Open Government Communities
Does Design Affect Participation?

Høgenhaven, Thomas

Open Government Communities

Copenhagen Business School

HANDELSHØJSKOLEN

Open Government Communities

Does Design Affect Participation?

Thomas Høgenhaven

PhD Series 31-2013

LIMAC PhD School
Department of Operations Management

PhD Series 31-2013

Open Government Communities
Does Design Affect Participation?

Thomas Høgenhaven

Advisor: Kim Normann Andersen

Department of IT Management
LIMAC PhD School
Copenhagen Business School
2013

Thomas Høgenhaven
Open Government Communities
Does Design Affect Participation?

1st edition 2013
PhD Series 31.2013

LIMAC PhD School is a cross disciplinary PhD School connected to research
communities within the areas of Languages, Law, Informatics,
Operations Management, Accounting, Communication and Cultural Studies.

Preface and Acknowledgments

After graduating from Political Science in 2009, I started working with online communities in the company, Better Collective. Although government and online communities are quite different fields, I was curious about the intersection of the two. I have been fortunate to satisfy my curiosity by studying this very intersection for the last couple of years through the industrial PhD program. The research presented is the output of this process.

As this dissertation is funded through the industrial PhD program, I have divided my time between Copenhagen Business School and Better Collective. The combination of academic and practice has been very fruitful, as it has helped me conduct rigorous research without getting lost in the infamous ivory tower.

A significant part of this work has been conducted at the Information Science department at Cornell University in 2012. As a visiting research fellow at the department, I gained many valuable inputs through discussions with professors and other PhD students, PhD courses, department presentations, and weekly meetings in the CeRI (Cornell eRule Initiative) research group comprising very inspiring academics such as Professor Cynthia Farina, Professor Claire Cardie and Assistant Professor Dan Cosley. These meetings mainly evolved around the research conducted on CeRI's open government platform, regulationroom.org. The discussions, meetings, presentations, and courses have helped shape the outcome of this dissertation. Especially in terms of research design and methodology, where this research is closer to the quantitative American Human Computer Interaction (HCI) tradition than it is to the qualitative Scandinavian Information Systems (IS) tradition.

Another important part of this PhD process has been the two Work-In-Progress (WIP) seminars held at CBS. The feedback from the first WIP seminar helped me focus in terms of theory and subject matters. Moreover, Helle Zinner Henriksen adviced me to write a monograph rather than an article-based dissertation. The monograph format has helped me focus and thoroughly examine the research

question. The feedback provided in the second WIP seminar helped me understand how far the dissertation had progressed and pinpoint remaining work.

With a single exception (the literature review in Chapter 4 has been submitted to First Monday, and is currently pending review), I have focused on writing this monograph rather than papers. A well-known disadvantage of the monograph format is a lack of ongoing critique from journal reviewers.

However, this does not mean my work hasn't been critiqued at this point. The work has been subject to constructive critique in many ways: Substantial parts of Chapter 2 and 4 have been thoroughly reviewed in the *Advanced HCI* course at Cornell University. The feedback helped me focus more on the important issues, find relevant readings to develop the chapters further, and improved my understanding of academic writing to a great extend. The research approach, Chapter 5, has been thoroughly discussed and pilot tested with other PhD students at Cornell University. These pilot studies helped me understand the American IRB system and, more importantly, helped me develop my statistical skills. The research designs have also benefitted a great deal from the two methodology courses I took at Aarhus University: *Statistics and SPSS* with Kim Berg Johannessen, and *An Overview of the Experimental Methods* with Professor Rose McDermott. An early iteration of the lean experimentation framework (Chapter 10) was presented in a breakfast talk at the Information Science department at Cornell University. The critique by the department faculty has helped elaborate the framework significantly.

The aforementioned is obviously in addition to the constant critique from the many meetings with my advisor, Kim Normann Andersen.

Summary

Governments in more than 55 countries have signed the international *Open Government Partnership* and are currently implementing open government initiatives, aiming to make governments and public sectors more collaborative, participatory, transparent, and technology-driven. If successfully implemented, such open government initiatives can improve democracy, efficiency, and innovation.

As history demonstrates, it is hard to build sustainable online participation. Merely 25% of online communities gather more than 1,000 members in their lifetime. Most of the other 75 % fail due to lack of participation. Many open government communities have shared or are likely to share the same destiny. Giving citizens, companies, and non-governmental organizations the chance to participate in government does not necessarily mean they will do it. Consequently, open government communities face a participation challenge.

Current research shows that the design of the community plays a critical role in participation. Some design patterns foster participation while other patterns discourage it. Existing research also demonstrates that insights from the social sciences can be translated into design ideas and thereby help solve the participation problem.

In this dissertation, social psychology theories are translated into hypotheses about how different open government community design patterns can improve participation. Thus, social psychology theory is used to enrich information system research by increasing the understanding of the motivation to participate. Through online field experiments, four different theories are applied to a Danish open government community, K10. These four field experiments, with respectively n = 3,959, n = 84, n = 29, and n = 453, comprise the empirical data of the dissertation.

With effect sizes varying from small to large, the results of the four experiments are mixed but encouraging: social psychology-inspired design makes it possible to alter and improve participation in K10. The implications for practice are that

open government communities can increase the probability of success by altering the community designs. Despite these encouraging findings, one challenge remains: it is unknown to which degree the findings from this dissertation extrapolate to other open government communities.

To overcome this challenge, I introduce the lean experimentation framework. This experimental framework makes it possible to effectively examine if a given design pattern – inspired by social psychology or other social sciences – have the desired effect in any given open government community. This framework will hopefully serve as a base for future studies and thus help further the academic understanding of what makes open government communities succeed and fail.

Resume (in Danish)

Over 55 lande har tiltrådt det internationale initiativ *Open Government Partnership* og er nu i gang med at implementere open government initiativer, som skal gøre politik og forvaltning mere samarbejdende, deltagende, transparent og teknologidrevet. Såfremt open government initiativerne bliver implementeret rigtigt, kan de forbedre demokrati, efficiens og innovation.

Historien viser, at det er udfordrende at opbygge og fastholde brugerdeltagelse på internettet. Beskedne 25 % af alle online communities tiltrækker mere end 1.000 medlemmer i deres levetid. Størstedelen af de resterende 75 % bliver lukket som følge af manglende deltagelse. Mange open government communities har allerede lidt, eller risikerer at lide, samme skæbne. At give borgere, virksomheder og NGO'er mulighed for at deltage er ikke ensbetydende med, at de deltager. Open government communities står derfor overfor en svær udfordring: at skaffe tilstrækkelig deltagelse til at kunne overleve.

Eksisterende forskning viser, at designet af communities er afgørende for deltagelse. Nogle designmønstre øger deltagelsen, mens andre reducerer den. Forskningen viser endvidere, at samfundsvidenskabelige teorier med succes kan omsættes til designmæssige ideer og dermed kan hjælpe med til at løse udfordringen med at opnå tilstrækkelig deltagelse.

I denne afhandling bliver social psykologiske teorier omdannet til hypoteser om, hvordan forskellige open government community designmønstre kan øge deltagelsen. Social psykologi bliver dermed brugt til at berige informationssystemfeltet ved at øge forståelsen af medlemmernes motivation til at deltage. Gennem felteksperimenter på internettet bliver fire forskellige social psykologiske teorier appliceret på K10, et dansk open government community. Disse fire eksperimenter, med henholdsvis n = 3.959, n = 84, n = 29 og n = 453, udgør afhandlingens empiriske data.

Effektstørrelserne i de fire eksperimenter varierer fra små til store. Dermed er resultaterne blandede, men opmuntrende: social psykologisk-inspireret design

muliggør ændrede og forbedrede deltagelsesmønstre i K10. I praksis betyder det, at open government communities kan øge deres sandsynlighed for at overleve ved at ændre på designet. På trods af disse opmuntrende resultater, er der stadig en væsentlig udfordring: det er uvist, hvorvidt resultaterne kan ekstrapoleres til andre open government communities.

For at imødekomme denne udfordring introducerer jeg lean experimentation frameworket. Dette framework muliggør effektivt at teste, om et designmønster – inspireret af social psykologien eller andre samfundsvidenskaber – har den ønskede effekt i et specifikt open government community. Frameworket vil forhåbentligt blive anvendt i fremtidig forskning og dermed hjælpe til med at fremme forståelsen af, hvilke faktorer, der får et open government community til henholdsvis at overleve og uddø.

Chapter Overview

Part 1: Open Government

Chapter 1:   Introduction

Chapter 2:   Open Government

Chapter 3:   Case Description

Part 2: Theory And Methodology

Chapter 4:   The Social Psychology of Online Communities

Chapter 5:   Research Design and Methodology

Part 3: The Four Experiments

Chapter 6:   Experiment 1

Chapter 7:   Experiment 2

Chapter 8:   Experiment 3

Chapter 9:   Experiment 4

Part 4: Discussion and Conclusions

Chapter 10:   Discussion

Chapter 11:   Conclusion and Perspectives

Bibliography

Appendices

Abbreviations

CSS	Cascading Style Sheet
CTR	Click Through Rate
HCI	Human Computer Interaction
HTML	Hyper Text Markup Language
ICT	Information and Communication Technology
IS	Information Systems
IT	Information Technology
PHP	Hypertext Preprocessor
SQL	Structured Query Language
URL	Uniform Resource Locator
WYSIWYG	What You See Is What You Get

Table of Content

17

Part 1

Open Government

Chapter 1: Introduction

"Another failure: The Municipality of Copenhagen wasted 7 million on a Facebook copy" (Thomsen 2011). Thus read a headline in a Danish IT media in 2011. Copenhagen, the biggest municipality in Denmark, has invested heavily in open government communities over the last couple of years. In 2008, the municipality launched a social network for Copenhagen-citizens called kbh.dk (Friis Sørensen 2007). The community platform carried a DKK 7,000,000 price tag in development. The municipality aimed to gather 125,000 users after two years. But after nine months, only 1,800 had signed up and fewer participated. The community is now closed (Meister 2010; Meister 2011). The Municipality has also invested DKK 5,000,000 in an international community platform for the COP 15 summit called World Climate Community.[1] This community has also been a massive failure and is now closed (Meister 2011). Finally, Copenhagen also had a forum on their website, kk.dk.[2] Due to lack of participation and engagement this forum has been quietly removed from the website.[3]

It is not just the Danish capital that has problems with online participation. The entire Danish citizenry has a single entry point to the public sector called borger.dk.[4] The site is owned and operated by the Danish government, regions, and municipalities. In order to facilitate e-democracy and polls, a debate section was launched on the site in 2010 (Digitaliser.dk 2010).[5] But in the 2010-2012 there had been almost no activity in the community, thereby undermining the potential of the initiative. Today, the debate section has been removed from borger.dk.

