In the three cases that we have discussed on this section we controlled for the frequency of use and the price as factors that might affect the decision to give as well as the size of the tip. We summarize these findings as well as connections with the literature in the discussion section that follows.

5.5 Discussion

Summary of findings

Returning back to the theoretical discussion on Section 5.2 we revisit the three cases of tipping as it has been described for offline environments and used in our case: (a) Reciprocation of the Service Provided, (b) Adherence to moral codes in order to attain social approval (or avoid social disapproval) even in an online anonymized environment and (c) Tip as an incentive for future service.

Evidence of Tipping used as a reciprocation of the service in an online environment

The results presented in the previous section concerning the case of reciprocity provided that in 71.30% of our subjects (8538) has tipped due to reciprocity. The marginal effect of the constructed quality index to the size of the tip was 0.6178. While the significance obtained from the ex-post estimator is not on the highest level it still can explain the cases falling into the 99.5% confidence interval. This can be attributed to the fact the initial clustering of the three cases made was not exact due to the fact that membership on the first cluster was not exclusive. This was done in order to keep cases where relatively high quality and high tip could also be included.

Number of Cases	Percentage	Use of Tipping
8538	71.30%	Reciprocity
2544	21.2%	Social Norm
892	7.4%	Strategic Interaction

Table 5.11: Summary of the Cases identified in our dataset

Essentially this was due to the fact that we needed to control for other factors that affect tip in general as we saw in Section 3.2 and in particular the effect of the price and previous history both in service use and individual interaction. This provides a ground for the importance of extrinsic rewards as a motivation factor for enhancing interaction on online communities. The prospect of tipping in that case works as a mechanism for improving quality which is highly evident on the cases that we analyzed in this study. That makes the contribution of this study solid with the findings of Benabou and Tirole [2003] from the perspective that explicit stated rewards can have a negative effect on the long run while a non explicit stated reward (as the tip in our case) can have a positive effect on performance which in that case is indicated by the quality index.

Existence of social norms in online environments

What is interesting from this finding is that social norms typically seen in offline environments do exist on a highly anonymized setting as the one in Google Answers. In fact in our study we found that 21.2 % of the cases adhere to a social norm where tip is given regardless the quality of the answer given. One probable explanation can be the cultural background of the users. In US for example for every service transaction customers are expected to provide a tip as a percentage of the price of the food. Surprisingly if we revisit the numbers from Table 5.4 then the average price/tip ratio is 11.68% and accounting for the case of reciprocation and strategic interaction the ratio is well around 10% making it almost the same with the reported price/tip ratio that is used in every restaurant in the US. This is a result that might need additional exploration in online environments and it might be that in such cultural background social norms are so strong that become internalized especially in collective settings [Adler and Kwon, 2002].

Tipping as an incentive to future service

In that case the initially formulated argument that an individual tips strategically in order to maximize the benefit that he/she receives from the service founds solid ground with the analysis that we conducted. We were able both to justify such a case as well as to show with the empirical data gathered that such a behavior exists on an online and anonymized setting where we control for other influential factors of tipping such as the price and the frequency of use. In particular the ex-post marginal effects estimators used for that case provided that gradually decreasing the tip by 0.013 units will have an increase of 0.45 units in the service quality (as measured by the users judgment and the time to get an answer).

Connections with the literature

The context of GoogleAnswers has also been studied by other researchers in the information systems literature. The first paper that appeared with a study of social interactions in GoogleAnswers is the one by Edelman and Draft [2004] where the relationship between earnings and ratings is examined. Rafaeli et al. [2005] and Rafaeli et al. [2007] have published a series of papers where they tackle the effect of price and social incentives (such as previous participation) as a factor that affects the outcome of the interaction and the sustainability of activity. Regner [2004] approaches GoogleAnswers from an economic perspective and in particular on the way such a standard principal-agent relation can induce evidence of social or other regarding preferences in such an online environment. Regarding the service model papers from the information and library science literature as in [Cahill, 2007] and [von Retzlaff, 2006] have addressed possible implications for the use of librarians and library users as possible agents where libraries could provide the facilitator by connecting librarians and users on such type of environment. Our study is positioned differently in the literature from the fact that it tackles with (a) User tactics and behavior – why users in such settings provide tips ? and (b) what are the implications for online communities (from a service viewpoint) where such an incentive reward is used. To this end the study presented on this chapter contributes to the exploitation of such a model from a user perspective as to how he/she can facilitate better service by strategically motivating servers (researchers that are willing to carry out the task on a payment basis).From a service design perspective the use of such mechanisms can enhance the activity as it has also been discussed in [Rafaeli et al., 2007].

From a contribution to the service literature, undoubtedly most of it deals with cases of restaurant tipping. What this study contributes is that is studying tipping on an online environment where the possibility to disentangle the reasons why someone will tip is much better than in field settings.

5.6   Conclusions and further research

This chapter connects with the research objective of this dissertation in multiple ways. First, it presents a complex environment where social capital contributions are made only by one category of users who, in turn, can be incentivized by both explicit (price) and implicit (tip) monetary rewards. The effect of extrinsic rewards can be seen on the high quality (both perceived and observed) received by the customers who make the outcome of the interaction more valuable. Such an environment is interesting to study, both from an economics and social interaction perspective, since the trading of the economic commodity in this case is information produced by searching the vast index of a popular search engine.

One particular extension of the study presented here could be on understanding the users' desire or willingness to pay for information contextualizing his/her preferences by category. In the analysis we have seen the general factors that affect tipping when in the context of information, it can play a role (e.g., both tipping and price were higher in categories such as business and money and computers and internet in contrast with the other general categories).

Another possible extension would be to study the other's judgment of the quality of the contributions provided and the way in which the communication code affects it. Such a study is the one presented in the following chapter where regardless of the incentive mechanism used for deriving the contributions from the users/contributors, other users can evaluate the usefulness of such contributions and the way they affect their own choice decision in an environment where the commodity is codified information (books) which are context independent.

Bibliography

P. S. Adler and S. W. Kwon. Social capital: Prospects for a new concept. *Academy of Management Review*, 27:17–40, 2002.

T. Amemiya. Tobit models: A survey. *Journal of Econometrics*, 24:3–61, 1984.

A. Ardichvili, V. Page, and T. Wentling. Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management*, 7: 64–77, 2003.

K. J. Arrow. The economics of information: An exposition. *Empirica*, 23:119–128, 1996.

R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211:1390–1396, 1981.

O. H. Azar. The history of tippingŮfrom sixteenth-century england to united states in the 1910s. *Journal of Socio-Economics*, 33:745–764, 2004a.

O. H. Azar. The implications of tipping for economics and management. *Economics*, 30:1084–1094, 2003.

O. H. Azar. The social norm of tipping: A review. *Journal of Applied Social Psychology*, 37:380–402, 2007.

O. H. Azar. What sustains social norms and how they evolve? the case of tipping. *Journal of Economic Behavior and Organization*, 54:49–64, 2004b.

R. Benabou and J. Tirole. Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70:489–520, 2003.

G. E. Bolton and A. Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90:166–193, 2000.

B. Butler, L. Sproull, S. Kiesler, and R. Kraut. Community effort in online groups: Who does the work and why? *Leadership at a Distance: Research in Technologically-Supported Work*, 2007.

K. Cahill. Worth the price? virtual reference, global knowledge forums, and the demise of google answers. *Journal of Library Administration*, 46:73, 2007.

R. B. Cialdini and M. R. Trost. Social influence: Social norms, conformity, and compliance. *The handbook of social psychology*, 2:151Ű92, 1998.

B. Edelman and P. Draft. Earnings and ratings at google answers. *Unpublished Manuscript*, 2004.

E. Fehr and U. Fischbacher. The nature of human altruism. *Nature*, 425:785–791, 2003.

L. Festinger. A theory of social comparison processes. *Human Relations*, 7:117, 1954.

H. Gintis. The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. *Journal of Theoretical Biology*, 220:407–418, 2003.

B. Gu, P. Konana, B. Rajagopalan, and H. W. M. Chen. Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research*, 18:68, 2007.

KHEE Lee and CB Hatcher. Willingness to pay for information: An analyst's guide. *Journal of Consumer Affairs*, 35:120–140, 2001.

M. Lynn and A. Grassman. Restaurant tipping: an examination of three rational explanations. *Journal of Economic Psychology*, 11:169–181, 1990.

M. Lynn and B. Latane. The psychology of restaurant tipping. *Journal of Applied Social Psychology*, 14:549–561, 1984.

M. Lynn and M. McCall. Gratitude and gratuity: a meta-analysis of research on the service-tipping relationship. *Journal of Socio-Economics*, 29:203–214, 2000.

M. Lynn and K. Mynier. Effect of server posture on restaurant tipping. *Journal of Applied Social Psychology*, 23:678–685, 1993.

R. E. Nisbett and T. D. Wilson. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84:231–259, 1977.

J. Preece. *Online Communities: Designing Usability and Supporting Socialbilty*. John Wiley & Sons, Inc. New York, NY, USA, 2000.

J. Preece, B. Nonnecke, and D. Andrews. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20:201–223, 2004.

S. Rafaeli and DR Raban. Experimental investigation of the subjective value of information in trading. *Journal of the Association for Information Systems*, 4:119–139, 2003.

S. Rafaeli, G. Ravid, and V. Soroka. De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, page 10, 2004.

S. Rafaeli, D. Raban, and G. Ravid. Social and economic incentives in google answers. In *ACM Workshop Sustaining Community: The role and design of incentive mechanisms in online systems, Sanibel Island, FL USA*, 2005.

S. Rafaeli, D. R. Raban, and G. Ravid. How social motivation enhances economic activity and incentives in the google answers knowledge sharing market. *International Journal of Knowledge and Learning*, 3:1–11, 2007.

T. Regner. Why voluntary contributions? google answers! 2004.

RR Reno, RB Cialdini, and CA Kallgren. The transsituational influence of social norms. *Journal of personality and social psychology*, 64:104–112, 1993.

B. J. Ruffle. Gift giving with emotions. *Journal of Economic Behavior and Organization*, 39:399–420, 1999.

T. S. H. Teo, V. K. G. Lim, and R. Y. C. Lai. Intrinsic and extrinsic motivation in internet usage. *Omega*, 27:25–37, 1999.

H. R. Varian. Pricing information goods. *Coyle's Information Highway Handbook: A Practical File on the New Information Order*, 1997.

L. von Retzlaff. E-commerce for library promotion and sustainability: how library technicians can market themselves and their library's services online. *Australian Library Journal*, 55:102, 2006.

E. Walster, G. W. Walster, E. Berscheid, W. Austin, J. Traupmann, and K. Mary. *Equity: Theory and Research*. Allyn and Bacon, 1978.

S. Whittaker, L. Terveen, W. Hill, and L. Cherny. The dynamics of mass interaction. *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 257–264, 1998.

Evaluating Content Quality and Usefulness of Online Product Reviews

Online product reviews are an important resource for consumers in online market-places because they provide a useful source of support information during the purchase of goods. Furthermore, in some online marketplaces consumers have the opportunity to assess the usefulness of a review by using a dichotomous evaluation form provided by the online marketplace. These evaluations produce a usefulness score, which can be calculated as a fraction of helpful votes out of the total votes that a review has received. This enables testing hypotheses regarding the factors that affect the usefulness of reviews which may, in turn, be used as metrics if evidence supporting the connection between the factor and the usefulness is found. This chapter reports an empirical study test applied to a large dataset of reviews collected from the United Kingdom section of the popular online marketplace, Amazon, to explore the connections between readability and usefulness. The results of these evaluations point out that usefulness is significantly affected by the qualitative characteristics of the review as measured by readability.

6.1 Introduction

Undoubtedly, the use of the Web as a source of information affects several areas of human activity, and the way that transactions in markets occur is no exception [Malone et al., 1987]. The ability to provide more relevant information to consumers has also influenced the product-choice process in electronic markets. Nonetheless, the most profound advantage to merchants of the product-choice and purchase process in online marketing is that it enables encapsulating and promoting the opinions of their customers for the products that they have purchased. This process has led to a massive number of online reviews that consumers may use to become better informed about the product or service that they are considering purchasing.