---

[1] The community was available at http://www.worldclimatecommunity.org/.

[2] Was available at http://kk.dk/forums.

[3] This explanation was offered by the municipality in a private correspondence.

[4] Translates literally as citizen.dk.

[5] The community was available at http://borger.dk/debat.

The examples above illustrate the difficulties involved with cultivating a successful online government community.[6] In fact, the chance of sustaining an online community over time is low due to under-contribution, and lack of participation is a serious threat (Preece and Schneiderman 2009). Over 9,000 online communities are leveraging Mediawiki software – the software behind Wikipedia. These communities have a median number of 7 contributors in their entire lifetime (Kraut & Resnick 2012). In most cases, a community comprising 7 contributors can hardly be called a success.

A Deloitte survey of over 100 businesses' attempts to build online communities shows that 35% of the communities studied had fewer than 100 users and only 25% had more than 1,000 users (Farzan et al. 2011). According to Farzan et al. (2011) "most business efforts to build online communities failed, even when firms spent over $1 million on the effort, primarily because of difficulties attracting people to the community and retaining them" (Farzan et al. 2011: 1). And failure can be expensive: Many of these companies have spent more than one million dollars on the community

Are these examples really representative for online communities? What about all the successful websites that largely rely on user participation, e.g. YouTube, Facebook, Twitter, and Wikipedia? Even the poster child of successful online communities, Wikipedia, considers lack of participation a threat. In a 2010 letter to the board, The Wikimedia Foundation sees "the health of the editing community

---

[6] Two other examples include Michigan State University and smoke-free.gov. Michigan State University has a branch seeking to be "responsible for bringing the knowledge generated by Michigan"s institutions of higher learning directly to the residents of the state." (Lampe et al. 2011: 3). In order to identify the community's needs and involve stakeholders, the university set up an online community, AdvanceMichigan. The aim was to collect feedback from a diverse set of Michigan residents (Lampe et al. 2011). Despite ambitious outreach attempts, the project failed due to lack of participation (Lampe et al. 2011).

Social support can play a critical role in smoking cessation (Mermelstein et al. 1986). In order to facilitate such social support, the US Government initiated an online community called Smokefree.gov. Despite the good intentions, the online community suffered from under-participation and was therefore terminated (Stoddard et al. 2008; Kraut & Resnick 2012).

as a major risk area for the Wikimedia movement" (Wikimedia Foundation 2010). Moreover, although the English Wikipedia has over 3.6 million articles, the Korean Wikipedia only has 160,000 articles, thereby severely lowering the quality of the encyclopedia (Kraut & Resnick 2012). Creating participation is only the first challenge. Maintaining sufficient participation is also reported a concern for other popular online communities (Sarkar et al. 2012).

1.1. Open Government Communities

Over the last ten years, the web has become increasingly social. It has developed from one-way communication to what we know as the social web, the read-write web, and web 2.0. Today, the most popular websites are leveraging user generated content implicitly or explicitly: Google, Facebook, YouTube, Blogger, Wordpress, Wikipedia, Twitter, LinkedIn, eBay, Craigslist, IMDB, Amazon, and Flickr.[7] All of these websites are successful at least in part because they manage to motivate users to deliver most, if not all, the content.

The success of some online communities has shown the potential value of co-creation, leading to an interest from businesses and government agencies alike (Cullen & Morse 2011). It is therefore hardly surprising that policy makers around the world seek to increase civic participation in government (Osiomo 2008; OECD 2008; Danish IT and Telecom Agency 2011). This movement is called open government, and is largely enabled by recent technological innovations and dissemination.

Open government is defined as

> *"government where citizens not only have access to information, documents, and proceedings, but also can become participants in a meaningful way. Open government also means improved commu-*

---

[7] See for example http://www.google.com/adplanner/static/top1000/ and http://latimesblogs.latimes.com/technology/2011/12/google-facebook-youtube-are-2011-most-visited-websites.html for data about the most visited websites.

> *nication and operations within the various branches and levels of government"*

<div align="right">(Lathrop & Ruma 2010: xix)</div>

As described further in Chapter 2, successful open government creates value in many ways. Open government adds *value to citizens* through (1) increased influence in government, (2) increased civic competencies as a result of participation, and (3) more efficient government and value for taxes. Open government adds *value to companies* by (1) increased privatization, and (2) better return of tax money due to increased efficiency. Finally, Open government adds *value to government* by (1) gaining access to more expert and experiential knowledge, (2) get help with tedious work through crowdsourcing, (3) communicate positive stories and successes, and (4) increased transparency and accountability to fight corruption. But these value propositions can only be realized if citizens, companies and NGOs participate in the open government initiatives.

Open government is a break with the new public management (NPM) paradigm, which has characterized the relationship between governments and citizens since the 1980ies (Hood 1995). Where NPM sees government as service producers and citizens as consumers, open government, as well as new public governance, sees citizens as co-creators of government through active participation (Linders 2011). This is "a radical reinterpretation of the role of policy making and service delivery in the public domain" (Bovaird 2007: 846). Similarly, Linders (2011) sees this development as "a new social contract that empowers the public to play a far more active role in the functioning of their government." (Linders 2011: 167). The move toward open government is not limited to a particular country. As of June 2013, 55 countries have committed to or are planning on committing to the Open Government Partnership (Opengovpartnership 2013).

1.1.1. The Participation Challenge

Most governments in OECD countries are struggling to motivate citizens to use digital self-service solutions instead of analogue. So far, the adoption and usage of these digital services is low (Andersen 2011; Magro 2012). And with open government, citizens not only have to use digital self-service, but are expected to contribute time and knowledge (Linders 2011). Building and sustaining participation in the internet has proved hard for online communities, and it is hardly going to be much easier in the public sector. Particularly not after 20-30 years of NPM, in which the public sector ethos has been replaced with consumer relations (Hood 1995), entailing citizens, companies and NGOs expect not to offer their time and knowledge, but to receive services paid via taxes.

Open government gives citizens increased rights to participate. But giving rights to participate does not equal actual participation. In fact, using these rights is the exception rather than the norm (Farina et al. 2011). Hence the participation challenge. If citizens are not persuaded and motivated to participate in open government communities, these communities will be of little or no use and thus waste resources and potential democratic improvements (Linders 2011). They will end up like the unfortunate examples mentioned in the introduction of this section.

Participation in open government communities is a complex and multi-faceted phenomenon. One of the contributing factors is the community design. An open government community needs to be designed and tailored to increase users' motivation to contribute resources such as free time and expertise (Fayard & DeSanctis 2008). A few examples can illustrate how design affects participation in online communities: As documented further in section 4.3, Drenner, Sen, and Terveen (2008) changed the design on MovieLens and require users to contribute to the community upon registration. This increases post-registration participation on almost 500 %. Cosley et al. (2006) test different recommender algorithms on MovieLens. The best-performing algorithm increases participation with more than 300 %. Cosley et al. (2003) find that the information conveyed through the design influences other users' participation. Users shown positive

information before participating are more likely to be positive themselves. Finally, Beenen et al. (2005) demonstrate that highlighting a contributor's uniqueness increases participation with 18% - 40%. At the same time, a poor design decision can effectively kill an online community. The participation at Answerbag[8] dropped around 50 % following a redesign the users disliked (Gazan 2011).

The positive and negative examples mentioned above have one thing in common: technology and design influences participation in online communities. The current information system (IS) research demonstrates a clear causality in the design-participation relationship: the design of the online community is crucial for the participation – both in terms of quantity and quality (Wasko, Faraj and Teigland 2004; Ma and Agarwal 2007; Venters and Wood 2007; Zhang and Watts 2008; Moon and Sproull 2008; Shen and Khalifa 2009).[9]

Although the literature is quite diverse, most, if not all, online community research finds that it is difficult to start and maintain participation (Butler 1999; Nonnecke & Preece 2000; Nonnecke et al. 2006; Rashid et al. 2006; Burke et al. 2009; Lampe et al. 2010). Despite multiple calls for research in the open government domain (Perry, 2000; Preece and Schneiderman 2009; Johnston 2010), research on the participation problem is scarce. Governmental and international policy makers all over the world are implementing open government policies in order to increase participation, but at the same time, many open government communities fail due to lack of participation.[10] This is a paradox that needs to be solved in order to succeed with open government. At this point, neither infor-

---

[8] http://www.answerbag.com

[9] It can certainly be argued that the *design -> participation* causality co-exists with a *participation -> design* causality. The latter constitutes an important part of the participatory design movement and Scandinavian IS in general. This second causality is, however, outside the scope of this dissertation, although it is briefly discussed in Chapter 10.

[10] Preece and Schneiderman (2009) note "that many web sites fail to retain participants, tagging initiatives go quiet, and online communities become ghost towns. Many government agencies are reluctant to even try social participation, fearing public uprising, pornography, or slander." (Preece and Schneiderman 2009 15).

mation system research nor practice has sufficient knowledge about why citizens, companies, and NGOs contribute to open government communities. This is the **research gap** I seek to address in this dissertation.

1.1.2. Social Psychology: A Potential Solution to the Participation Challenges

Although the social web is fairly new, academic research on social interactions is not. As described further in Chapter 4, several social science disciplines hold the potential to facilitate improvement in participation in open government communities. From the set of disciplines, social psychology is selected to serve as a foundation in design tactics leading to increased user contributions. The social psychology field is well developed and serves as a sound foundation for information system research into online social interactions (Beenen et al. 2005; Cosley et al. 2006; Chen et al. 2010; Burke, Marlow, and Lento 2009; Kraut and Resnick 2012). It is therefore hypothesized that social psychology theory also can help increase participation to online communities in the government sphere.

Social psychology theories can be used to generate design ideas to be tested in an open government community. The theories are applied through experimental methodology, in which a construct of each of the theories is used in controlled settings. The purpose of this research is to improve open government communities, not evaluate the theories.

The selected theories are applied to a Danish open government community called K10. [11] K10 is an online community for people involved with two public benefit programs called "early retirement pension" and "flexjob". The community was launched in 2005. Although it is owned by a Danish citizen, K10 operates in the government sphere as the discussions evolve around legislation, requirements, benefit rates, terminology, rulings, and rights to repeal and complain. As of June 2013, K10 has over 12,000 users, whereof 6,000 have written a post. Please refer to Chapter 3 for a detailed description of K10.

---

[11] http://www.k10.dk

1.2. Research Question

In this dissertation, I examine and answer the following research question: Can social psychology theory help increase participation in open government communities?

As thoroughly described in Chapter 5, this dissertation relies on experimental design and analysis. Four different experiments are conducted and reported, and thus constitutes the empirical data in this dissertation. An integral part of this methodology is the proposition of research questions and hypotheses to be examined with the means of statistical tests. Each of the four experiments is guided by a research question and set of statistical hypotheses.

1.2.1. Research Question and Hypotheses in Experiment 1

It is difficult to know how often to contribute to an open government community since there is no objectively "right amount" of activity. One way to determine the socially accepted and appropriate level of contribution is by comparing yourself to other users. Social comparison theory contends that people compare themselves to others to help identify appropriate behavior and evaluate capabilities, opinions and actions. This experiment therefore seeks to leverage social comparison theory to increase participation in the open government community, K10.

The following question asked is: *How does exposure to social comparison information affect users' participation in K10?* Derived from social comparison theory, a null hypothesis and six alternative hypotheses are proposed in order to answer the research question.

H_0 Receiving social comparison information has *no* effect on subsequent participation.

H_a Receiving social comparison information has an effect on subsequent participation

H_{a1}	Receiving social comparison information reduces the contribution level among above-mean contributors
H_{a2}	Receiving social comparison information increases the contribution level among below-mean contributors
H_b	Receiving non-comparative information has an effect on subsequent participation, but smaller than comparative information.
H_{b1}	Receiving non-comparative information increases the contribution level among above-mean contributors
H_{b2}	Receiving non-comparative information increases the contribution level among below-mean contributors

1.2.2. Research Question and Hypotheses in Experiment 2

In order to attain certain behavior, it can be helpful to define goals. As described in depth in Chapter 2, a characteristic of open government is that participation is voluntary. This means that goals cannot be combined with financial incentives normally used within public and private organizations.

But can goals have an effect on participation in an open government community even though the goals are not coupled with sticks or carrots? Goal-setting theory predicts that they can. In this experiment, the subjects are asked to commit to a participation goal. The aim of the experiment is to examine if the subjects live up to their goal commitment. If the subjects assign themselves high goals and subsequently live up to them, such goal assignment might help open government communities overcome the participation challenge.

The question asked in this experiment is: *How does a self-assigned goal to a number of contributions to K10 affect subsequent participation?*

In order to answer this research question, one null hypothesis and one alternative hypothesis are proposed. The hypotheses are derived from goal-setting theory.

H_0	Setting a participation goal has no effect on subsequent participation
H_a	Setting a participation goal makes subjects strive to achieve the goal

1.2.3. Research Question and Hypotheses in Experiment 3

A person's beliefs about own capabilities matter and influence behavior. If one believes she has the chance to succeed in a specific situation, she will approach the task differently than she would if she did not hold such beliefs. This is at least what self-efficacy theory asserts. According to this theory, people like to have an effect on their environment and are more motivated to do certain actions when they know these actions have an effect on other people.

If self-efficacy theory holds validity in the open government community domain, it might be possible to increase participation, and thus help solve the participation challenge by increasing the users' self-efficacy. This leads to the following research question: *How does knowledge of other users' gratitude for previous contributions affect future contributions?*

In order to answer this research question, one null hypothesis and one alternative hypothesis are proposed:

H_o	Knowledge about users gratitude for the subjects' previous contributions does *not* affect the subjects' future contributions
H_a	Knowledge about users gratitude for the subjects' previous contributions affects the subjects' future contributions

1.2.4. Research Question and Hypotheses in Experiment 4

Social identity theory argues that people seek to enhance the groups they belong to in order to enhance their own self-image. Moreover, people often prefer to contribute to groups comprising similar others. It is therefore expected that the benefiters of a contribution affect users' participation level.

If it is clear to a user who is benefiting from her participation, she might partici-
pate more. In order to examine this belief, the following research question is
proposed: *How does benefit of contribution affect subsequent behavior on K10?*

Based on social identity, one null hypothesis and two alternative hypotheses are
tested:

H$_0$ The group constructs have no effect in terms of engagement and contri-
           bution behavior.

H$_a$ Subjects assigned to the *benefit other* groups will engage and contribute
           more than subjects assigned to *benefit self* and control group.

H$_b$ Subjects assigned to the *small benefit other* group will engage and con-
           tribute more than subjects assigned to *large benefit other* group.

1.3. Operationalizing the Research Question

The aim of the remainder of the dissertation is to answer the overarching re-
search question:

*Can social psychology theory help increase participation in open government
communities?*

The next sections explain how this research question is operationalized through
the research design. As the literature review in Chapter 4 shows, at least 12 dif-
ferent social psychology theories have been used to examine and increase partic-
ipation in non-governmental online communities. The research question is in-
tended to examine to which extend this theoretical body also can be used to ex-
amine and increase participation in communities operating in the government
sphere.

As mentioned, the dissertation comprises four different experiments. Each ex-
periment relies on a different social psychology theory. Ideally, this dissertation
research provides exhaustive answers to this ambitious research question. How-

ever leveraging 12 different theories are outside the scope of a single dissertation. The aim is to answer the research question by translating several, but not all theories, into design patterns that subsequently are tested on an open government community. Based on two criteria, fit to case and granularity level, four theories are selected and operationalized Chapter 4.

Dissertation Structure

Research question	Chapter 1
Open government community: K10	Chapter 2-3
Social psychology theory	Chapter 4
Research design and methodology	Chapter 5
Experiment 1 Experiment 2 Experiment 3 Experiment 4	Chapter 6-9
Theoretical, methodological & empirical insights	Chapter 10
Answer research question	Chapter 11
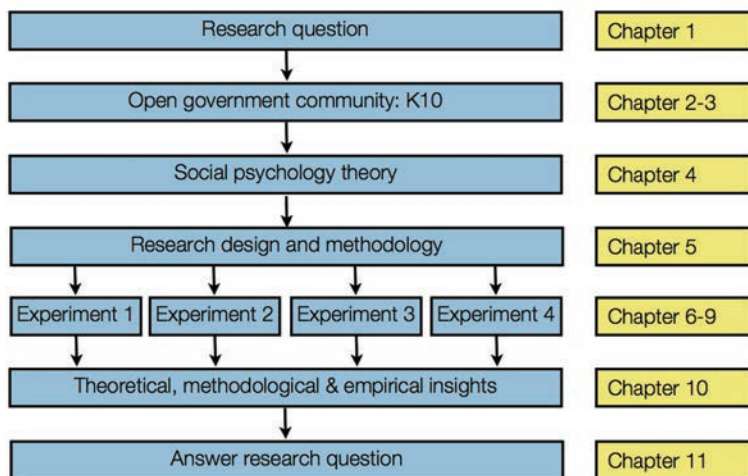
Figure 1: Dissertation structure and operationalization of the research question

The process is illustrated in Figure 1. The figure shows that four theories are applied to a case, which generates some specific results of the experiment. The sum of the results from the four experiments is expected to generate new and unique insights into the appliance of social psychology theory to open government communities, and thus answer the research question.

As explained Chapter 4 and Chapter 5, experimental methodology is considered to hold a great potential when it comes to online community design. To generate as many relevant insights as possible, three different experimental designs are used, and four different proxies are applied to conduct the intervention. The three experimental research designs applied are between-subject design, within-

subject design, and split-plot design. The four proxies used to make changes to the independent variable are email, internal community messages, survey, and advertisement in Google.

The aim of this diversity in the experiments is to explore the opportunities and limitations experimentation offers in open government communities. The aim is not to make statistically valid claims about which research design and proxy is the best in open government research. The high level aim is rather to explore how experimentation in general aligns with open government research by exploring various designs and proxies.

1.3.1.1. Knowledge Used to Answer the Research Question

When answering the research question, two types of knowledge gained from the four experiments are used. The *first* type of knowledge is direct empirical knowledge. This type of knowledge is outcome-based and reported directly in each experiment in Chapter 6-9. This knowledge comprises significance, p-values, effect sizes and other substantial results. I.e. to which extend the theory-based intervention causes a change in the behavior, and to which extend the experimental design works in the context.

The *second* type of knowledge is indirect meta-knowledge stemming from conducting the experiments. This type of knowledge is process-based and collected during the experimentation phase. This type of knowledge comprises meta-insights, reactions, problems, concerns, opportunities, and other elements not directly visible in the quantiative results.

Combined, these two types of knowledge help answer the research question with very specific quantitative data and at the same time offer a broader and more holistic interpretation of the experiments and this methodology's potential in the open government domain.

1.4. Scientific Communities and Contributions

As this dissertation operates in different scientific domains it is relevant to describe how the different domains are leveraged, how they interact, and how this dissertation contributes to these scientific communities.[12]

Gläser (2006) argues that a researcher is a member of a scientific community when he contributes to the shared knowledge base.[13] However, researchers are facing increasing difficulties in delineating scientific and technological fields (Zitt & Bassecoulard 2006), in part because research fields are increasingly interdependent and complex, but also because the vocabulary and the institutional structure in emerging fields change quickly (Zitt & Bassecoulard 2006).

Due to these complexities and interdependences in emerging research fields, this research adheres to more than one research field.[14] In this dissertation, I refer to and communicate with three research communities: one primary, and two secondary research communities. The primary community is information systems, and the two secondary are open government and social psychology.

---

[12] Researchers are organizing into scientific communities where they share a set of subjects and communicate with each other (Kornfeld & Hewitt 1981). This form of organization enables knowledge aggregation and scientific progress at a faster rate. The downside to this method of organization is that it entails parallelism.