The influence of online reviews on an individual's choice of a certain product on the internet has been the subject of several recent studies in the literature [Chevalier and Mayzlin, 2006, Pavlou and Dimoka, 2006, Hu et al., 2006]. The effect of online reviews in purchasing behavior or awareness [Duan, Gu, Whinston, 2008] has been acknowledged in these studies. Additionally, users of Web 2.0 applications have been found to be more strongly influenced by the opinion of peers [Hee Shin, 2008], which reinforces the value of online reviews. Nelson [1970] defines the concept of an experience good as a product or a service, the quality and utility of which can only be determined upon consumption. Therefore, in order to make a decision to consume or purchase this product or service, a consumer has to rely on previous experiences that will provide an indicator whether this product or service is worthy for purchase or not. One example of an experience good is a book where the utility that the consumer receives by reading this book can be extracted only after reading it. Therefore, in experience goods such as books, the producers (e.g., the publishers) often use reviews by authoritative sources such as literature experts to provide an opinion and endorse the book, so that by trusting these sources the consumers will continue to purchase this product.

However, the inclusion of prior experience in the promotion of experience goods

can actually pose a problem for consumers, mainly due to following factors:

- The cost to the producers of publishing experiences by previous consumers, especially in cases where negative views might reach new customers, makes the producers unwilling to do so due to the risk of negative word of mouth [Richins, 1983].

- The obvious search costs that arise for a consumer in order to search for, acquire, and evaluate the prior experiences [Stigler, 1961] are high.

- The variability between the different versions of the same product may trouble a consumer (e.g., an mp3 player with a large set of characteristics vs. a simpler version of the same mp3 player) during a purchase decision.

The development of Internet marketplaces where consumers can establish interaction has undoubtedly affected the way that a review (as an expression of prior experience) influences the way consumers make a choice about a product or a service based on prior experiences [Clemons et al., 2006, Dellarocas, 2003]. First and foremost, the use of online mechanisms for reporting and categorizing reviews of a product or a service, in conjunction with the development of modern search engines, has eliminated the search costs for the consumers [Lohse and Spiller, 1998]. Online marketplaces such as the popular bookstores Amazon.com[1] and Barnes and Noble[2] enable a consumer interested in purchasing a product to read a series of reviews about it. Furthermore, apart from the description of the experience derived from the purchase of the reviewed product, a consumer is also able to rate the usefulness of the product, usually on a standard Likert scale.

Mining and processing the characteristics of online reviews can greatly enhance the adoption of online marketplaces for future purchases [Hu and Liu, 2004]. The latter comes into connection with an important field from the marketing literature,

---

[1] http://www.amazon.com
[2] http://www.barnesandnobles.com

which has to do with the referral value of a specific product. In particular, in word-of-mouth scenarios [Brown and Reingen, 1987] consumers refer a product or a service to fellow consumers usually with enthusiasm (if they are satisfied), or regret if they are unsatisfied). The extent to which the referral value of the information provided by a consumer of a product might affect the opinion of a future customer still remains an issue to identify [Borenstein and Saloner, 2001].
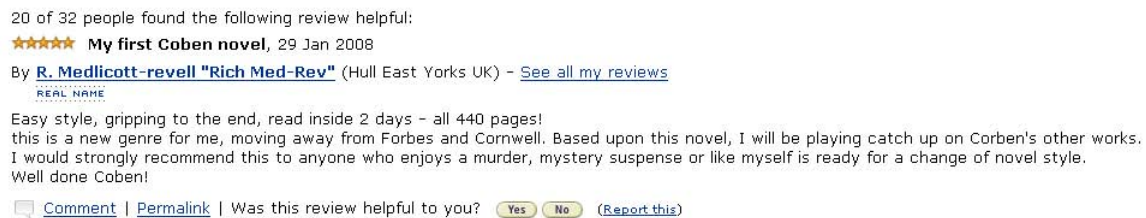
Figure 6.1: The interface of the review evaluation mechanism used in this study

Several approaches for exploiting consumer reviews can be found in the literature. For example, Miao, Li and Dai [2009] developed a review retrieval system combining information retrieval and data mining techniques.

In this study we tackle the qualitative or textual characteristics of online consumer reviews, concretely, using readability as a key element in the transfer of past experiences to other customers. The fundamental assumption that is made in this study is that a review submitted by an individual reflects his or her experience of using the product. Therefore, the underlying assumption is that individuals who submit reviews have consumed the product and are in a position to report their own personal experience or judgment of the product since they have already experienced it (negative or positive depending on the value of the review rating). Furthermore, the review acts as a "justification" of the rating so the potential buyer can evaluate whether the review was fair or not. This evaluation is reflected in the text that is submitted along with the rating. In addition, online marketplaces use a way of meta-rating unfair reviews so that interested buyers can evaluate how helpful a review was during the choice process. In such a case, the review text acts as the main source of evaluation of the

usefulness of a specific review by other consumers.

However, reviews by individual consumers often express a personal view of their experience with the product and thus might differ from the expectations of the interested buyer. For example, it might be that a reviewer expects a book to contain more action, whereas an interested buyer does not care about that specific characteristic. Nevertheless, in order to evaluate the usefulness of the review, someone has to read it first. Therefore, the style and the readability of a review might actually play a role in how its usefulness is evaluated.

In this chapter we seek to evaluate how the style and the comprehension of a review, as depicted by a readability test, might affect its usefulness - the number of people that found this review useful out of the total number of people that read and evaluated it. In order to investigate this issue, we employ the readability metrics applied to a dataset of reviews with their meta-evaluations collected by the bookstore section of Amazon in the United Kingdom. Figure 6.1 depicts the interface of the review evaluation mechanism that was used in this study.

The major contribution of this study is to provide an evaluation of the impact that the qualitative characteristics of a review might have on whether consumers that are interested in buying a product or a service from an online web store consider this codified piece of information as useful or not. We show that apart from the review score that a particular review provides, consumers also evaluate its importance by how closely this review matches their communication code, which is denoted by the way the review has been written. This communication code is depicted by the qualitative characteristics of the review text which are measured by readability tests.

To this end, this chapter is structured as follows: Section 6.2 provides a background on readability tests and the way they are calculated, as well as on the meaning of their norms. Section 6.3 presents the actual analysis of the dataset to which we applied the readability tests, following a discussion of the results and their implications 6.4. Finally, Section 6.5 summarizes the conclusions of this study and the challenges for

future research.

6.2 A background on readability tests

The concept of readability describes in general terms the effort that is needed by a person to understand and comprehend a piece of text [Zakaluk and Samuels, 1988]. In a more formalized manner, a readability test is i a formula that is the result of a linear regression applied to subjects regarding the reading ease of different pieces of text that the subjects were asked to comprehend using specific instruments. The objective of a readability test is to measure on a categorical or continuous scale how difficult it is for readers to comprehend a piece of text in conjunction with the linguistic characteristics of that text. A readability test can only provide an indication of how understandable a piece of text is based on its syntactical elements and style. Most of the readability tests in the literature represent a school grade level that is required in order to comprehend the piece of text provided. Apart from their actual application in the field of education, readability tests have also been applied to cases where the subject is required to read a piece of information to make a decision or comprehend the logic behind the actual outcome, such as in the case of software documentation [Lehner, 1993].

It is fair to say that the attention an interested buyer might give to a review can to a large extent be associated with its readability. In that case, the assessment of a review by a readability test can provide an indication whether someone who evaluated a particular review as useful was actually able to comprehend the piece of text that was submitted with the review. On the other hand, we might expect to see that readers of some reviews that were not considered helpful might have been affected by the readability of the text as well as its content.

However, the use of a readability test has some major limitations which we should take into consideration during the analysis of the results of this study. In particular,

Readability Measure	Score Range	Measurement Implications
Gunning-Fog Index	1-12	Indicates the grade level of the education scale. The lower the grade the more readable the text
Flesch Reading Ease Index	0-100	Scores above 40% make the text understandable by practically everyone. As the value of the index decreases, the comprehensiveness of the text becomes more difficult.
Automated Readability Index	1-12	Indicates the grade level of the education scale. The lower the grade the more readable the text
The Coleman-Liau Index	1-12	Indicates the grade level of the education scale. The lower the grade the more readable the text

Table 6.1: The readability tests that we used in this study

the result of a readability formula cannot tell us whether the content of the review expresses personal views on the product or contains some gender, class, or even cultural bias. Furthermore, to avoid cultural background variance, and to a large extent language proficiency, we collected the reviews only from the United Kingdom store of the online marketplace in order to have only native English speakers and as much geographical concentration of the population as possible (population from one country)[3].

Readability tests to study qualitative characteristics of several types of texts have been applied to several areas in information science, and a large set of readability indexes has been developed over the years [Paasche-Orlow et al., 2003]. For our study we selected four major readability texts which individuals on various educational levels have used extensively to evaluate the readability of a piece of text.

Table 6.1 lists the readability tests that we used in our study. These are the Gunning-Fog Index, the Flesch/Kincaid Reading Ease, the Automated Readability Index (ARI) and the Coleman-Liau Index. The major reason for selecting these readability metrics is the availability of software to provide reliable measurement of this indexes (The GNU style command). All four tests evaluate the readability of a text by consistently decomposing the text to its basic structural elements, which are then combined using the empirical regression formula. An important issue of a readability test is that it can be used to evaluate only texts of a certain length since a reader's ability to comprehend a text also involves cognitive properties that are beyond the scope of this study. The logic behind the calculation and the norms of these instruments is described in the sections below.

6.2.1 The Gunning-Fog Index

The Gunning-Fog index [Gunning, 1969] provides a measure of how well an individual with an average high school education is able to comprehend the evaluated piece of

---

[3]Furthermore, the Readability tests used in the analysis section of this study have been developed only for the English language, and language proficiency undoubtedly affects their validity.

text.The approach to compute this index was the following:

- For each review we calculated the average number of words per review sentence on a 100+ word review passage. This gives as the average sentence length (L).

- We then obtained the number of difficult words (D)-that is words that have more than three letters-by excluding proper nouns, compound words, and common suffixes.

- We finally added the average sentence length to the number of the difficult words.

The following equation describes the empirical relation in the Fog Index.

$$Fog = 0.4 \times \left(\frac{Words}{Sentence} + 100 \times \left(\frac{N(complex\_words)}{N(words)} \right) \right)$$

An obvious difficulty in measuring the Fog index for a given text is the evaluation of the number of complex words. In our analysis we considered a word as complex if it had more than two syllables.

6.2.2 The Flesch Reading Ease

The Flesch Reading Ease index [Flesch, 1951, Kincaid et al., 1975] is a readability test that uses as a core linguistic measure the number of syllables per word and the number of words per sentence in a given text. The Flesch test is used to evaluate the complexity of the text in order determine the number of years of education needed for someone to understand it. The following equation describes the calculation of the Flesch-Kincaid score for a given text:

$$FK = 0.39 \times \left(\frac{total\_words}{total\_sentences} \right) + 11.8 \times \left(\frac{total\_syllables}{total\_words} \right) - 15.59$$

The variables *total_words, total_sentences* and *total_syllables* denote the total number of words, sentences, and syllables, respectively, found in the text. For calculating

the Flesch score of a particular review we decomposed the text into sentences, then words, and finally into syllables, which were combined using the constants presented in the formula above. It can be easily inferred from the mathematical expression that the sorter is the number of words per sentence—the fewer words per sentence, the better the readability score that the Flesch test will provide.

6.2.3 The Automated Readability Index

The Automated Readability Index (ARI) differs from the Gunning-Fog and the Flesch-Kincaid tests in that it uses simpler metrics to evaluate the readability of a typical English language text. In order to calculate the ARI for a given review we first calculated the total number of characters (excluding standard punctuation such as hyphens and semicolons) and the total number of words.

$$ARI = 4.71 \times \left(\frac{characters}{words} \right) + 0.5 \times \left(\frac{words}{sentence} \right) - 21.43$$

The calculations for the ARI involved the same steps as for the Fog and Flesch indexes where additionally the number of characters, that is the review length, had to be calculated as well. The ARI can provide an indication of the impact of the review length on the readability of the review.

6.2.4 The Coleman-Liau Index

The Coleman-Liau Index [Coleman and Liau, 1975] is similar to the Automated Readability Index, the only difference being that the second part of the formula considers a more careful selection of the textual characteristics of the evaluated piece of text. The CL index has been developed specifically for machine-based scoring, thus the calculations that it involves are quite tiresome to do by hand. The following formula

describes the Coleman-Liau index.

$$CL = 5.89 \times \left(\frac{characters}{words} \right) - 0.3 \times \left(\frac{sentences}{word} \right) - 15.8$$

The calculation of the Index considers fragments of sentences of 100 words multiplied by a constant (0.3).