[13] According to Gläser (2006), scattered research turns into a scientific community when research is shared in common knowledge bases through journals, conferences, and/or workshops.

[14] It is necessary to rely on different scientific communities in order to solve complex challenges. For example, Fedorowicz & Dias (2010) note that the domain of digital government relies heavily on the information system (IS) discipline.
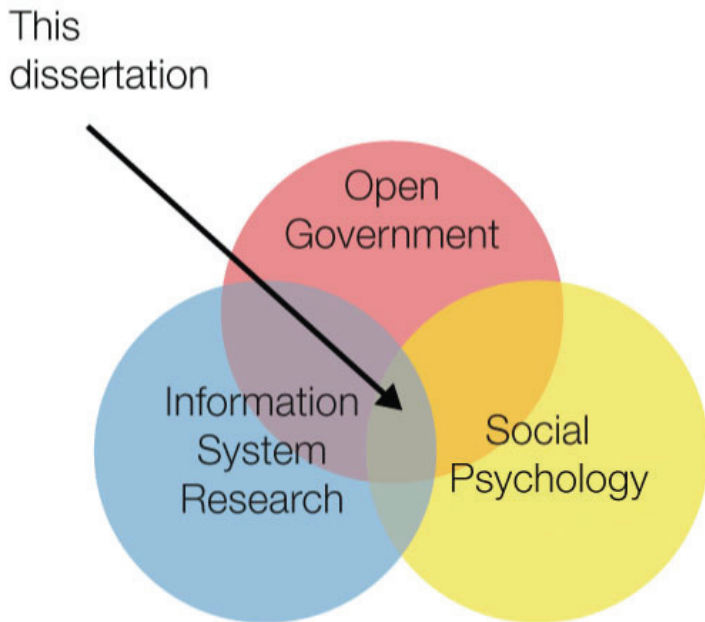
Figure 2 shows where this dissertation is located in an interdisciplinary field comprising three scientific domains.

1.4.1. The Information Systems (IS) Community

The research question of this dissertation revolves around the design of open government communities. The examination of how design affects interactions in online communities is usually conducted in the Information Systems (IS) field. IS research is therefore considered the primary community, entailing that this dissertation principally aims to contribute to the IS knowledge base. IS can be de-

fined as "the development, use and application of information systems by individuals, organizations and society" (Baskerville and Myers 2002).

IS research is often relying on reference disciplines in order to understand, explain and improve information systems (Baskerville and Myers 2002). In this dissertation, the IS field is therefore enriched with social psychology theory in order to understand and improve participation. IS research comprises many sub-disciplines[15] and perspectives, one of them being Human Computer Interaction (HCI) (Zhang et al. 2002).[16] HCI in the IS field "are concerned with the ways humans interact with information, technologies, and tasks, especially in business, managerial, organizational, and cultural contexts" (Zhang et al. 2002: 334).

The HCI community arose in the 1970ies.[17] In the early days, HCI was seen as an applied science, aiming to use psychology and cognitive sciences to improve design (Carroll 2003). Thus, HCI does not have a tradition for developing own theories, but rather relying on reference disciplines. The theoretical aspects have been strengthened over time, which is often seen as a sign of maturity (Clemmensen 2006). Since the 1990ies the HCI field has become increasingly broad, fragmented, and inclusive (Carroll 2003).

One critical assumption in the HCI community is that design matters. In the context of online communities, it is assumed that user behavior in online communities at least in part is a response to the design (Wasko, Faraj and Teigland 2004;

---

[15] These sub-disciplines include Information systems management processes, Information systems development processes, Information systems development concepts, Representations in information systems, and application systems (Baskerville and Myers 2002).

[16] It should be noted that HCI often is considered a distinct research field.

[17] HCI arose as a response to the predominant waterfall model. Design was in the bottom of the waterfall, thereby being introduced too late in the process to have an impact (Carroll 2003).

Ma and Agarwal 2007; Venters and Wood 2007; Zhang and Watts 2008; Moon and Sproull 2008; Shen and Khalifa 2009).[18]

The research question of this dissertation falls within a classic HCI scope in which the aim of the research is to "validate novel design ideas and systems, usually by showing that human performance or work practices are somehow improved when compared to some baseline set of metrics" (Greenberg and Buxton 2008: 111). This type of research is applied to the open government domain in order to help solve the participation problem.

According to Caroll (2003), HCI

> *"lies at the intersection between the social and behavioral sciences*
> *on the one hand, and computer and information technology on the*
> *other. It is concerned with how people make use of devices and sys-*
> *tems… and how such devices and systems can be more useful and*
> *usable"*

> (Caroll 2003: 1)

This definition clearly conveys how broad and multifaceted the HCI field is. In a more specific description of the HCI field's scope, Hewett et al. (2009) delimit the scope to the following content: use and context of computers (including human social organization and work, application areas, and human-machine fit and adaptation), human characteristics (including human information processing, language, communication, interaction, and ergonomics), computer system and interface architecture (including input and output devices, dialogue techniques, dialogue genre, computer graphics, and dialogue architecture), development process (including, design approaches, implementation techniques, evaluation

---

[18] For example, Ma and Agarwal (2007) argue that "specific features … are also likely to facilitate or constrain how actors within the community relate and interact with each other." (Ma and Agarwal 2007: .44).

techniques, and example systems and case studies), and project presentations and examinations.

This dissertation is mainly concerned with voluntary human social participation to online communities. This means this dissertation only covers a small corner in the HCI field. The same applies to HCI methodologies. A plethora of methodologies are used in the field including experiments, statistical analysis, surveys, diaries, case study, interviews, focus groups, ethnography, usability testing, and participatory design (Lazar et al. 2010). The different methodologies differ in terms of ontology, epistemology, and examined data type, and they have different comparative advantages and disadvantages. Only one of the methodologies, experimentation, is used here. Again, this dissertation only is concerned with a small part of the IS and HCI fields.

The primary research outlets are the ACM CHI and CSCW conferences. Several major IS journals such as *MISQ* and *Information System Research* sporadically publish online community studies.

1.4.1.1. Contributions

A comprehensive HCI literature body on online communities exists. The experiments conducted in this dissertation help examine the boundaries of this research by extending the research into the open government sphere. Based on these experiments, the differences between governmental and non-governmental settings are discussed.

1.4.2. The Open Government Community

The open government domain is being developed in both theory and practice. Although different actors have different goals, most researchers and practitioners agree that the overarching goal of open government research is to examine the possibilities and implications of collaboration and transparency by using interactive web technologies in government. In this dissertation, open government provides the background and empirical subject matter that is studied. But the

primary contribution is to IS, as this perspective helps overcome the participation problem. Thus, IS provides a fruitful perspective to study open government communities.

Open government literature is published in various journals and conference proceedings. Many journals only have a couple of open government articles, e.g. *Governance:* An *International Journal of Policy, Administration, and Institutions*, and *Public Administration Review*. Two journals are consistently publishing open government articles: *Government Information Quarterly* and *European Journal of ePractice*. It is in these two journals a significant part of the open government knowledge base is currently built. Popular conferences and conference proceedings include *International Conference on Theory* and *Practice of Electronic Governance* (ICEGOV) and *Digital Government Research*.[19]

1.4.2.1. Contributions

The field of open government is rapidly emerging. Research is spread out across many different research and practice communities. As further described in chapter 2, the field is still evolving and settling, making it hard to describe in concise terms. This dissertation makes two contributions: (1) grounding the domain through a comprehensive literature review, and (2) a framework to conduct more experiments in theory and practice.

1.4.3. Social Psychology

The primary subject matter of this dissertation is open government. Social psychology, along with HCI, is used as a reference discipline to examine and improve open government by relying on established and validated theories.

Although the social web is fairly new, academic research on social interactions is not. The discipline is starting to adapt to the web, but the web is often used as a

---

[19] One of the leading research institutions is the Cornell eRule Initiative (CeRI). In 2012, I spent 6 months at Cornell University participating in the eRule initiative. I therefore feel strongly embedded in this part of the open government community.

mean to collect data rather than studied as a subject matter per se. Nonetheless, most social psychologist find the internet a legitimate research area (e.g. Reis & Gosling 2010) and find research into online support groups, social networks message boards, and blogs important (Zadro, Williams and Richardson 2004; Beenen *et al.* 2005; Zanbaka et al. 2007; Jonson 2008; Reis and Gosling 2010; Toma 2010).

Some of the most important social psychology publications include *European Journal of Social Psychology, Social cognition, Journal of personality and social psychology, Personality and Social Psychology Bulletin* as well as the classic *The Handbook of Social Psychology.*

1.4.3.1. Contributions

In this dissertation social psychology is used to examine and improve open government, meaning that it is primarily a one-way relation from social psychology to open government. But only primarily, since the mere application of social psychology theory is likely to give new insights into the validity boundaries of the given theories. However, this is more a benevolent unintended consequence rather than the goal per se.

Starting with the work of Beenen et al. (2004), many social psychology theories have been applied to online communities. This dissertation contains a comprehensive literature review of this research, in which it is described, categorized, and analyzed. The contributions include (1) clarifying how various social psychology theories have been used to understand and improve online communities, and (2) identifying gaps for further research.

1.4.4. Relations between the Three Scientific Communities

The arrows in Figure 3 below show that IS research is the primary community, and open government and social psychology are secondary.
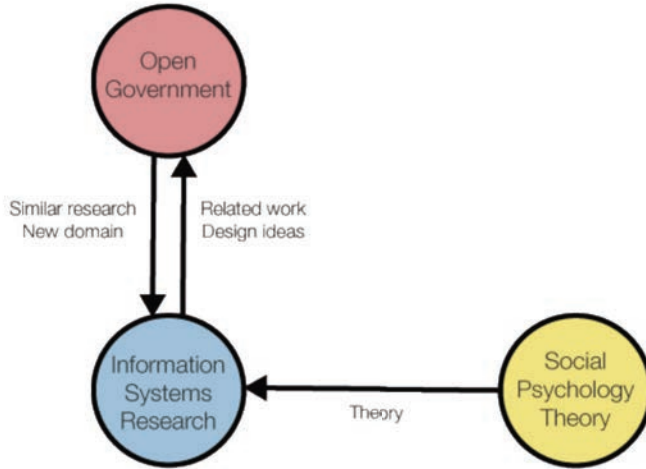
Figure 3: Relations between primary and secondary research communities

Figure 3 takes the point of departure in the primary scientific community, information systems.[20] The Figure illustrates how the research gap, located in the information systems domain, is addressed by leveraging open government and social psychology. Figure 3 shows the relationships between the research communities studied in this dissertation and is not a general model.[21]

1.4.5. Summary of Research Contributions

Table 1 summarizes how this dissertation seeks to learn from and contribute to three related domains: open government, information systems, and social psychology.

	Information Systems	Open Government	Social Psychology
Learning from	Research designs	Issues and challenges	Theoretical works

[20] For simplicity and relevance reasons, this figure only shows the relationship between information systems and the other two scientific communities. It does not mean that there are no relationships between other research communities.

[21] Such general model would be much more complicated and contain more links between the four nodes.

	Research ideas	Prior research	Methodology
Contributing to	Substantial knowledge about participation through four experiments	Grounding the domain	Literature review of application to online communities
	Replication (partly)	Substantial knowledge about participation through four experiments	
	Testing boundaries of existing findings	Lean Experimentation framework	
		Thorough case description	

Table 1: Summary of learning and contributions to the three research communities

1.5. Key Concepts

This dissertation positions itself in an interdisciplinary field. As different fields have different languages, focus points, and definitions, it is useful to define two key concepts used in the dissertation: online community and contributions.

1.5.1. Online Communities

Before *online communities* were *communities*. The term *community* is derived from the Latin words *com* (together) and *unus* (singularity), and describe a form of social organization (Delanty 2009). The term has been used for at least 2,300 years, as Aristotle used the community term to describe a spatial environment such as a city. The term community is quite ambiguous. As recent as 50 years ago George Hillery found 94 different definitions of community (cited in Fernback 1999). People are the only consistent trait in different community definitions (Fernback 1999; Reich 2010; Wilson 2010).

Although it is not particular clear what a community is, it is usually considered desirable because it creates stability, a place for belonging, solidarity, civic bonds, social identity, and recognitions (Delanty 2009). But communities also have their dark sides. Communities foster interactions which can be positive and negative (Gusfield 1978). Conflicts and violence is thus an integral part of com-

munities. And strong communities also tend to create tension against other communities (Delanty 2009).

Online communities have roots in the 1970s with the establishment of Listservers. Further technological improvements in the 1980s entail bulletin boards, usenet groups and synchronous chat systems such as Internet Relay Chat (Preece et al. 2003). A significant technological landmark is the 1991 invention of the web. The web was broadly adapted in the late 1990s, thereby allowing a more diverse online population that is closer to reflect the real – offline – population.[22] Today, the web is serving as the platform for most online communities (Preece et al. 2003), although it is rapidly assisted by and shifting to mobile applications.[23] Anecdotal research on online communities was published in the 1990s (Preece et al. 2003), but theory-based research is a rare sight before the turn of the millennium. The time frame included in the literature review in Chapter 3 is therefore starting in 2000.

As described above, there is no commonly accepted definition of community and putting online in front of it hardly reduces the complexity. There is no canonical definition of online communities (Stanoievska-Slabeva and Schmid 2001; Lee, Vogel and Limayem 2003; Porter 2004). Quite contrary, dozens of definitions exist, making it hard to know if different research is addressing the same domain.[24] The essence of online community definitions can be condensed to eight traits: People, purpose, time, space, interactions, technologically mediated, artifacts, and governance. In a recent book aiming to synthesize much of the re-

---

[22] With the establishment of Listservers in the 1970s, the first online communities were formed. Further technological inventions in the 1980s lead to new forms of online communities, including bulletin boards, usenet groups, and synchronous chat systems such as Internet Relay Chat (Preece, Maloney-Krichmar and Abras 2003). A technological landmark is the 1991 invention of the web, which was broadly adapted in the late 1990s and, lead to a proliferation of online communities.

[23] Foursquare (http://www.foursquare.com) and Instagram (http://instagram.com) serve as examples of online communities that are primarily mobile driven.

[24] A debated dichotomy is online community and social networks. However, "The difference between online community … and social network is more or less arbitrary. The rule of thumb is that that 'community' implies a tighter social bond and inclusiveness than 'network'." (Anttiroiko 2010: 20).

search into online communities, Kraut and Resnick (2012) propose the following definition:

> *"By 'online communities' we mean any virtual space where people come together with others to converse, exchange information or other resources, learn, play, or just be with each other. The term applies to many social configurations, from small close-knit groups to sites with millions of participants. Online communities may be supported by a wide variety of technology platforms, from email lists to forums, blogs, wikis, and networking sites. The common feature is ongoing interactions among people over time, with some of the interactions being technology mediated."*
>
> (Kraut & Resnick 2012: 1)

This is a broad definition of online communities, and covers all important traits of online communities, except one: artifacts. The reason why online communities are perceived as decisively valuable in society is that they create artifacts, such as photos, movies, music, text, metadata, product data, geo-data, and software.

As described earlier in this chapter – and thoroughly described in Chapter 3 – the community used as case for the experiments in this dissertation is called K10, and consists of people applying for government benefits in Denmark.

1.5.2. Contributions

Contributions are important since they often constitute the explicit value, i.e. the artifacts, of an online community (Olsson 2009). Olivera et al. (2008) define contributions as "voluntary acts of helping others by providing information." (Olivera et al. 2008: 23).[25] The act of contribution can take many forms, depending

---

[25] Although useful, this definition of contributions can be contested on two grounds. *First*, seeing contributions as "voluntary acts of helping others by providing information" comprises only explicit contributions, i.e. providing information to other users. A user who does not post anything is often referred to as a lurker (e.g. Nonnecke and Preece 2000). Although active participation is the

on the design of the community. Typical types of contributions include posts (Ludford et al. 2004; Nambisan & Baron 2009), articles (Clever et al. 2009), photos (Nov, Naaman and Ye 2008; Burke, Marlow and Lento 2009), video (Cha et al. 2007; Krumm et al. 2008), music (Cheliotis and Yew 2009), ratings (Rashid et al. 2006; Cosley et al. 2006), reviews (Clever et al. 2009), events (Farnham, Brown, and Schwartz 2009), tagging (Nov et al. 2008), and geo-local information (Lindqvist et al. 2011).

Following the definition of the two key concepts, online community and contribution, this chapter is concluded with an overview of the dissertation by briefly summarizing each chapter.

1.6. Dissertation Structure and Chapter Summaries

In the remaining part of **Chapter 1**, the dissertation structure is presented and each chapter briefly summarized.

Chapter 2 is an overview and grounding of the open government field. In this chapter, open government is defined and placed in a broader Egovernment context. In order to understand the aim of open government, two different and sometimes competing open government perspectives, collaborative democracy and deliberative democracy, are reviewed. The main part of the chapter is dedicated to understand the promises, critiques, and value propositions of open government.

---

most visible part of online communities, it does not follow that passive participation has no value to online communities (Rafaeli et al. 2004). Lurkers contribute indirectly by being the audience to the active members of the community, and by generating revenue through pageviews and clicks on ads (Rafaeli et al. 2004).

Secondly, contributions do not always help others. For example, uploading photos to a social network is a contribution to the common pool of information, but it does not *necessarily* help other users. If a photo is uploaded but goes unnoticed, it can hardly be categorized as helpful to others. The perceived value of any particular contribution is often unknown. Thus, contributions *potentially* help other.

After describing and analyzing the open government field, **Chapter 3** describes an open government community called K10. K10 is an online community for people involved with two public benefit programs called "early retirement pension" and "flexjob". K10 is used as the field for the four online field experiments conducted and reported in Chapter 6-9. Conducting experiments on a real open government community will help establish to what extend social psychology theory can help increase open government participation.

To leverage existing knowledge regarding behavior in social contexts online, **0** offers a literature review on the use of social psychology theory in online community research. This chapter relies on research from two scientific communities: Social psychology and HCI. In this chapter, 12 different social psychology theories are presented, and their appliance into online communities analyzed. Four of the theories are selected and operationalized for further use in this dissertation. The aim of this chapter is to identify theoretical insights and assumptions that can help set up fruitful online field experiments and thus help solve the participation challenge.

Theory and practice can be connected in multiple ways. **Chapter 5** describes how the social psychology theories are translated into HCI research. In addition to describing the overarching ontological, epistemological, methodological, and axiological positions of this dissertation, experimental methodology in online settings is thoroughly discussed. Four different intervention strategies are set up, which forms the foundation of the research conducted and reported in Chapter 6-9. Thus, the overarching methodological considerations are presented in Chapter 5 while the details that vary across the four experiments are in Chapter 6-9. The reason behind this structure is that some considerations apply to all four experiments while others depend on the specific experimental design.

The first experiment, reported in **Chapter 6**, seeks to answer this question: *How does exposure to social comparison information affect users' participation in K10?* The question is answered through a split-plot experiment containing 3,959 subjects. The intervention is made with emails, whereby subjects in different groups

are receiving emails with different participatory information. The experiment leverages social comparison theory.

The second experiment, reported in **Chapter 7**, is concerned with this question: *How does a self-assigned goal to a number of contributions to K10 affect subsequent participation?* Based on goal-setting theory, this experiment measures the effect of asking the subjects to give an expected number of posts in the following two-week period, thereby making an implicit contribution commitment. 84 subjects are answering through a survey and another 3,998 subjects are used as a control group.

In the survey reported in Chapter 7, the respondents are asked to name at least one other K10 user who is making particularly important contributions. Based on these answers, the experiment reported in **Chapter 8** is passing on this positive feedback to the subjects. The message is delivered via K10's internal community messages and email. Leveraging self-efficacy theory, this third experiment addresses the following question: *How does knowledge of other users' gratitude for previous contributions affect future contributions?* The experiment is a within-subject design with 29 subjects.

The fourth and final experiment is reported in **Chapter 9**. This experiment answers this question: *How does benefit of contribution affect subsequent behavior on K10?* In this experiment, advertisements for K10 are bought at Google.dk. Four different ad copies are used to rime the subjects with different behavioral messages. Based on social identity theory, the four groups respectively seek to highlight different benefits of making a contribution at K10: *benefit to self*, *benefit to small group*, *benefit to large group*, and *no benefit*. This experiment has 453 subjects.