6.3 Analysis and results

Having provided a background on the readability tests that we are going to use, we continue to the analysis of the reviews in our dataset in order to test whether the readability tests can actually give us an indication how the qualitative characteristics of a review influence its usefulness for a consumer.

6.3.1 Data collection and definition of variables

In order to apply the readability tests that we discussed in the section above, we developed a web crawler to capture the content of the book section of Amazon UK. The crawler consisted of two parts: (a) A web client to randomly pick items from the front page of the bookstore and (b) a client to the web-service interface provided by Amazon (AWS) where the data for the particular item were collected[4]. The list of books was stored in a relational database which we used for further processing of the reviews expressed in each individual product page. We omitted from the database those books for which the publication date was older than 6 months or had no rating. Furthermore, we excluded books at special offers or discounts to control for price or bargain effects.

The reason that we picked Amazon UK in order to obtain the dataset used in this study is the degree of language homogeneity among reviewers and consumers which

---

[4]The Amazon Web Services API is provided by Amazon to developers and resellers and is publicly accessible at `http://developer.amazonwebservices.com/`. The API version that was used was the 2008-04-07

255

Variable Code	Variable description
Productid	The id of the product that this review is written for. It is used to control for the publication date and other product characteristics
Summary	The summary / title of the review
Content	The actual content of the review. To be used for content analysis.
revieworder	The order that the review appears on the product review page.
Reviewpage	The page that the review appears (default setting is five reviews per page).
rating	The rating that this review justifies, measured on a 1-5 Likert scale.
totalvotes	The number of total votes that have been given to this review.
helpfulvotes	The number of votes that consider this review helpful.
reviewerid	The id of the customer used to control if the customer is a professional reviewer or not.

Table 6.2: The main variables of the initial dataset collected by using the web crawler

might play a role in the comprehension of a text. This is important because readability tests are useless if a reader is not a native speaker of the language in which the text is written. This is due to the fact that many languages differ in syntactical form, and the style of the language in the review might be totally different from the reader's native language.

Table 6.2 provides a description of the variables of our dataset. These variables can be categorized into two groups: the numerical expressions of the review (rating, totalvotes, helpfulvotes), and the textual or qualitative characteristics (summary,

content) including the identifiers and exposure (productid, reviewerid, revieworder, reviewpage).

A particular issue with the dataset that we collected was that of by-passing promotion-backed items such as bestsellers. Since these items are more accessible to the visitors to the online bookstore, there is always a selection bias towards the more visible items. This placement may result in a high exposure of recent product reviews in contrast to older ones. In order to avoid that bias, the web crawler kept a list of the frequency of the items that were displayed in the front page and randomly chose items listed by categories.

Our dataset contains in total seven variables and two identifiers. The *reviewerid* actually provides the id of the customer in the online bookstore's central database. By using this identifier we can group the reviews by customer since a customer may have submitted reviews for more than one product, in this case books. The *productid* is the unique product identifier provided for a product. With this identifier we can group the reviews by product and check for variances between products of different categories.
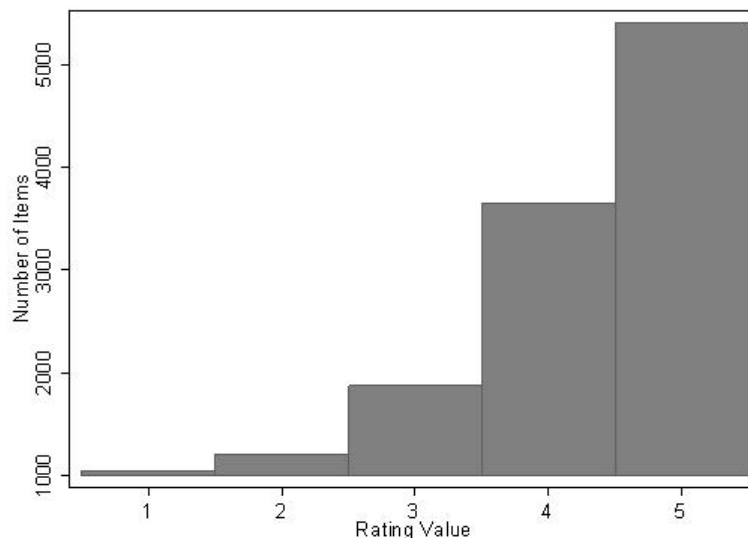
Figure 6.2: Distribution of the rating values among the items in our dataset. (Total of reviewed items/books: Nb=7320

We define the usefulness ratio of a review (UR) as the fraction of the votes that
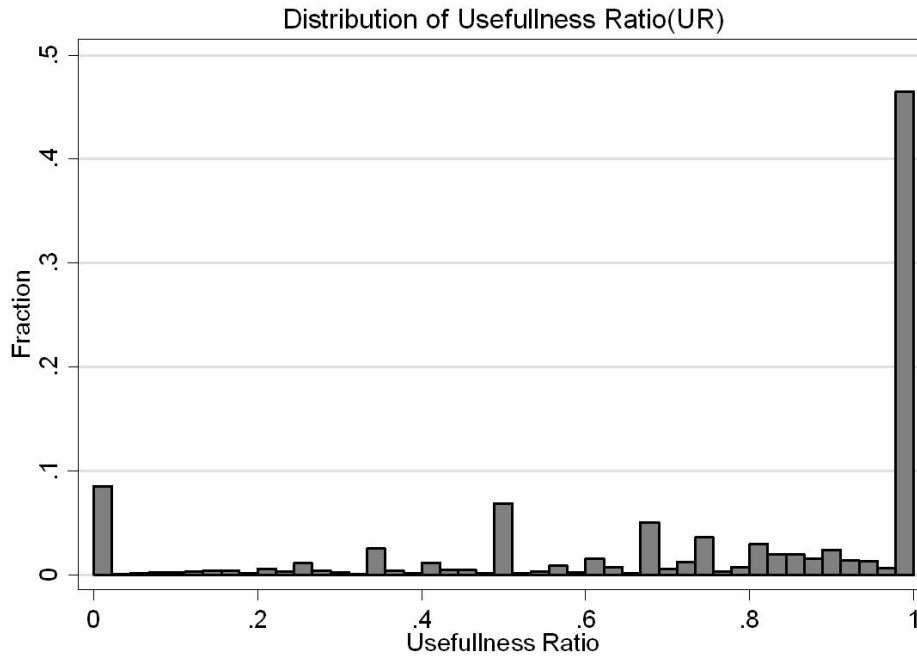
Figure 6.3: The distribution of usefulness scores on our dataset plotted by density (N=37221)

considered this review helpful *(helpfulvotes)* divided by the total number of readers that evaluated the usefulness of the review *(totalvotes)*. Thus we have the dependent variable for our analysis defined as:

$$UR = \frac{helpfullvotes}{totalvotes}$$

The usefulness ratio is in fact a measure of the quality of the review as considered by the readers themselves. From that definition it is easy to infer that the bigger the number of helpful votes a review receives from those that evaluated the review, the higher will be the usefulness ratio.

However, since the number of total votes that a review has (that is the minimum number of readers) may affect the consistency of the metric, we need to keep control for the exposure of this review since some reviews at a certain period of time receive more exposure than others. Typically, this exposure is affected by time since the
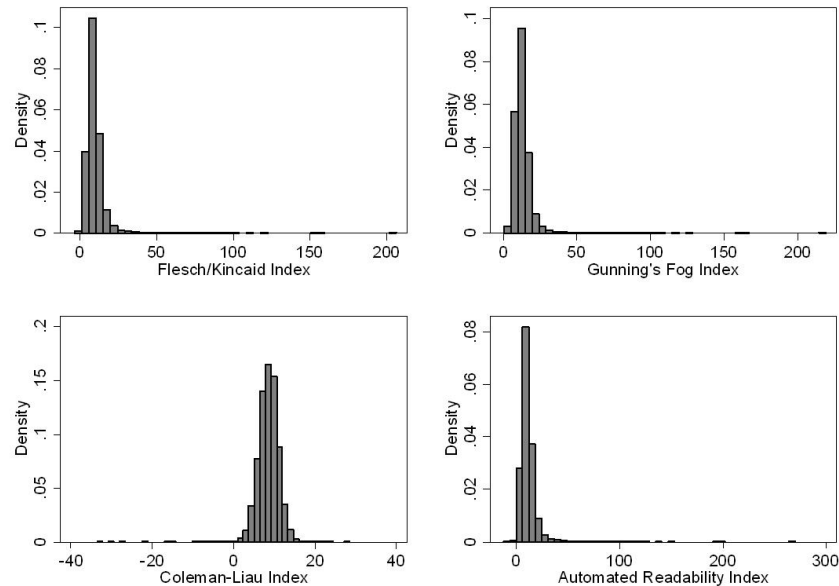
Figure 6.4: Histograms of the distributions of the four readability tests used on our dataset.

system displays first the most recent reviews for a product (in this case the book). In our study the particular exposure of a review was measured by keeping a set of two variables for the pagination results. In particular, the variable reviewpage indicates whether this review was at the first, second, or third page at the time the review was retrieved. The same applies to the revieworder, which controls the display order for a particular review on a particular page. Combining the two variables (reviewpage, revieworder) into a new compositional variable, we are able to control for the review exposure on the website during the time the review was posted.

For example, if a review appears on page 2 and was ordered as third in the page then the exposure value is 23, and so forth. It is generally assumed that reviews which appear on the top of a page get much higher exposure than a review that appears at the bottom since visitors' attention is captured by elements that are displayed in the beginning of the space under the product description.

On the other hand, we don't have a variable that justifies the actual exposure of a review and, in particular, the number of people that read the review. However, in
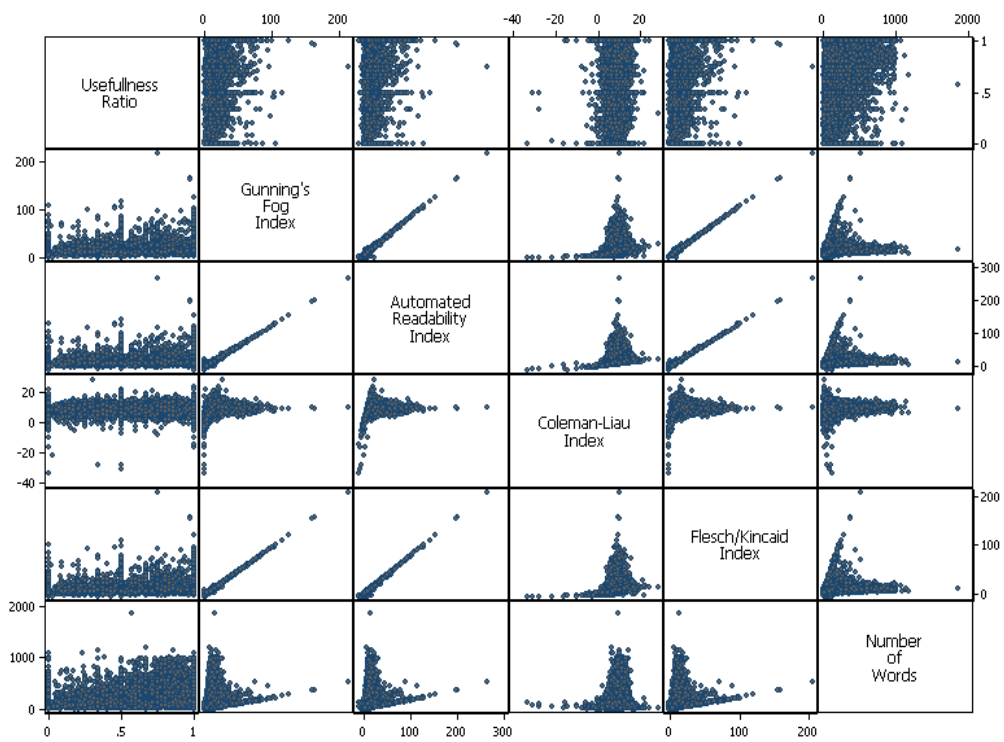
Figure 6.5: Scatter plot matrix of the usefulness ratio and the qualitative characteristics of the text of an online review.

order to hold our analysis to an acceptable level we assume that the total number of people that evaluated the usefulness of a particular review is the minimum number of the readers who read it. In that way we get an indication whether a review has been read by a high number of visitors since it is assumed that those two numbers are positively correlated.

The dataset consists of 38,366 reviews where the total votes (totalvotes) were greater than zero, which means that the reviews on our dataset have been evaluated for their usefulness at least once. Figure 6.3 provides an overview of the distribution of usefulness scores on our dataset. It is interesting to note that around 47% (total of: 17,695) of the reviews have received a perfect score from the readers, which provides that around half of the reviews were very highly acclaimed by their readers. The result is that for this group of particular reviews, the number of helpful votes is the same as the number of potential buyers that have read the reviews. On the other hand, we find that approximately 9% of the votes (total of: 3,292) found the reviews to be totally non-useful for their readers, receiving an absolute 0 of helpful votes. As can be observed in Figure 6.3, much of the variance in the usefulness score happens between the 0.8 and perfect (1).

In total, 7,320 books are covered by the reviews. Figure 6.6 depicts the distribution of the average usefulness score per rating scale value for the items covered by our dataset.

Figure 6.2 presents the distribution of the rating scores that were given to all reviews on our dataset. It is interesting to note that more than 70% of the reviews are highly positive (rating>=3). This might be explained by the fact that most of the reviewers were more than satisfied with the books they read and therefore provided a review. It is likely that the unsatisfied customers were not willing to report their experience with that particular book.

The variable we used to apply the readability tests was the *review's content* as presented on the website, encapsulated by the content and summary variables re-
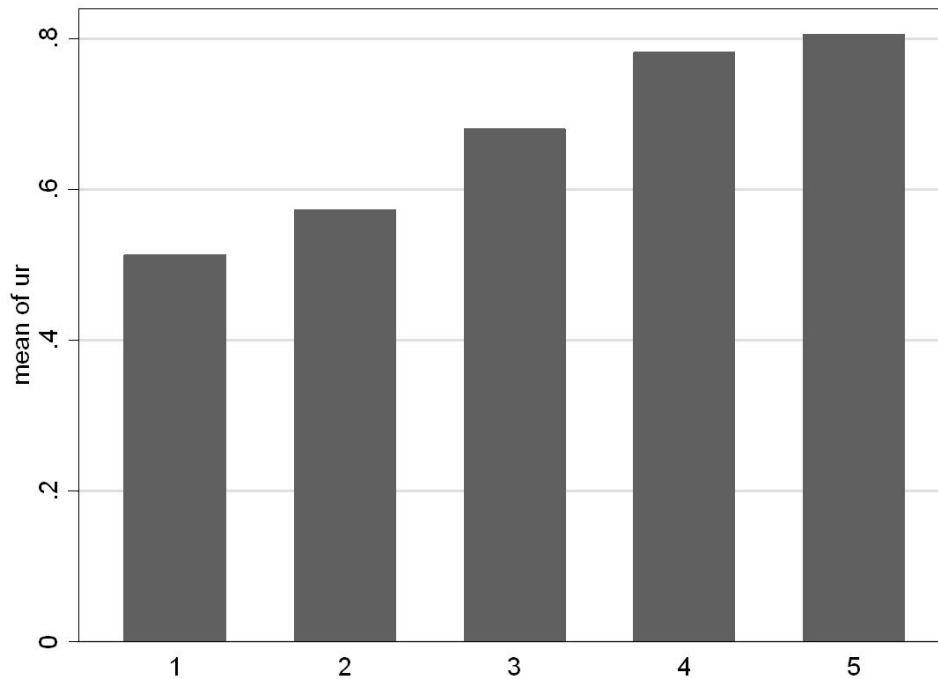
Figure 6.6: Distribution of the rating scores and average usefulness in our dataset

spectively. To obtain the results of the readability formulas described in Section 2, we used the "style" command part of the GNU-Dict package[5]. Figure 6.3 provides a scatter plot matrix showing the relation between the readability scores and the number of words provided in the review text.

As aforementioned, the usefulness of a review is measured as the percentage of the votes cast as helpful divided by the total votes on the usefulness of this review.

Table 6.3 provides the descriptive statistics for the main variables on our dataset. On average the rating of a review was 4 (a number that can also be derived from Figure 5) with the average text length of 160 words. The usefulness ratio had a mean value of 0.750 or 75.0% providing that three out of four reviews were considered useful by the visitors to the website. On average a review received (9.16-7.63 = 1.53), almost two negative votes out of the average nine consumers that read this

---

[5]The version of the GNU Dict that is currently available is 0.7 and can be downloaded from http://www.gnu.org/software/diction/diction.html

Variable	Mean	Std. Dev.	Min	Max
Usefulness Ratio	.750	.322	0	1
Helpful votes	5.03	7.63	0	230
Total Votes	6.61	9.16	1	372
Realorder	44.03	104.58	11	1735
Rating	4.08	1.23	1	5
Number of Words	160.07	135.84	3	1854
Coleman-Liau Index	8.55	2.44	0	28.5
Automated Readability Index	26.41	3.40	0	263
Gunning Fog Index	.99	.010	0	214
Flesch-Kincaid Level	9.81	6.25	0	206

Table 6.3: Descriptive Statistics for the variables that we use to estimate the usefulness of a review (N=37221)

review. As aforementioned, we consider the number of total votes as the minimum number of consumers that read the review and considered it in their decision to select the particular item.

6.3.2 Analysis and results

In order to evaluate the reliability of the scale we are using to evaluate the usefulness of the review, we used Cronbach's alpha statistic [Cronbach, 1951] to check whether the usefulness score of a particular review could be reliably evaluated by the readability tests that we run against its textual properties. In our calculations for the dataset of N=37221 observations we obtained a scale reliability coefficient of 0.87(87%) which is well beyond the minimum requirement of 0.7 [Nunnally, 1978]. The average inter-item covariance that we obtained for four items (the usefulness score against the readability tests) was $\approx$ 6.98, suggesting that no scale item should be removed. The high reliability score of our scale indicates that our dataset holds for internal consistency and can be used to further examine the impact of the items of which it is composed.

Table 6.4 presents the inter-item correlation matrix that we obtained from running Pearson product moment correlations between the items in our dataset. The correlation coefficients were obtained by doing a pair-wise correlation between the variables and asking for a confidence interval of 1% (P<0.01). By looking the sign of the coefficients in the first column we can get some interesting information. In particular, the higher the exposure of the review, (realorder) the lower its usefulness. In fact, by looking more carefully on the relation between the usefulness ratio and the exposure of a review we see that the more exposed the review, the fewer helpful votes the review will receive where at the same time the coefficient of total votes is positive.

The standard length of the review (measured in words) as a pure textual property seems also to play an interesting positive role on the usefulness of the review, since the consumers tend to evaluate as more helpful a review that is more detailed so the review length will be longer.

It is interesting to note that the four readability tests we have used have a high inter-item correlation, which can be explained by the fact that they evaluate the same piece of text. As can be seen from the distribution of the readability scores obtained (Figure 6.2), the variance of the scores follows a similar distribution apart from the Coleman-Liau index which seems to be more sensitive to the syntactical elements.

As can be seen in Table 6.5, we ran three different types of regressions to check whether the UR is affected by the following variables: rating, helpfulvotes, realorder, and the readability tests (fogscore, fleschscore, ariscore). The variable totalvotes was dropped from the regression models due to multi collinearity. As can be inferred from the coefficients of all three models (standard least squares, logistic, and truncated regression) the readability scores as well as the length of the review affect the usefulness ratio in a highly significant way ($p < 0.001$).

In order to test whether the characteristics of a review actually differ depending on the usefulness ratio, we split the dataset into several categories based on the usefulness ratio that the reviews received as well on their rating value.

In particular, the dataset was split into a total of three group categories by the following criteria:

1. If the review's usefulness score was less or more than 0.5 (i.e., the amount of the helpful votes the review received was less or more than the number of those that read the review and didn't consider it useful)

2. If the review's usefulness score was less than 0.25 and more than 0.75 (i.e., the amount of the helpful votes the review received was less than one quarter or more than three quarters of those that read the review and didn't consider it useful).

3. If the review's rating score was less or more than 3, which, considering the 5 value Likert scale, split the dataset into reviews that had low (1-2) and high (4-5) rating.

265

	Ur	Rating	Helpful	Total	Coleman-Liau	ARI	Gunning Fog	Flesch/Kincaid
Ur	1.0000							
Rating	0.2767*	1.0000						
Helpfulvotes	0.2261*	0.0513*	1.0000					
Totalvotes	0.0221*	-0.0656*	0.9074*	1.0000				
Coleman-Liau Index	0.1245*	-0.0036	0.0895*	0.0610*	1.0000			
Automated Readability Index	0.0653*	-0.0131*	0.0689*	0.0552*	0.4157*	1.0000		
Gunning Fog Index	0.0629*	-0.0139*	0.0651*	0.0530*	0.3921*	0.9873*	1.0000	
Flesch/Kincaid Index	0.0654*	-0.0155*	0.0682*	0.0556*	0.3930*	0.9949*	0.9933*	1.0000

Table 6.4: Pearson inter-item correlation matrix between the elements in our dataset ($*P < 0.01$)

	OLS	TOBIT	TRUNCREG
helpfulvotes	0.008***(0.000)	0.009***(0.000)	0.003***(0.000)
Realorder	-0.000***(0.000)	-0.000***(0.000)	-0.000***(0.000)
Rating Score	0.069***(0.001)	0.072***(0.001)	0.062***(0.001)
Number of Words	0.000***(0.000)	0.000***(0.000)	0.000*** (0.000)
Coleman-Liau Index	0.011***(0.001)	0.012***(0.001)	0.006***(0.001)
Automated Readability Index	-0.011***(0.002)	-0.013***(0.002)	-0.002 (0.002)
Gunning Fog Index	-0.006**(0.002)	-0.006*(0.002)	-0.005**(0.002)
Flesch/Kincaid Index	0.020***(0.003)	0.022***(0.004)	0.007**(0.003)
Constant	0.314*** (0.010)	0.275***(0.011)	0.501***(0.008)
R-squared (OLS): 0.152 , R-squared (Tobit): 0.1737 Confidence Level: *$p<0.05$, **$p<0.01$,*** $p<0.001$, N=36586, N-truncated: 33485			

Table 6.5: Regression Results of the usefulness ratio (UR)

Grouping Category	Group	Number of Observations	Percentage of the dataset	Split Criterion
Groupid-A	A_1	5,844	17.45	*Ur* <0.5
	A_2	27,641	82.55	Ur>=0.5
Groupid-B	B_1	3,980	10.88	*Ur* <0.25
	B_2	22,511	61.53	Ur>0.75
Groupid-C	C_1	5,025	13.73	Rating <3
	C_2	27,730	75.79	Rating$\geq$3

Table 6.6: Sampling and selection procedure for the groups used in our analysis

Since splitting the data into two groups gives us two independent samples from the same dataset, we are able to use a statistical test to find how significant the difference is between the means of the characteristic in which we are interested in these groups.

The usefulness of a review is affected by the rating that the review has received

Having split our dataset into two grouping variables, we are able to test the relation between the usefulness ratio of a review (UR) and the rating that this review has received. In particular, we are interested in identifying—by comparing the means of the ratings—whether the usefulness ratio has any relation to the rating that a review has received. Figure 6.7 shows the distribution of the usefulness ratio for each of the values of the rating scale (1-5).