The four experiments are summarized in **Chapter 10**, and thus serve as the foundation for the discussion. The four experiments offer rich insights into the behavior behind contributions to online communities in the open government sphere. But they do not offer a complete picture. As the need for further research

in the area seems vast, an experimental framework is proposed in this chapter. This framework, lean experimentation, seeks to create a process of experimentation that examines assumptions through the experimental proxies used in chapter 6-9. Making experimentation easier is seen as one of the most effective ways of improving open government communities, as they can be conducted at low cost in both research and practice. The remainder of chapter 10 discusses other implications of the research in open government.

Finally, the conclusions in **Chapter 11** offer a detailed summary of the substantial, theoretical and methodological research results. Based on these insights, a number of suggestions for further work in research and practice are proposed.

1.7. Chapter Summary

Policy makers in governmental and international organizations are implementing open government policies in order to increase citizens' participation in government in online environments. At the same time, many open government communities fail due to lack of participation – even though significant resources are invested in them. This is a paradox that needs to be solved in order to succeed with open government. At this point, neither research nor practice has sufficient knowledge about what motivates citizens (and other non-governmental actors) to contribute to open government communities. The aim of this dissertation is to help solve this participation problem.

Current research shows that social science disciplines such as social psychology hold a potential to improve the design and participation in online communities. It is unknown to which degree this is true for open government communities. This research gap leads to the following research question: *Can social psychology theory help increase participation in open government communities?*

The aim of this dissertation is to answer this research question. In order to do this, three scientific communities are involved in the work: open government, information systems research, and social psychology. Information system is the

main community and the other two are predominantly used as reference disciplines to facilitate improvement in participation to open government communities.

Chapter 2: Open Government

On his first day as president of the United States of America, Barack Obama released *Memorandum on Transparency and Open Government*, stating:

> *"My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government."*

> (Obama 2009: 1)

Since his first day in office, President Obama has taken several measures to implement open government in federal agencies (Farina et al. 2010; White House 2011). The most important step is probably the Open Government Directive from 2009, setting deadlines for agencies to publish information online, improve information quality, create a culture of open government, and create an enabling policy framework plan for open government.

It is not only the US who is heading towards open government. The European Union started the effort back in 2003 with a directive to re-use public sector information (Janssen 2011). But it is not limited to the EU and USA either. Open government is a global phenomenon comprising over 55 countries from five continents (Nam 2011; Open Government Partnership 2013).

The open government movement is also progressing at an intergovernmental level as institutions such as the EU and OECD are promoting the development of open government (Osimo 2008; OECD 2008; United Nations 2008). Other international institutions such as the World Bank have implemented open government communities by offering much of their data for free.[26]

---

[26] The World Bank's data is available at http://data.worldbank.org/.

On a high level, the aim of this chapter is to describe why open government matters and is worth writing a dissertation about. More specifically, this chapter seeks to contribute to open government in research and practice by grounding the field and describing the key concerns. Grounding the field helps understand what open government is, why it matters, and how research can contribute to successful open government implementation.

The chapter progresses in the following way: It starts with a definition of open government and a description of how it relates to egovernment. Subsequently, the advantages and disadvantages of open government are examined and the value propositions described. Based on these value propositions, it is clear that open government has a significant potential to citizens, companies, and governments. But in order to succeed, the participation problem needs to be addressed. This is the gap this dissertation seeks to address. The chapter is concluded with a description of how open government is implemented today.

2.1. Defining Open Government

In political science, there is a sharp distinction between administration and politics, between the executive branch and the legislative branch. But this distinction is often lost in interdisciplinary open government research (e.g. Anttiroiko 2010), possibly because it is rarely rooted in political science.[27] Consequently, some scholars and practitioners use open government to describe politics, while others use it to address public administration. Others are more stringent and use it only in terms of public administration (e.g. Christopher Hood 2007; Nam 2011; Farina et al. 2011). In this dissertation, I am concerned with open government in the sphere of public administration. It focuses on how citizens can participate in public administration, and excludes how citizens can be more engaged in electoral and parliamentary processes (e.g. Missingham 2011). As will be discussed

---

[27] From a political science perspective, this makes the research focus somewhat ambiguous. For example, in an anthology on open government, the editors, Lathrop and Ruma (2010), use the term to describe both changes in politics and administration. One explanation of this lack of distinction is that the line between e-democracy and e-government is blurry (Harrison et al. 2011).

later in this chapter, democratic processes are indeed a part of this scope, but political elections are not.

At this point in time, there is no canonical open government definition.[28] A large number of terms with partially overlapping meaning are being used e.g. citizen sourcing, collaborative government, wiki government, co-governance, e-participation, do-it-yourself government, isocratic administration, government as a platform, government 2.0[29], and open government (Linders 2011). Many

---

[28] An emerging research field can usually be characterized as a fragmented adhocracy (Sanford & Rose 2007). This applies very much to open government, where the nomenclature is unclear.

[29] A point of confusion is the usage of *open government* and *government 2.0* respectively. Some scholars use government 2.0 (e.g. Anttiroiko 2010; Nam 2011) while others use open government (e.g. Hood 2009; Coglianese 2009). Other scholars use the two terms interchangeably (e.g. Chun et al. 2010). A potential difference between government 2.0 and open government is where it is being used. Government 2.0 definitions often stem from practitioners (see for example Nam 2011 that draws on DiMaio 2009 and O'Reilly 2010), whereas open government draws on academia (e.g. Hood 2009; Jaeger and Bertot 2010; Farina et al. 2011). Nam (2011) argues that open government is the end, and government 2.0 is the technology to achieve this end. The focus on government and technology is the primary difference between the two terms, insofar there is any difference.

In this dissertation, I use the term open government, not government 2.0, for five reasons: *Firstly*, government 2.0 will probably be outdated by the time this dissertation is submitted, even though the subject matter stays the same. Professionals and scholars both have an incentive to be the first to coin government 3.0, whether or not such a term is deemed necessary. Both academics (Anttiriko 2010) and consultants (Di Maio 2009) have started to write about government 3.0. *Secondly*, many people are already fed up with web 2.0 and government 2.0 hype (Skiba 2006; Best 2006). A reason to avoid government 2.0 is therefore to avoid unnecessary negative connotations. *Thirdly*, some of the open government technologies are not adapted from web 2.0, but novel innovations. This includes. e-rulemaking platforms such as Regulation Room (http://regulationroom.org). *Fourthly*, the term of open government is increasingly adapted after President Obama used it in his Memorandum on Transparency and Open Government. *Fifthly*, and most importantly, open government highlights the more fundamental aspect of the phenomenon, namely that it is about openness in government and not just technology. As the focus in this dissertation is on citizen behavior rather than technology per se, open government seems to be the more appropriate term. Like Chun et al. (2010), I consider the terms to be synonyms and include definitions of both in this section.

different definitions are used in academia, on blogs, and in practice. As explained by Schuler (2010), the idea of open government is still being constructed.[30]

2.1.1. A Definition

The term open government has been around since the 1950s. It has been used historically in terms of freedom of information, transparency, and anti-corruption (Parks 1957; Birkinshaw 1997; Parycek & Sachs 2010), but is gaining new meanings in the age of the internet.[31] Although different scholars and practitioners highlight different aspects of the term, there are three important common traits in open government definitions: (1) participation and collaboration, (2) open data, and (3) technology enabled.

2.1.1.1. Participation and Collaboration

Defining open government, Schuler (2010) highlights the aspects of *participation and collaboration* when arguing that "The goal is to change the nature of governance, particularly the relationship of 'ordinary citizens' to the government" (Schuler 2010:.93).[32]

2.1.1.2. Open Data

In their definition of open government, The United Nations (2010) highlight the aspect of *open data* as they write that "the notion has assumed greater definition through its association with government as a 'platform' or provider of data and services for others to exploit as they see fit." (United Nations 2010: 16). Open data serves as an independent subject matter, but is here seen as only one of

---

[30] Making matters more complicated, the web is a moving target evolving quickly (Fogg 2003). A definition from 2006 might not accurately describe open government today.

[31] A 'historical' definition of open government is: "the notion that the people have the right to access the documents and proceedings of government." (Lathrop and Ruma 2010: xix). Contemporary technology is giving new meanings and opportunities to the term.

[32] In the same way, O'Reilly proposes the following open government definition: "the use of technology-especially the collaborative technologies at the heart of Web 2.0-to better solve collective problems at a city, state, national, and international level" (O'Reilly 2010:.12).

three parts in open government. It is here worth noting that the provision of data is a means, not the end (Allison 2010; Meijer & Thaens 2010; Bass & Moulton 2010; Swartz 2010).

2.1.1.3. Technology

Finally, in its contemporary sense of the term, open government is enabled by *technology*. A historically important part of open government is freedom of information requests. These requests could be made without directly relying on technology. In 2013, it is hard to see any modern societies not relying on electronic databases in order to retrieve such information, as technology facilitates openness in terms of speed, scale and costs Moreover, through proactive disclosure of data on the internet, most freedom of information requests can be made obsolete (Coglianese 2009; Miller 2010; White House 2011). As Malamud (2010) concisely puts it, "Today, *public* means *online*" (Malamud 2010: 46). Some scholars offer technology-centric definitions of open government. For example, Eggers (2010) define the term as "A form of digital revolution that transforms government" (Eggers cited in Mergel 2010: 7).[33]

2.1.1.4. The Definition

The definition used here covers all three criteria described above and is written by Lathrop and Ruma (2010): Open government is

> *"government where citizens not only have access to information, documents, and proceedings, but also can become participants in a meaningful way. Open government also means improved communication and operations within the various branches and levels of government"*

> (Lathrop & Ruma 2010b: xix)[34]

---

[33] Similarly, Robinson et al. (2009) argue that technology is enabling open government by removing barriers.

[34] Harrison et al. (2011) also use this definition. Although a single citation does not seem like a lot, it is as canonical as it gets in an emerging field.

2.2. Egovernment and Open Government

Although open government is highly interdisciplinary, it can be placed within the egovernment field (Chun et al. 2010), which has been around for more than a decade. Open government broadens and extends the traditional egovernment field in three ways: purpose, functions, and scope. This is illustrated in Table 2.