Graphs by Rating Score
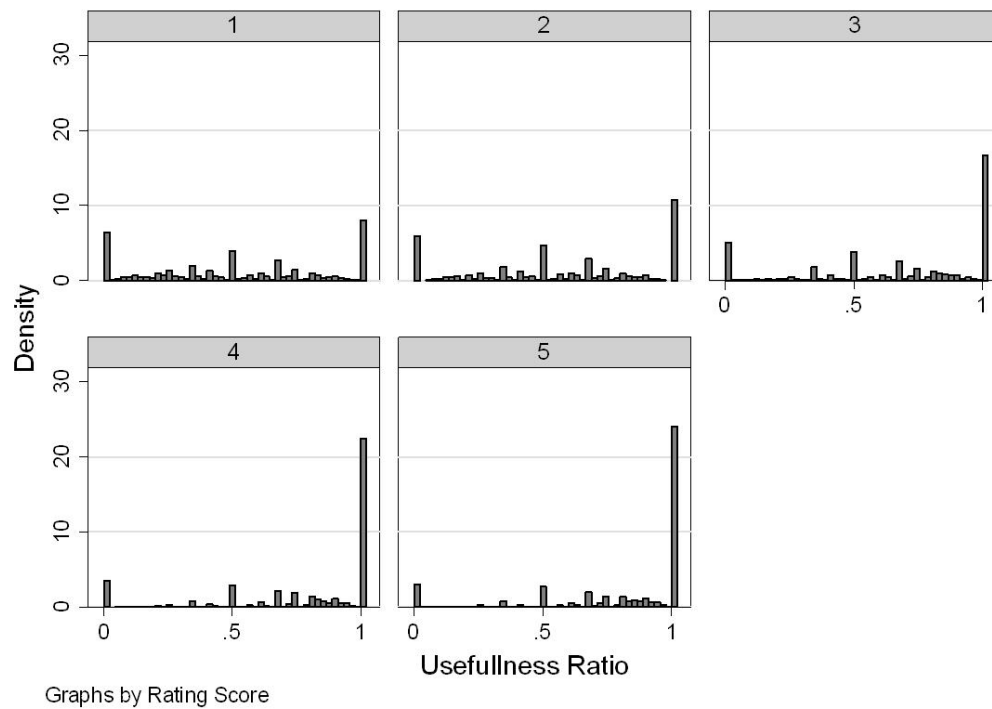
Figure 6.7: The distribution of usefulness ratio for each of the values of the rating scale that a particular item was evaluated

As aforementioned, we selected a nonparametric test to test whether the mean value of the rating is the same across the groups that contain high and low UR. For the grouping variable groupid-A the groups are split equally by the UR (groupid-A =1 if the UR is less than 0.5 and groupid-A =2 otherwise). We selected the Mann-Whitney test to compare the mean value of the rating between the two groups. The reason for choosing this particular test is that the Mann-Whitney test is a nonparametric statistical test that, unlike parametric tests, (t-test) does not rely on the assumption of normality (the distribution of the rating among the groups follows the normal distribution).

For the grouping category Groupid-A we ran the Mann-Whitney test for a total of N= 34002 observations. The Z value that we obtained from the test was Z =-39.407, which results in a P value of P=0.000, which is highly significant at three degrees of freedom, so we can reject the hypothesis that the mean of the rating is the same for reviews that have a high and low usefulness ratio.

In order to verify the above result in case the rating plays no role in determining the different values of the UR also in the case of the upper and lower limit (since most of the UR is concentrated on values of 0 and 1), we used the second grouping (Groupid-B) which splits the dataset in two parts, which in fact are the first and fourth quintile of the UR values. We ran again the Mann-Whitney test for a total of N= 26884. The Z value obtained from the test was Z=-35.547 which corresponds to a P value of P=0.000 providing that the hypothesis that the rating is the same between the two groups is rejected.

Both results affirm that, indeed, the rating is affected by the usefulness ratio of the review. In fact, as can be seen from Figure 6.8 the value of the UR is increased depending on the rating scale.

The usefulness of a review is affected by the qualitative characteristics of the review

In order to test whether the usefulness of a particular review is affected by the qualitative characteristics of the review text, we followed the same procedure for both grouping categories (Groupid-A and Groupid-B), which we used for testing the relation between the usefulness ratio and the rating.

Running the same test for the first grouping category (Groupid-A), we obtained a Z value of Z= -27.433, which provides a P value of P=0.000. Again, this value is highly significant at three degrees of freedom. The hypothesis that the mean of the review length is the same when the UR is high or low is rejected providing that the length of the review text also affects the usefulness of the review.
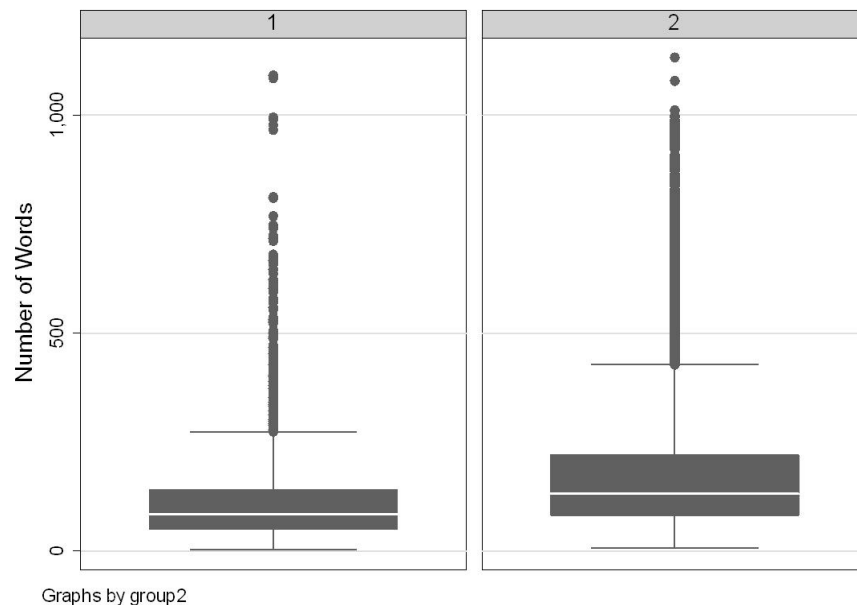
Figure 6.8: Average word length comparison between the two groups (groupid-B=2 if *ur* > 0.75)

For the second grouping variable, the same test gives us a P value of P=0.000 (Z= -37.050) which also rejects the null hypothesis. The results of the tests confirm that the difference displayed in Figure 6.8 is significant and provides that the qualitative

characteristics of the review (in that case the review text[6]) also affect the usefulness ratio of the review and are positively correlated (higher review usefulness implies that the review text will be longer).

The rating of a review is affected by the qualitative characteristics of the review

Running the Mann-Whitney test for the qualitative characteristics of the review text and using the third grouping variable (Groupid-C), we obtained a Z value of Z = -3.097 resulting to a P value of P=0.0020. This provides that we can reject the null hypothesis and confirm the significance of the relation between the qualitative characteristics of the review text and the actual rating value that a particular review has.
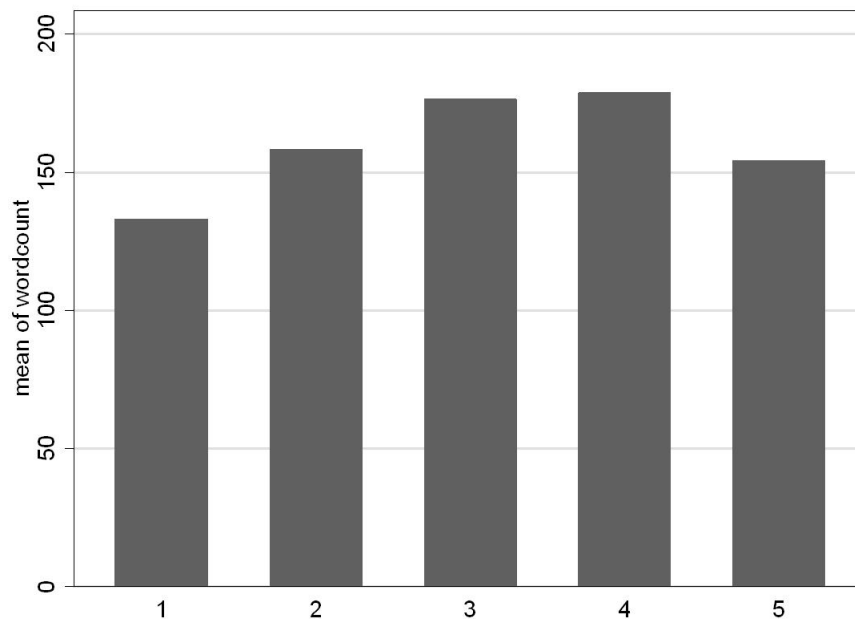
Figure 6.9: Distribution of the average length of the review text following the rating of the review

In fact, Figure 6.9 can be used to compare the tendency of the review text across the values of the rating scale that is used when a review is submitted (1-5). It is

---

[6]The results of the Mann-Whitney test for the Fog, Flesch, and ARI indexes also provide the same results (P=0.000 for Z values of -14.903, -12.547 and -3.728)

clear that reviews with positive ratings tend to contain more text (as depicted by the number of words contained in the review text).

6.4 Discussion

The results of the tests, along with the interpretation of the regression coefficients that we obtained in Section 6.3.2, provide us with some interesting insights into the relation between the usefulness of an online review as considered by the visitors to the online marketplace and the actual qualitative characteristics of that review. By our hypothesis testing of the data we were able to verify the following:

6.4.1 The usefulness of a review is affected by its positive or negative rating value

Going back to Figure 6.7, we see a significant (as confirmed by the tests) tendency of the usefulness ratio towards reviews with higher rating. That might be explained by the fact that consumers (as visitors of an online information resource) tend to read appraisals of a product first (the fact that a review is marked with 5 stars also increases attention from a usability point of view). From the definition of the usefulness ratio, the higher the rating value of a review, the higher the number of helpful votes that a (future) customer will give it. At least in our dataset, reviews with a rating value above three had a higher amount of usefulness ratio with a perfect score resulting in the fact that indeed the higher the number of helpful votes the review received, the higher was its usefulness ratio. This finding implies that customers react to positive and negative reviews differently, which also confirms results from the study done by Hu et al. [2008].

6.4.2 The usefulness of a review is affected by its qualitative characteristics

The word length and the readability scores (as a result) have confirmed that the style of the text in a review also provides an indicator of why a consumer considers a review to be highly useful. This can be explained by the fact that consumers evaluate a critique by how well it is justified and whether it provides them with as much information as possible in order to form their own views about the quality of a particular product (in our case a book) and reduce the uncertainty about its quality.

6.4.3 The rating that the review provides is affected by its qualitative characteristics

Our results indicate that there is a clear relation between the value of the rating that a review provides and its qualitative characteristics in terms of review length. From these results we can imply that consumers who are satisfied with the book they read want to express more of their personal opinions in their reviews, which makes the standard case from word-of-mouth scenarios that excited customers are often willing to provide more information about their experience and reflect their excitement in their judgment of the product or service they have consumed.

6.5 Conclusions and further remarks

The main result of our study is that when a particular review is considered useful to the potential buyers of a product or a service, this has something to do with the qualitative characteristics of the review justification as a piece of text. By employing readability formulas we were able to analyze the reviews in our dataset and provide a set of results in connection with the usefulness of the particular review.

In this study we have focused on the content-specific characteristics of the review text. However, one of the limitations of this approach is that we were not able to as-

sess whether a review was written in a way that expressed a personal opinion about a product or a service. This limits the study because we know from the marketing literature that potential consumers tend to associate themselves with other consumers who express a more personal experience about the product that might influence the potential consumers' choice process [Bettman and Park, 1980]. Another limitation of this study is its inability to check the actual reliability of the readability tests by cross validating whether the tests actually measure the readability of a review written on a website, since the readability tests do not take into account usability factors (e.g., the position of the text on the screen, etc.).

The empirical results of this study also contribute to the ever-growing literature on the importance of online reviews as an advantage to online marketplaces over traditional markets where the codification of information related to the products or services can actually help future buyers to evaluate the quality of an experience good (in our case books) by reading the judgments provided by other customers. In relation to the research question pursued in the context of this dissertation, we showed that the qualitative characteristics of online reviews communities and online communities in general are an important element that affects the perceived usefulness of the online community output.

The study also shows that the qualitative characteristics of online reviews are a rich source of information toward understanding the way consumers evaluate information about products in an online marketplace, which can be combined with evidence coming from studies regarding quantitative aspects as perception based on accumulation of negative reviews [Lee, Park and Han, 2008]. The results of this study represent a point of departure to extend the analysis further by incorporating cognitive characteristics of consumers as captured by their reviews.

Bibliography

J. R. Bettman and C. W. Park. Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, 7:234, 1980.

S. Borenstein and G. Saloner. Economics and electronic commerce. *Journal of Economic Perspectives*, 15:3–12, 2001.

J. J. Brown and P. H. Reingen. Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14:350, 1987.

J. A. Chevalier and D. Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43:345–354, 2006.

E. K. Clemons, G. G. Gao, and L. M. Hitt. When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, 23:149–171, 2006.

M. Coleman and T. L. Liau. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60:283–284, 1975.

L. J. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16:297–334, 1951.

C. Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49, 2003.

R. F. Flesch. *How to Test Readability*. Harper, 1951.

R. Gunning. The fog index after twenty years. *Journal of Business Communication*, 6: 3, 1969.

M. Hu and B. Liu. Mining and summarizing customer reviews. *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, 2004.

N. Hu, P. A. Pavlou, and J. Zhang. Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of online word-of-mouth communication. *Proceedings of the 7th ACM conference on Electronic commerce*, pages 324–330, 2006.

Nan Hu, Ling Liu, and Jie Zhang. Do online reviews affect product sales? the role of reviewer characteristics and temporal effects. *Information Technology and Management*, 9, 2008.

J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. 1975.

F. Lehner. Quality control in software documentation based on measurement of text comprehension and text comprehensibility. *Information Processing and Management*, 29:551–68, 1993.

G. L. Lohse and P. Spiller. Electronic shopping. *Communications of the ACM*, 41:81–87, 1998.

T. W. Malone, J. Yates, and R. I. Benjamin. Electronic markets and electronic hierarchies. *Communications of the ACM*, 30:484–497, 1987.

P. Nelson. Information and consumer behavior. *Journal of Political Economy*, 78:311, 1970.

J. C. Nunnally. *Psychometric theory*. McGraw-Hill New York, 1978.

M. K. Paasche-Orlow, H. A. Taylor, and F. L. Brancati. Readability standards for informed-consent forms as compared with actual readability. *New England Journal of Medicine*, 348:721, 2003.

P. A. Pavlou and A. Dimoka. The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17:392–414, 2006.

M. L. Richins. Negative word-of-mouth by dissatisfied consumers: A pilot study. *Journal of Marketing*, 47:68–78, 1983.

G. J. Stigler. The economics of information. *The Journal of Political Economy*, 69:213, 1961.

B. L. Zakaluk and S. J. Samuels. *Readability: Its Past, Present, and Future.* International Reading Association, Newark, 1988.

Part III

Findings and Conclusions

CHAPTER 7

Conclusions and retrospect

7.1 Discussion

The goal of this thesis was twofold. First, the objective was to understand the behavioral nature of contributions to online communities, and second, it was to elaborate on the factors that affect these contributions. In Chapter 1 we outlined the context in which we perceive an online community, that is, as an online social space where participants form exchange relations by sharing a common purpose. In this context the nature of a contribution can be either a participation to a discussion thread or (depending on the social policy that the online community facilitates) the addition of information in this online social space. Therefore the contributions, that the title of this thesis is referring to are contributions of information provided by participants in an online community.

These contributions were tackled in this thesis from the perspective of social capital. As mentioned in the introductory chapters, the main reason for doing so was that by definition the concept of social capital focuses on the relationships between the

individuals in a social structure [Burt, 2005], considering each individuals' attributes as important factors that affect these relations, relations with the emphasis on the availability of existing social resources. In this dissertation the social resource that is tackled is information.Therefore, the empirical studies that were presented in the empirical part focused on the individual attributes, and then moved to the case of availability of social resources in an online environment (information) by also examining the impact of implicitly and explicitly formed social relations. In relation to social capital research, this thesis contributes to the examination of the above characteristics in the context of an online community, theorizing that the interactions between the participants (facilitated by Computer Mediated Communication) over the social resources (in that case information) constitute virtual forms of social capital.

Although the above conclusive aspect is not new in the literature of online communities [Huysman and Wulf, 2006], the factual contribution of this dissertation relies on the extensive study of all the aforementioned aspects of social capital in online communities. Chapter 3 in particular, provides a study on the attributes of online social relations by utilizing the public goods game to study these characteristics under the perspective of shared social resources. The case of an online social dilemma as encapsulated by the public goods game is then extended in Chapter 4 where we study the impact of social relations implicitly formed in the context of a social resource that is the information available on the Yahoo!Answers online service. In Chapter 5, from the perspective of social capital, this thesis contributed in two ways. The first was to provide empirical evidence on the existence on social norms (in that case tips) from offline environments to online environments; the second was the evaluation of the nature of social interactions when bound to these norms where participants strategically manipulate their behavior (by making the use of a social norm) in order to gain more from other participants. Chapter 6 evaluated the quality of the social resources, thus completing the perspective of social capital to behavioral characteristics, relations and social resources.

Another aspect that was outlined in the introductory chapter is the case of motivational factors. As has been discussed in the theoretical part of this dissertation, our view of motivational factors is related to the expectations of the participants and their endogenous characteristics as expressed in other activities of everyday life. Literature refers to these characteristics as social norms since these are expressions of behavior that are imposed to individuals by the societal structure in which they are active and taking part[Cialdini and Trost, 1998]. One particular research issue that was addressed in the introductory chapters was whether social norms can be found on the internet where no sanctioning mechanisms can be easily imposed and the anonymity of an alias or pseudonym protects the defiant from its offline social environment.

While the study of these characteristics of online communities seems to be framed in research oriented only settings, there is a considerable set of practical implications that can be theorized from the contributions of this thesis. The practical contribution of this thesis targets the design of software for online communities. From such a perspective, the implications of this thesis target software for online communities not as a platform for communication, but as a platform that can enhance the social and behavioral dynamics that became evident in the empirical part of this dissertation.

Table 7.1 summarizes the findings and implications of the empirical studies. We provide a description of the findings in subsequent sections.

7.1.1 Importance of Signaling Mechanisms

The design of Online Communities is not a software or systems implementation issue; rather, it is a design of effective social interaction taking advantage of the abilities of the internet and WWW to facilitate interaction, regardless of the limiting factor of time and space. In this way, the thesis provides grounds for the importance of the implementation of social monitoring and interaction mechanisms that can boost activity in the context of an online community. One particular finding from Chapter 4 is that a mechanism whereby participants can signal their level of commitment to

Research Questions	Findings and Implications
Study 1: What are the characteristics of participants in online communities in socioeconomic terms? Are frequent participants also cooperative? Do they have some form of altruism embedded in their behavior? Can these two states be disentangled?	Symmetric effects between cooperation and participation frequency. The more someone participates in an online social activity the more cooperative it becomes.
Study 2: Do users with continuous presence receive many more responses than do users that have just started using the service? Do users post an answer to a question before getting an answer to the question that they themselves posted on the system ? Do users get an account of the number of times the user who posted the question has provided answers to the community?	Participants do care about the participation and contribution rates of the other participants in order to participate.
Study 3: What is the effect of economic incentives and repeated interaction on service promptness in the context of an online community such as the one in GoogleAnswers ?	Supporting evidence for the existence of social norms in an online context. Significant amount of participants contributing strategically in order to get better service in the future with considerably less effort.
Study 4: What is the impact of the qualitative characteristics of a contribution on the perceived usefulness of the online community?	Qualitative characteristics of content have a significant effect on the perceived usefulness of the community for the participants.

Table 7.1: Summary of the Research Questions and Findings in the empirical Part of this dissertation

the community (in terms of the effort they provide in the community, in our case the amount of information) can help them receive answers in shorter time than those do who commit less. This provides an incentive then for those who wish to get more benefit from the community by contributing more.

Having other participants benefit from this signaling mechanism has important usability implications as well [Lampe et al., 2007]. For example, how should the activity level be measured? Should the users be penalized by reading only and not contributing? This can have a counter argument anchored in the knowledge barriers that are imposed on someone who wishes to participate. However, this dissertation has shown that the availability of such a mechanism in an online community can greatly enhance participation and therefore make the online community accumulate more members and information.

7.1.2 Identification of the Behavioral Characteristics

Another important dimension that this thesis has explored is the importance of considering the behavioral characteristics of the participants as a factor that can make an online community successful. The findings from Chapter 3 provide a valuable guideline for considering behavioral characteristics of online community participants taking a latent construct of online sociability as a reference point. The positive symmetric relation that was evaluated in the analysis of the empirical study provides an interesting implication for practical scenarios as well, which is that the more social the online community members become, the more they tend to contribute to the community. Theoretically, this can be attributed to efficacy characteristics also mentioned in the literature by Tedjamulia et al. [2005] and Kavanaugh et al. [2005].A sense of efficacy is an important factor found to be related to participation in social activities.

7.1.3 Ability of the participants to interact strategically

The practical implication of interacting strategically essentially confirms the empirical evidence from the study of online social networks where status seeking participants seek to increase their status by accumulating as many connections as they can. In this case their strategic motive is to have better access to resources than other fellow participants, and in order to achieve that, they devise a strategy. This can depend on the importance of the nature of the social resource to each participant. If the social resource is, for example, access to the labor market (as in the case of Linkedin.com) or the internet music scene (e.g., MySpace), participants tend to act more strategically in the way that they interact with their fellow participants.

Therefore, it is important for the community software mechanism to provide the ability for the participants to track the activity of others in order to compete with different ways to gain status or go higher in the hierarchy of the online community. In relation to incentivized actions, this can be either in an intrinsic mode (e.g., receiving the status of an expert in a high profile community of programmers) or by getting an extrinsic reward by answering a question (e.g., in the case of Yahoo!Answers). Furthermore, the community participants who receive more benefits, that is, those that make a greater effort in providing material to the community, would be able to give some extrinsic forms of rewards (e.g., standard payments or gift coupons) to those who would like to undertake their task (e.g., question about a specific problem).

7.1.4 Importance of the quality evaluation mechanisms

Quality evaluation is another important factor when it comes to motivating activity in online communities. With the growing amount of unsolicited information posted to online fora by automated software agents (spambots), an important issue is the evaluation of the quality of the material available in a community. For example, much discussion has been tackled in the online community literature regarding the importance of moderation mechanisms. A special case of a moderation mechanism can be

the case of user assisted moderation, where users can collaboratively filter out any non relevant or non important information, thus making the material available for the online community participants valuable. In the final chapter of part II of this dissertation we discussed the case of online reviews and whether the usefulness score that was assigned to them by other participants had anything to do with their qualitative characteristics.

The main practical implication of this finding relates to the importance of a quality evaluation mechanism for an online community. The empirical findings of Chapter 6 confirm a symmetric relation between high quality reviews and good qualitative measures, thus proving that the results from the social filtering mechanism provide a evidence of the actual quality of the reviews (in terms of readiness), as measured by the readability indexes that were constructed for that purpose.

7.2 Conclusions

Having provided a discussion of the implications of this dissertation, we revisit the research question formed in the introduction and we summarize the limitations and the future research in the following sections.

7.2.1 Retrospect

This dissertation encompassed four empirical cases related to the study of behavioral characteristics from the perspective of motivating contributions in online communities. In connection with the research question developed in the introductory part of this thesis, our aim was to understand the nature of contributions in online communities and identify the factors that enhance and sustain them. In order to better understand these behavioral factors, we adopted a bottom up approach. First, we studied the nature of the contributions in an online setting using a controlled environment with declared extrinsic rewards (payoffs) positioned in the case of the cooperative contri-

bution mechanism implemented by the public goods game. Having identified some of the behavioral characteristics that sustain contributions in an online community, such as communication and social interaction, we continued to study an application of the cooperative contribution mechanism in the context of Yahoo Answers. Here we analyzed how past activity (in terms of contributed effort and perceived benefits) had an influence on the realm of a purpose oriented online community such as the one in Yahoo Answers. In Yahoo Answers those that contribute effort, however, are not compensated in an extrinsic form, but receive intrinsic forms of motivation, such as reputation signals, etc.

Extrinsic rewards might have an effect in that context and that was the purpose of studying such interactions in a different environment Extrinsic rewards might have an effect in that context and this was the purpose of studying such interactions in a different environment which was operated by Google called Google Answers. In Google Answers those who were contributing effort were compensated using extrinsic rewards and, in particular, the fixed price that someone was willing to pay if the question that was submitted received an acceptable (by an agent) answer. This is, in fact, a standard Principal-Agent mode of operation where the principal has a pre-declared price for a task and the agent takes on the task or not [Regner, 2004]. Furthermore, if the principal was more than satisfied with the agent's performance, he/she was awarded a further reward in the form of a tip. We studied the factors that affected tipping in the GoogleAnswer's platform in order to find out whether extrinsic forms of motivation are efficient for sustaining activity and increase the overall turnover of the community, both in terms of volume (in that case the volume of answers produced) and participation levels.

Contributions might also be affected by the (perceived) value of the information available and the way it is expressed. This was the focus of the fourth empirical study presented in the previous chapter. This approach relates content quality with the perceived usefulness that a contribution might have in the context of an online com-

munity that is formed around information goods (in the case of the previous chapter: books).

Chapter 5, on the other hand, adopts a market perspective which is dictated by the context of study. Participants in GoogleAnswers (askers) do have a strong willingness to pay, due to the fact that processing of information available for retrieval requires high levels of cognitive ability. This is also related to the standard problem of query formulation from the information retrieval perspective [Aula, 2003].