	Traditional Egovernment	Open Government
Purpose	Deliver offline services online	Create transparency, participation, and collaboration
Functions	Digital presence Simple communication Information provision Transactions One-stop shopping	Participating Collaborating Transparency
Scope	Public sector information systems	Public sector information systems Non-governmental information systems in a governmental sphere

Table 2: Purpose, functions, and scope of egovernment and open government

2.2.1. Purpose

Egovernment can be defined as "The employment of the Internet and the world-wide-web for delivering government information and services to the citizens." (UN 2002 quoted in Yildiz 2007: 650). The purpose of egovernment is to deliver the information and services online in addition to or instead of offline. It is about digital accessibility (Chou et al. 2010; McDermott 2010; Harrison et al. 2011). *Open government*, is, on the other hand, aiming to facilitate participation, collaboration, and transparency (op cit.). All the services and information provided in the egovernment definition are still needed as open government often relies on this basic infrastructure.

2.2.2. Functions

The traditional egovernment stage models (e.g. Layne & Lee 2001; Andersen & Henriksen 2006) are primarily concerned with digital presence, simple communication, information provision, and transactions (ibid.), but not with participa-

tion and collaboration. For example, the final stage in Layne and Lee's (2001) stage model is horizontal integration, providing one stop shopping for the citizens. But in open government citizens will do more than shop; they will co-produce the open government information systems and content in them.

The Public Sector Process Rebuilding Using Information Systems (PPR) model proposed by Andersen and Henriksen (2006) comes closer to the aspiration of open government by giving the ownership of data to the citizens and having data mobility across platforms. However, open data is only the beginning in open government, not the end (Allison 2010; Meijer & Thaens 2010; Swartz 2010; Bass & Moulton 2010).

2.2.3. Scope

Traditional egovernment research is concerned with IT systems owned by governments (Heeks and Bailur 2007). For example, Brown and Brudney (2001) distinguish between three categories of egovernment: Government-to-Government (G2G), Government-to-Citizen (G2C), and Government-to-Business (G2B). Non-governmental actors are only included insofar they have contracts with a given government, delivering information systems.

The scope is not as clear-cut when it comes to open government, as openness makes the borders vague. In open government, non-governmental actors are included liberally (Pollitt 2012). This openness is obviously the point of open government, but it does make it harder to propose a clear scope of the subject.

Government agencies are still central actors but they are cooperating with - and occasionally challenged by - non-governmental organizations and citizens (Pollitt 2012). The interactions are increasingly complex; citizens can contribute content to government owned sites, public servants can reach out to stakeholders and citizens and comment on non-governmental sites, and citizens and public servants can meet at neutral ground via social media. In the most radical interpretation of open government the scope covers potentially the entire internet.

There are three scenarios where the openness in open government results in a broader scope than the one applied by traditional egovernment: (1) non-governmental content on governmental websites, (2) government information used on non-governmental websites, and (3) government employees contributing to non-governmental websites.

2.2.3.1. Non-governmental Content on Governmental Websites

In is increasingly common that the role of web property owner is separated from the content creator role. This is the core of web 2.0 in which regular users produce content on commercial websites they do not own. Citizens can post content such as comments on many government websites. This content is thus created by non-government actors, but placed on governmental websites. It is therefore necessary to understand non-governmental actors in order to understand participatory governmental websites.

2.2.3.2. Government Information Used on Non-governmental Websites

When governments provide data feeds for non-governmental actors to use freely, the traditional egovernment scope cannot grasp end products such as user interfaces. Understanding the end product is crucial as egovernment, and open government, needs to be evaluated from the citizens' point of view (Nam 2011).

The open government scope encompasses all websites and apps that interact with government.[35] This can include online forums,[36] e-rulemaking platforms,[37] information portals,[38] brainstorm collaboration[39] etc.. Whether the website or app is operated by a governmental or non-governmental organization is irrele-

---

[35] For example, a majority of comments regarding current rulemaking is scattered around the web, e.g. on blogs, newspapers, and social media, and thus not confined to government websites.

[36] E.g. http://k10.dk, http://digitaliser.dk, and http://govloop.com.

[37] E.g. http://regulationroom.org and http://regulations.gov.

[38] E.g. http://borger.dk and http://usa.gov.

[39] E.g. http://sammenomaarhus.dk and http://opengov.ideascale.com.

vant. Linders (2011) call this Citizen To Citizen Government. This is governance with little or no government involvement: "Citizens help themselves and one another. The government plays no active role in day-to-day activities but may provide a facilitating framework" (Linders 2011:169).

From this perspective, it is worth noting that the White House's Open Government Directive demands that the agencies' action plans "should include descriptions of and links to appropriate websites where the public can engage in existing participatory processes of … [the] agency" (The White House 2009: 9). But it does not specify if these websites are owned by the government or non-governmental actors. Moreover, non-governmental actors operate some of the US open government flagship initiatives. For example, Regulation Room[40] is used in US Department of Transportation's formal rulemaking process while it is being owned by Cornell University. Similarly, The U.S. Department of Energy's flagship project is Open Energy Info,[41] which is operated by Alliance for Sustainable Energy (White House 2011; White House 2013). Taking it even further, UK Prime Minister David Cameron aims to empower communities to govern themselves (Linders 2011). Thus, open government also includes non-governmental websites operating in the governmental sphere. Government (e.g. domain.gov) and non-government (e.g. domain.org) initiatives are different, but they are all open government when their subject matter is government (Noveck 2010).

2.2.3.3. Government Employees Contributing to Non-governmental Websites

It is not sufficient to limit interactions to own website properties. In both traditional egovernment and open government, there is a *build it and they will come expectation* (Høgenhaven & Andreasen 2011). This approach to the internet is problematic and undermining for successful egovernment and open government, and it "is no longer enough for the organization to focus on building a better

---

[40] http://regulationroom.org

[41] http://openei.org

website…" (Farina et al. 2011: 395). [42] Instead of seeing government websites as islands on which public agencies govern,

> *"Government needs to go where their citizens are. Since people typically use commercial entities rather than government websites to find and discuss agency related info (ex. health or safety issues), government agencies that reach out to these social spaces to post helpful information can expect to have more success in enticing people back to their agency website for more information and details"*

> (Mitchell 2010: 28)

2.3. Participation in Open Government

It is difficult to generally and objectively decide if a contribution is useful, as this is group and context dependent (Moon & Sproull 2008: 496). This being said, the literature on online communities often assumes that more contributions are better than few contributions. This is the basic assumption when there is no *a priori* knowledge about the usefulness of any given contribution. There are clear exceptions to this assumption: redundant contributions, trolling, and flaming are likely to have no or negative impact. More often than not it is possible to filter out or delete poor contributions with little effort. This can be done through automatic algorithms or collective filtering. This supports the general claim that a high contribution level is better than a low.

---

[42] Farina et al. (2011) further argue that government agencies need to not only be proactive, but also reactive to relevant commenting around the internet: "The strategy must provide for both proactive push and reactive response and, perhaps most important, it must be able to adapt to a broad range of events and circumstances that even the most foresighted planning will be unable to anticipate or control." (Farina et al. 2011: 396). A similar delimitation discussion can be found in the eParticipation literature. Sanford and Rose (2007) note that eParticipation can be initiated by non-governmental actors, although most of the literature is concerned with government initiated initiatives.

Solomon and Wash (2012) find that new users are affected by the existing content on websites. For example, if new users see opinionated content, they tend to write opinionated content themselves. New users, thus, are likely to reproduce what is already there. On one hand, this means that more participation will fall inside the community norms and values, thereby making the contributions valuable to the community. However, , the contributions will probably not add a lot of diversity. There is a declining value of additional participation, but it is still likely to be valuable.

In open government public participation is necessary and desirable. If no one is participating in open government, it fails. But how much participation is necessary? And does it always make sense to try to increase it? The answers to these questions largely depend on what one considers to be the aim of open government. To some open government proponents the aim is to improve democratic processes. To other proponents the goal is to improve the output of decision-making. In other words, some are concerned with process, others with outcome.

Noveck (2010) uses these two aims to distinguish between deliberative democracy and collaborative democracy. Deliberative democracy is, as suggested by Habermas (1994), concerned with process and procedure. Collaborative democracy is, on the other hand, concerned with outcome (Noveck 2010). Table 3 provides an overview of the differences between deliberative and collaborative democracy in terms of focus, aim, participation and concerns.

	Deliberative democracy	Collaborative democracy
Focus	Process oriented	Outcome oriented
Aim	Improve democracy	Improve outcome
Participation	As many as possible	As many as it takes to improve decision – too many create overload
Concerns	Creating digital divides	Getting bad inputs in processes

Table 3: Deliberative and collaborative democracy

2.3.1. Collaborative Democracy

Proponents of collaborative democracy are concerned with attaining the best possible output. Participation is considered to be positive insofar it supports this goal. But there can also be *too much* participation, thereby making collaboration harder and inflicting too many poor opinions on the decision-making process.[43]

The collaborative democracy perspective can sound elitist, as it is concerned with too much "unqualified" participation. And not all proponents of collaborative democracy agree that too much participation can be an issue. More participation is better when scalable filtering models are successfully applied (O'Reilly 2010). In idea brainstorms (e.g. sammenomaarhus.dk and http://opengov.ideascale.com), citizens can vote on other citizens' suggestions and thereby filter out the poor ones. According to crowdsourcing theory (Surowiecki 2005; Brabham 2008), more participation from users with diverse backgrounds is always better. The more information available, the better are the decisions made.

2.3.2. Deliberative Democracy

Whether or not collaborative democracy proponents see too much participation as a potential problem, they agree that participation is only a means, not the end. But to deliberative democracy proponents, participation is the end. This research has its roots in Habermas (1994), and has been examined in online settings since the early 1990ies (e.g. Dutton 1992). Through increased participation it is possible to improve the democracy, as citizens are increasingly involved. This is considered to be a value, regardless of how it affects the outcome of decisions.