7.2.2 Revisiting the general research question

After summarizing the key findings and the implications of the four empirical studies we are in a position to revisit the general research question: ***What are the main driving factors that affect contribution in Online Knowledge Communities?***

The obvious conclusion reached is that behavioral characteristics of the users are a key element affecting participation; further, the facilitation of social interactions through the community mechanism is an important element for the success in terms of the sustainability and evolution of an online community. The research presented in this dissertation has examined the nature of the motives that affect participation and, in particular, the effect of extrinsic and intrinsic rewards as an important factor that drives this participation.

Another issue with the research question framed above is the case of the unit of analysis. In this thesis the unit of analysis for the research question pursued was the characteristics of the individuals since we intended to analyze interaction between individuals and not collective characteristics of an online community (which is evident in other approaches on how communities sustain and evolve) [Boccaletti et al., 2006, Hansen, 2002, Jackson, 2003].

This dictated the approach to the research question to be of behavioral nature since we were interested in the individual motivational characteristics and not in the group properties that might be formed during the formation of an online community.

287

Part I of this thesis offered previous theoretical and empirical work concerning the issue of participation in online communities. The research summarized in this part essentially provided the ground for the four empirical studies presented in Part II of this dissertation, where each particular case was addressed in the setting of an online community. Chapter 3, however, was not a study that took place in the context of an online community; rather, it was a controlled reconstruction of the dilemma of participating or not in an online community, thus using two concrete framings (GIVE and TAKE) as a model of the participation (contribute to the community or get benefited by it). The findings suggest that users who participate more frequently in an online community tend to give more than those that have a more spontaneous participation rate.

The research findings also have more context specific implications to the research question framed above. In connection with the theoretical background presented in Part I of this dissertation, we provide the related argumentation as to the contributions and the empirical findings of this thesis.

7.2.3   Summary of the findings and the implications of the empirical studies

As discussed in Chapter 1 of this dissertation, we position the findings of this thesis by revisiting the framework of Snyder and Cantor [1998] in relation to the motivational factors that affect participation in online communities.

Value expressiveness is positioned as a way of expressing values about other's actions and concerns, something that was evident in the studies presented in chapters 3 and 4. On the other hand utilitarian functions were related to the study of Chapter 5 where we provided empirical evidence that participants also think strategically when they interact with other participants in order to receive the maximum benefit with the least effort.

The function of a participant being socially adjusti??ve, on the other hand, relates

Function	Supporting Evidence on Research Findings
Value Expressive	1,3
Utilitarian	1,3
Social Adjustive	2
Knowledge Seeking	2

Table 7.2: Motivational Factors and the supporting evidence provided by this dissertation

very much to the existence of social norms, which was evident in the study of Chapter 5, where participantsŠ tension to comply with an offline social norm also had an influence in their online social behavior.

Research Finding 1: Participants on online communities do care when other participants participate or not

The study presented in Chapter 4, having as a context of study the realm of Yahoo!Answers presents empirical evidence that participants do care about the contributions of other participants (and provide an answer). That is in essence a confirmation in online settings of the general literature of social preferences, where participants do feel envy about other participants receiving more benefit with less effort, and therefore are not willing to contribute. Social preferences also relate with the so called group mediation factor as the one tackled in the Collective Effort Model [Karau and Williams, 2001]. This might lead to an increase of expectations of contributions by those that have already contributed a level of effort that is higher than the average contributed effort. For example, in the study of chapter 4, the empirical evidence suggests that participants tend to decide to reciprocate by the level of previous effort made by those who ask for an answer to a posted question, that is, whether to participate in the thread or not.

Although this is not observable by users in a direct way, the ability of the software

to provide information (in the user's profile page) about how much a user has benefited by the community provides the ground for a discussion whether the users' do actually care about the actions of the other participants in that particular online community. In the time period that we undertook this analysis, the results suggested that (a) a high level of contribution resulted in a shorter time to get an answer and (b) a low level of contribution resulted in a higher time to get an answer. This confirms the argument by Kollock [1999] as to the change of the production function of an individual in an online setting, based on the perceived benefit or effort that this individual will conceive by his/her participation on the online community.

Research Finding 2: A High degree of social Interaction leads to higher contributions in the online community

The study presented in Chapter 3 has provided empirical evidence which suggests the existence of symmetric effects between cooperation and participation frequency in online communication activities. The more someone participates in an online social activity the more cooperative he/she becomes. This, in essence verifies findings from the social capital literature and, in particular, the literature related to quantifications of social capital with respect to social and organizational activities [van der Gaag, 2005, Quan-Haase and Wellman, 2004, Wellman et al., 2001]. This finding opens up the question of whether social norms are sustained on a highly anonymized setting such as on the internet. On such terms, although social interaction is anonymous from the perspective of personal interchange, structural relations are sustained and the hiding behind pseudonyms still provides an identification of the actions of each individual in the context of a group.

This finding suggests that exploiting social interaction in the realm of an online community will have a positive effect on the activity of the community and the resulted contributions, since "normative" social influence will become an important factor in dictating increased participation.Dholakia et al. [2004] report a case as a result

of episodes of social interactions in the context of a virtual community of consumers. Such a case of social influence relies on the provision and further support of mechanisms (such as those that are provided in the realm of Yahoo!Answers)where participants will be able to trace other participantsŠ activities, as the first finding suggests.

Research Finding 3: Content quality is an important factor for the perceived value that the community contributes to a participant

This research finding supports the knowledge seeking perspective from Snyder and Cantors' framework. In particular, the results from the study presented in 6 to some extent, the significance of the answer length variable in the case of Chapter 4) support this behavioral characteristic in the direction that the quality of content enhances participation and enhancement of contributions. This is also evident in Wikipedia where contribution of content has an effect on the participant's reputation, as the study by Ciffolilli [2003] argues. This, in fact, can be related, to some extent, to the reputation of the individual contributors. It is expected that highly reputable members of an online community will contribute high quality content and have an extra incentive to contribute in order to maintain their status index. This is also connected with the volume of social interactions that take place in the community. A high level of social interactions and/or a high number of participants makes the importance of a status index significant. An individual then becomes self motivated to contribute high quality content to maintain this status index, which also results in a continuous cycle of participation, since the empirical evidence from chapter 6 and 3 suggests that high quality content makes the perceived value / benefit for an individual important.

The findings of Chapter 6 are also particularly connected with a study by Rashid et al. [2006] where it was shown that the perceived value of information displayed had an effect on the participation by community members. These findings hold in the context of an online experience sharing community however to some extend it can be generalized that the quality of content has an effect on the personal attitudes of the

users towards the online community [Curien et al., 2006].

7.2.4 Additional contributions and discussion

This thesis also contributes empirically. First, this is done by providing empirical evidence as to what social mechanisms support cooperation in the context of an online community. This can be used as a factor to consider the design of more effective online community software that will allow for the creation of sustainable and evolving online communities. A particular issue in online communities, as has been highlighted in the related literature, is the phenomenon of lurking or participation in absentia. Online communities seem to suffer from that factor, since interaction is low, regardless of the number of registered participants. Empirical evidence that this dissertation provides is that the provision of direct (user-to user) social interaction mechanisms can enhance the activity in the online community. While this seems to be obvious, from a design perspective this thesis provides empirical evidence for the support of this direction to the design of online community software.

Furthermore, the ability of participants to rate the quality of the content provided in the online community is another factor that affects the participation in an online activity as it does not pose any interpersonal barrier, such as the lack of expertise. This partially enhances the level of social activity since it does not involve a direct contribution but an indirect contribution as to the improvement of the existing content status and quality. The classification of activities, as presented by Koch and Wörndl [2001] and discussed in the introductory part of this dissertation, connects empirically with the findings of this dissertation since the empirical evidence that we presented highlights the importance of the facilitation of interaction as an important element of an online community. The support for this case has been extended to the way that dyadic (and thus no hierarchical) form of interaction can be enhanced with the presence of social cooperation mechanisms (e.g., the display of an index highlighting the contributed posts as to highlight contributed effort).

The use of experimental methods in information systems research

Although the core focus of this dissertation has been to examine the relation between information exchange and behavior in the context of an online community, it also uses a broader methodological paradigm related to the use of experimental methods in information systems research. The use of such a detailed methodology also addresses the recent critiques of the relation of information systems research to the current practice, as well as the call for research in the area of information systems to focus more directly on the technological artifact [Orlikowski and Iacono, 2001, Benbasat and Zmud, 2003].

The significant potential of experimental methods in contrast to other information systems research methodologies is the prescriptive (rather than descriptive) nature of the research inquiry which provides explicit guidance to information systems developers as to what will make a technology useful (rather than focusing on whether it is useful or not). For example, the celebrated TAM model [Davis et al., 1989, Venkatesh et al., 2003], even though it identifies predictors for technology adoption in an organizational setting (e.g., utility, ease of use etc), it does not provide guidance on what technological features will make a technology easy to use and increase its usefulness. Other information systems research theories such as adaptive structuration theory [DeSanctis and Poole, 1994] or fragmented institutionalism [Lamb and Kling, 2003] are only useful to retrospectively explain why a technology was successful or not because it provides little practical guidance in either designing or managing new technological interventions.

An additional important case related to this dissertation is the switch between the organizational to a user oriented setting, where the technology provided here is mainly used as a platform and not as a stand for accomplishing a specific task in an organizational setting. From this perspective, the empirical studies presented here consider the organizational affiliation as an enabler to engage in communication with other users and not as a boundary which limits their interaction with actors from the same

organizational setting.

From an experimental methodological point of view, the empirical studies conducted in this dissertation are in the category of artefactual field experiments, where the subject recruitment procedure is done in the field (without a concrete procedure for recruiting participants in the study, but by observing their behavior in natural occurring settings). While the empirical study presented in Chapter 3 had a stratified subject recruitment procedure, the empirical studies that are presented in the other chapters of the empirical part follow a higher exogenous subject attrition to the context of study than the initial one. Whereas this can be criticized in terms of selection effects, the advantage of this is that subjects are non-informed about observation which, in turn, avoids biases such as the Hawthorne effect. As long as the internal validity of the measurement holds, then it is methodologically safe to argue that the subject selection has a non significant effect on the outcomes of the study.

The last but not least important issue with the methodology of the empirical studies is the case of the external validity. While the study of Chapter 3 was carried out in an abstract setting (with the purpose of emulating the case of a social dilemma), the other three studies considered an applicable context producing prescriptive findings (as seen in Table 7.1). From that point of view, the use of experimental methods in this dissertation aligns the methodological approach with the recent developments in the IS field.

7.3 Where do the findings apply?

As mentioned in the opening chapter and also in Chapter 2, the definition of an online community is characterized by a high degree of imprecision. The studies presented in this dissertation consider the case of structured and dyadic form of interaction in the context of online community. This, in essence, provides the assumption that participants in an online community are often engaged in communication with other

participants enacting the role of the facilitator in a more abstract way. However, there might be cases where the online community has a more direct form of engagement, as, for example, in an online Blog. There is a growing set of literature arguing that the leaderŠs involvement as a facilitator can enhance the activity of an online community [Butler et al., 2007, Koh et al., 2007],as in an offline setting [Bonjean, 1963]. Web logs or blogs, for example, act as an enabler of this direction. Blood [2004] discusses the implications of widely available blogging software as an enabler of the facilitation of this kind of social interactions.

The perspective of knowledge sharing dilemmas, as argued by Cabrera and Cabrera [2002], poses another application of the findings of this thesis. Table 7.3 summarizes the results obtained by the empirical studies with the interventions proposed by Cabrera and Cabrera [2002].

Objective	Results
Encourage Communication	Positive/Symmetric
Increase frequency of Interactions	Positive/Symmetric
Increase expectations of others' participation	Asymmetric
Publicize information about members' contributions	Asymmetric
Increase Sense of group identity	Positive/Symmetric

Table 7.3: Objectives and research findings on the Knowledge Sharing Dilemma approach

Increasing the sense of group identity will have the result of an increased sense of group efficacy [Hsu et al., 2007] which will make the basis for the imposition of social norms stronger. The supporting evidence for the existence of social norms appearing in the study presented in Chapter 5 pushes this direction further where, in fact, participants do seek social approval or on the contrary seek to avoid social disapproval when participating in an online community. For example, the discussion around a par-

ticular topic on an online blog has a linear mode of evolution (typically threads are not formed due to software constraints because other than an author's postings appear as comments and not threads). A particular example can be the case of the Slashdot online community. Findings from recent empirical studies [Lampe and Resnick, 2004, Poor, 2005] can back the suggestion that the social norms in these cases evolve from the use of sanctioning mechanisms, such as those used by moderators. Moderation, however, is a form of sanctions and not an exact enabler of social norms. This thesis provides evidence that even in online anonymized settings, social norms do exist and evolve and sanctioning mechanisms tend only to represent social disapproval which in the case of the Slashdot system is distributed to the other members.

Social disapproval can also be expressed in the case of free riding. The evidence, provided by the studies presented in Chapters 3 and 4, also provides ground for the existence of social preferences as in the case of users caring about the activity of other users. This has an effect on participation rates. The literature of social influence also confirms these findings [Cialdini and Trost, 1998, Marsden and Friedklin, 1993] since strong social influence will cause the participants to further seek social approval for their actions.

Encouraging communication leads to another direction connected to the case of reciprocity (as discussed in Chapter 5). As the evidence of Chapter 5 suggests, increased communication with the presence of strong dyadic interaction leads to a higher degree of contribution (as is depicted by the voluntary contribution of the tip).

The study presented in Chapter 6 relates to a lesser degree to the structure of interaction in the context of an online community, but goes directly to the output (in that case the usefulness of the content as ranked by the other participants). Barriers to entry might also pose an issue due to the fact that in the case of a community formed around online reviews, expertise is needed in order to contribute content. That is, in essence, the need that the study presented in Chapter 3 addresses. In a very abstract setting where expertise factors are stripped down, we were able to disentangle the

relation between participation and willingness to contribute.

7.4 Limitations

The limitations of this thesis are coupled with two very specific characteristics of the research presented here. The first has to do with the nature of the research question and the second with the context in which we undertook the empirical studies as presented in part II of this dissertation. In particular the communities (and the subsequent datasets) that were examined under the prism of the research question have characteristics that are unique in the context of an online community. tThe two field studies (Yahoo Answers and Google Answers) dealt with a form of community that was functioning as an information market In this context, there was evidence of the formation of principal-agent relations where a participant was posting a specific question and another participant was replying. In the case of GoogleAnswers, the pool of participants that were acting as agents was controlled, as Google was pre-evaluating the skills of these agents in order to avoid high variation to the quality of service provided. Thus, extrinsic rewards had an effect due to the fact that there was high degree of effort involved in undertaking these tasks. On the other hand, the realm of Yahoo!Answers was acting as a repository of public goods where participants were contributing their endowed experience to the various subjects directed clearly by intrinsic motives. However, what would have been the effect if Yahoo!Answers had involved the reward of monetary outcomes in order to enhance the participation?

A naive approach in that case would suggest that extrinsic rewards would enhance participation. However, extrinsic motivation does not always have the desired result due to the social context in which the rewards are applied. For example, there might be an issue where extrinsic rewards undermine the social value of contributions, thus making participants unwilling to participate at all. The topic is known in the literature as the "motivation crowding effect" and was first introduced by Titmuss [1971] in his

book *The Gift Relationship.* Titmuss' research context tackled the blood contributions and the social value that underlines such a case. A similar behavioral approach is evident also in the psychology literature as the "Cognitive Evaluation Theory" [Deci et al., 1999]. This in essence provides that extrinsic rewards are not the panacea for enhancing activity or increasing the participant's perceived usefulness of an online community.

The above, intrinsically provides also the limitations of the study presented in Chapter 3, where the nature of incentives was presented as variable monetary payoffs dependent on the cooperation equilibriums arising in the different groups formed with participants. However, an issue that was not tackled in the experimental setting was the cultural influences of the environment and the subjective value of information on that perspective. It might be that information, or the perceived benefit, received by a participant of an online community is invaluable and, in that case, the value of information can have a significant effect on participation.

Chapter 6 addresses the choice of the selection of an experience good, however, the context of study is limited to a community formed around product reviews where, as mentioned earlier, expertise might pose a barrier for contributions. For example, inexperienced customers will not choose to post their opinion, either due to lack of expertise or to avoid disapproval expressed by the ratings of the others. Furthermore, the quality measurement scales that were used in the empirical study of chapter 6 were "blind" to the behavioral elements of expression that might be captured by a written review (e.g., bad temper due to the product malfunctioning).

Another interesting approach to evaluate the value of the benefits obtained by an online community participant is the study of Chapter 5, where gratitude was expressed with the use of voluntary excess rewards (tips). In this regard, this dissertation also provides evidence of an effect of strategic behavior where some participants were able, by showing a high degree of gratitude in the beginning, to obtain much more benefit with less effort (as depicted by the sum of the declared price and tip) across

time. This should also be taken into account in the design of social interaction mechanisms in an online community where strategic behavior will be able to be formed either in an extrinsic or intrinsic form.

7.5 Topics for future research

As aforementioned, this thesis provides empirical evidence as to the factors that affect cooperation in online environments. However, given the constraints of this thesis, a number of points of potential interest to the present thesis have not been dealt with, and are thus posed for further research. Chapter 1 dealt with the perspective of online communities as virtual forms of social capital which, in essence, is a broader research area itself. Future research could contribute to the understanding of other behavioral characteristics of participants in online communities, such as in the case of social norms. Strategic behavior of participants in online communities, on the other hand, complements this direction, since the theoretical argument, as Kollock [1998] suggests, for an individual is to maximize his/her benefit by participating in such a collective structure.

However, future research should rely on already widely used systems and technologies. A particularly interesting point would be to evaluate whether the findings from this dissertation hold for broader definitions of an online community, where participants do not form interactions over a shared purpose but over a broader aspect of activities. For example, the open source developers might not share the same ideology (e.g., to develop free software) since some of them look for benefits from the established business model already existing in the market.

Another possible extension of the research presented in this thesis is the application of the findings to the enhancement of contributions in peer to peer networks. Antoniadis et al. [2004], for example, study the impact of different extrinsic rewards for contribution to peer to peer systems and their effect on the overall utility of the

network. Peer to peer networks also suffer from the problem of lurking where security concerns have to be addressed, as pointed out by Davidson [2003]. The study of lurking behavior is a research stream that is becoming more and more important in relation to online communities [Preece et al., 2004]. Possible extensions of the research presented in this thesis should also consider lurking as an important factor for the impact of the motivational factors addressed in this dissertation.

Bibliography

P. Antoniadis, C. Courcoubetis, and R. Mason. Comparing economic incentives in peer-to-peer networks. *Computer Networks*, 46(1):133–146, 2004.

A. Aula. Query Formulation in Web Information Search. In *Proceedings of IADIS International Conference WWW/Internet*, volume 1, pages 403–410, 2003.

I. Benbasat and R.W. Zmud. The identity crisis within the IS discipline: defining and communicating the discipline's core properties. *Management Information Systems Quarterly*, pages 183–194, 2003.

R. Blood. How blogging software reshapes the online community. *Communications of the ACM*, 47:53–55, 2004.

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424:175–308, 2006.

C. M. Bonjean. Community leadership: A case study and conceptual refinement. *American Journal of Sociology*, 68:672, 1963.

R.S. Burt. *Brokerage and Closure: An Introduction to Social Capital*. Oxford University Press, 2005.

B. Butler, L. Sproull, S. Kiesler, and R. Kraut. Community effort in online groups: Who does the work and why? *Leadership at a Distance: Research in Technologically-Supported Work*, 2007.

A. Cabrera and E. F. Cabrera. Knowledge-sharing dilemmas. *Organization Studies*, 23: 687–710, 2002.

R. B. Cialdini and M. R. Trost. Social influence: Social norms, conformity, and compliance. *The handbook of social psychology*, 2:151Ű92, 1998.

A. Ciffolilli. Phantom authority, self–selective recruitment and retention of members in virtual communities: The case of Wikipedia. *First Monday*, 8(12):1396–0466, 2003.

N. Curien, E. Fauchart, G. Laffond, and F. Moreau. Online consumer communities: escaping the tragedy of the digital commons. *Internet and Digital Economics, Cambridge University Press, Cambridge*, page 201Ű219, 2006.

A. Davidson. *Peer-to-Peer File Sharing Privacy and Security*. Center for Democracy and technology, 2003.

F. D. Davis, R. P. Bagozzi, and P. R. Warshaw. User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, 35:982–1003, 1989.

E. L. Deci, R. Koestner, and R. M. Ryan. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125:627–668, 1999.

G. DeSanctis and M.S. Poole. Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, pages 121–147, 1994.

U.M. Dholakia, R.P. Bagozzi, and L.K. Pearo. A social influence model of consumer participation in network-and small-group-based virtual communities. *International Journal of Research in Marketing*, 21(3):241–263, 2004.

M. T. Hansen. Knowledge networks: Explaining effective knowledge sharing in multiu-nit companies. *Organization Science*, 13:232, 2002.

M. H. Hsu, T. L. Ju, C. H. Yen, and C. M. Chang. Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expecta-tions. *International Journal of Human-Computer Studies*, 65:153–169, 2007.

M. Huysman and V. Wulf. IT to support knowledge sharing in communities, towards a social capital analysis. *JOURNAL OF INFORMATION TECHNOLOGY*, 21(1):40, 2006.

M. O. Jackson. A survey of models of network formation: Stability and efficiency. *Group Formation in Economics: Networks, Clubs and Coalitions*, 2003.

S. J. Karau and K. D. Williams. Understanding individual motivation in groups: The collective effort model. *Groups at work: Theory and research*, pages 113–141, 2001.

A. Kavanaugh, J.M. Carroll, M.B. Rosson, T.T. Zin, and D.D. Reese. Community Net-works: Where Offline Communities Meet Online. *Journal of Computer-Mediated Com-munication*, 10(4), 2005.

M. Koch and W. Wörndl. Community support and identity management. *Proceedings of the seventh conference on European Conference on Computer Supported Coop-erative Work*, pages 319–338, 2001.

Joon Koh, Young-Gul Kim, Brian Butler, and Gee-Woo Bock. Encouraging participation in virtual communities. *Communications of the ACM*, 50:68–73, 2007.

P. Kollock. *The economies of online cooperation*. Routledge, London UK, 1999.

P. Kollock. Social dilemmas: The anatomy of cooperation. *Annual Reviews in Sociology*, 24:183–214, 1998.

R. Lamb and R. Kling. Reconceptualizing users as social actors in information systems research. *Management Information Systems Quarterly*, pages 197–236, 2003.

C. Lampe and P. Resnick. Slash (dot) and burn: distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 543–550. ACM New York, NY, USA, 2004.

C.A.C. Lampe, N. Ellison, and C. Steinfield. A familiar face (book): profile elements as signals in an online social network. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 435–444. ACM Press New York, NY, USA, 2007.

PV Marsden and NE Friedklin. Network studies of social influence. *Sociological Methods & Research*, 22:127, 1993.

W.J. Orlikowski and C.S. Iacono. Research commentary: desperately seeking the" IT" in IT research-A call to theorizing the IT artifact. *Information Systems Research*, 12 (2):121–134, 2001.

N. Poor. Mechanisms of an online public sphere: The website slashdot. *Journal of Computer-Mediated Communication*, 10, 2005.

J. Preece, B. Nonnecke, and D. Andrews. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20:201–223, 2004.

A. Quan-Haase and B. Wellman. How does the internet affect social capital. *Social Capital and Information Technology*, page 113Ű135, 2004.

Al M. Rashid, Kimberly Ling, Regina D. Tassone, Paul Resnick, Robert Kraut, and John Riedl. Motivating participation by displaying the value of contribution. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 955–958, Montreal, Quebec, Canada, 2006. ACM.

T. Regner. Why voluntary contributions? google answers! 2004.

M. Snyder and N. Cantor. Understanding personality and social behavior: A function-alist strategy. *The handbook of social psychology*, 1:635Ű679, 1998.

S.J.J. Tedjamulia, D.L. Dean, D.R. Olsen, and C.C. Albrecht. Motivating Content Contri-butions to Online Communities: Toward a More Comprehensive Theory. In *Proceed-ings of the 38th Hawaii International Conference on System Sciences*, volume 193, 2005.

R. M. Titmuss. The gift relationship: From human blood to social policy. *London and New York*, 1971.

M.P.J. van der Gaag. *Measurement of Individual Social Capital*. s. n, 2005.

V. Venkatesh, M.G. Morris, G.B. Davis, and F.D. Davis. User acceptance of information technology: Toward a unified view. *Management Information Systems Quarterly*, pages 425–478, 2003.

B. Wellman, AQ Haase, J. Witte, and K. Hampton. Does the internet increase, decrease, or supplement social capital?: Social networks, participation, and community com-mitment. *American Behavioral Scientist*, 45:436, 2001.