The main concern from a deliberative democracy perspective is to promote equal participation and bridge digital divides. For if not everyone has the same

---

[43] This is especially a problem when the participatory platforms do not scale well. This is, for example, the case when people participate by sending emails or post cards (Farina et al. 2011).

possibilities to participate, the democratic legitimacy and ethics are compromised (Sanford & Rose 2007).

2.3.3. Participation as Dependent Variable

The work done in this dissertation is concerned with addressing the participation challenge in open government. At this point, the aim is to increase the quantity of contributions. The dependent variable used in this dissertation is *number of contributions* made to K10. This is a simple, quantitative variable that counts number of posts written in the defined experimental timeframe.[44] This dependent variable is used because it is directly related to the participation challenge. For better or worse, this quantitative variable is the standard in much online community research, and is used in many papers included in the literature review in Chapter 3 (e.g. Ludford et al. 2004; Rashid et al. 2006; Nov, Naaman & Ye; Yang and Lai 2010).

Following the logic of Anderson's (2006) long tail hypothesis, there is a niche demand for basically everything. This niche is potentially very small and often ignored, but exists nonetheless. This entails that any given post is likely to be useful to at least a few other users, and thus desirable for the community. Moreover, the more answers that are given to any question, the more likely it is that the initial question is answered correctly and thoroughly. This means the more posts the better.

As the preferred dependent variable is quantitative and not qualitative,. methodologies such as netnography (Kozinets 2002), ethnography (Wilson and Peterson 2002; Nardi 2010), and discourse analysis (de Laat 2002) are unsuitable for this research. Variables based on these methodologies are useful for analyzing the content of open government communities, but offer little insights into the quantity of posts.

---

[44] The variable, post count, comprises both new threads and posts. Please refer to section 3.4 for more details on the difference between these

The simple post count dependent variable has two severe limitations: *First*, the variable does not give any information about the contribution quality or the perceived usefulness of the contribution. There have been several attempts to quantitatively assess the contribution quality (see for example Weimer, Gurevych, and Mühlhäuser 2007; Blumenstock 2008; Shah and Pomerantz 2010). The suggested metrics include text length (Weimer, Gurevych, and Mühlhäuser 2007; Blumenstock 2008; Shah and Pomerantz 2010), question and exclamation mark frequency (Weimer, Gurevych, and Mühlhäuser 2007), capital word frequency (Weimer, Gurevych, and Mühlhäuser 2007), number of references and links in the article (Stvilia et al. 2005; Blumenstock 2008), spelling error frequency (Weimer, Gurevych, and Mühlhäuser 2007), readability indices (Blumenstock 2008), number of edits (Stvilia et al. 2005), HTML markup (Weimer, Gurevych, and Mühlhäuser 2007), user profile data (Wu et al. 2010), user ratings of contributions (Drenner et al. 2008), historical data (Wu et al. 2010), time between answers (Aschoff, Schaer, and Schwabe 2011), and number of links pointing to the article (Stvilia et al. 2005).

The problem with these metrics is that their robustness across different kinds of online communities have yet to be demonstrated and validated (Aschoff, Schaer, and Schwabe 2011). Although the research cited above is promising, the external validity of the metrics is unknown (Farina et al. 2010). Moreover, most of this research is evaluating English language. This makes it even more problematic to apply these variables to K10, a community that operates in Danish.

Secondly, number of posts can be seen as a means, not an end to open government. As described above, the open government proponents who are concerned with collaborative democracy focus on better policy outcomes. In this case, the more is not necessarily merrier. From this perspective, an increased number of contributions might actually be worse because it clutters the system and might inflict noise on good data. But with the participation challenge in mind this argument is rather hypothetical. As most open government communities fail, the number one priority is to get critical mass. This means participation quantity becomes an end in itself.

Despite these limitations, post count is used as the primary dependent variable in this dissertation.

2.4. The Promises and Critiques of Open Government

To some people (e.g. Salamon 1995) increased participation in government is a utopia (or a nightmare). And one ought to wonder if large-scale citizen participation in government is even feasible? The short answer to this question is yes. It is already happening today, and has been doing so for decades. A good example of this is 911 help lines.

Police, doctors, and fire departments would not be able to allocate their resources effectively without citizen inputs. Despite modern technologies, emergency services cannot know where they are needed without citizen input (Eaves 2010). The 911 help lines serve as knowledge aggregators based on a few data points: issue and location.

Many policy-makers and researchers tend to be enthusiastic about new technological opportunities (e.g. O'Reilly et al. 2010; White House 2011; Nam 2011). For example, Nam (2011) only highlights the benefits of open government. But it happens far too often that technological idolization and technophilia ends in disappointment. There is a tendency to overestimate the benefits while underestimating the disadvantages and unintended consequences (Bekkers, van Duivenboden and Thaens 2006; Cammaerts 2008; Meijer and Thaens 2010). This has happened many times in egovernment literature already, and as Nam (2011) points out, we cannot expect that to be any different when it comes to open government. Anttiroiko (2010) argues that: "Even if Web 2.0 provides a lot of opportunities, it may also pose risks that we should be aware of in order to utilize its potential in a responsible and sustainable ways." (Anttiroiko 2010: 19).

Furthermore, the costs and benefits are unclear: "the full cost to the public bodies for opening up their data is not clear, and the benefits are difficult to calcu-

late" (Janssen 2011: 454). Therefore, there is a need to examine arguments for and against open government (Anttiroiko 2010; Cammaerts 2008; Meijer and Thaens 2010). Thus far in this dissertation, it has been assumed that open government is desirable. But this assumption deserves to be explored. The advantages and disadvantages of open government described here are at a general level. If a government agency implements specific open government measures, it is likely to encounter only some of them.[45] The advantages and disadvantages highlighted in the open government literature are summed up in Table 4.

---

[45] As Meijer and Thaens (2010) point out, open government policies need to be translated into specific strategies, aligning technology with specific agency goals. With the risk of being overly abstract, it is fruitful to assess general advantages and disadvantages to understand the phenomenon.

	Advantages	Disadvantages
Participation	More democratic processes in public agencies Better decisions and policy outcomes Civic education Increased legitimacy of decisions	Makes the digital divide deeper Attain and maintain participation Amateurism Unsteady supply Lack of accountability Increased politicization Internet culture
Efficiency	Reduced costs Crowdsourcing From retail to wholesale	Moving costs from problem-solving to management Scalability issues Opportunity costs Extra work
Privatization	Cheaper to produce data than products	Commodification Alienation
Transparency	Increased trust in government Reduced corruption	Might not imply action Citizens can misunderstand data Contradicts a need for secrecy
Government Culture	Cheaper to fail online than offline Makes successes more clear	Risk aversion Need for secrecy Can create perverse effects
Security and privacy	Allows citizens to see who knows what about him	Need for common identifiers to facilitate open government Data linkage: The sum of data tells more than the parts. Unforeseen data linkage can give too much information to enemies. The sum of data tells more than the parts.
Innovation	Better innovation in government More innovation outside government	Need to get non-governmental actors interested in participating

Table 4: Summary of open government's advantages and disadvantages

To structure the arguments for and against open government, seven different aspects are examined: participation, efficiency, privatization, transparency, government culture, security and privacy, and innovation.

2.4.1. Improved Democracy: Participation as Process

Participation can improve democratic processes through increased citizen involvement (Habermas 1994; Held 1997; Harrison et al. 2011). Opening up government in terms of inputs and processes increases citizens' opportunity for participation in politics as well as administration (White House 2011). Existing open

government communities show that it is possible to engage new stakeholders into the democratic process (Farina et al. 2011).

2.4.1.1. Pro: Civic Education

Nam (2011) argues that participation not only benefits the democratic process, but also helps educating the participating citizens.[46] Citizens should not be viewed as 'static actors' in open government processes, but as actors that can improve and benefit directly and indirectly through participation (Farina et al. 2011; Lampe et al. 2011; Farnham et al. 2012).[47]

2.4.1.2. Pro: Increased Legitimacy of Decisions

When citizens feel they have been involved in a decision-making process, they are more likely to accept the political processes and decisions as legitimate (Sanford and Rose 2007). By creating more participatory channels and opportunity, open government can help accomplish this.

2.4.1.3. Con: Motivation to Participate

As described in Section 2.3, the right to participate does not necessarily equal willingness to participate (op cit.). Most open government communities that

---

[46] Nam (2011) argues "civic education through participation in citizen-sourcing projects can deepen and enrich citizens' substantive knowledge of issues, broaden their understanding of key actors and the government's role, and hone their civic skills in using governance tool" (Nam 2011: 13).

[47] Farnham et al. (2012) find a significant correlation between young peoples' civic learning and the use of an open government community called Puget Sound Off. Based on the participation in Regulation Room, Farina et al. (2011) find that citizens who use Regulation Room gain a greater understanding of (1) rulemaking processes, (2) other citizens' perspectives, and (3) the work of relevant government agency. The scholars conclude "Regulation Room experience gives cause for optimism that Rulemaking 2.0 participants can gain new knowledge from their experience, and, furthermore, that some of these gains can result from "just" reading" (Farina et al. 2011: 459). Lampe et al. (2011) echo this finding as civic participation in the AdvanceMichigan project increases citizens' knowledge, empathy with other citizens, and sense of own political interests.

have been evaluated thus far show that the democratic gains have been moder-ate at best due to insufficient participation (Parycek & Sachs 2010).[48]

2.4.1.4. Con: Slacktivism

Online participation is often easier than offline. But easy participation often en-tails less impact (Christensen 2011).[49] Many of the online participation forms have very low, or zero, political or social impact. Such low-cost, low-effect partic-ipation is called slacktivism, referring to "activities that are easily performed, but they are considered more effective in making the participants feel good about themselves than to achieve the stated political goals" (Christensen 2011: 5).

2.4.1.5. Con: Increased Politicization

When including citizens in the process, it is easier for policy makers to gauge public reactions to new legislation early in the process (Farina et al. 2011). How-ever, this might lead to increased politicization and thus make the executive branch less expert driven and more prone to short-sighted populism.

2.4.1.6. Con: Digital Divides

Due to digital divides, not all citizens have the same prerequisites to contribute (Norris 2001; Jaeger and Thompson 2003; Yu 2006).[50] The implication of the

---

[48] Parycek and Sachs (2010) explain "The number of participants in existing eParticipation projects is too small to deduct democratic legitimization" (Parycek and Sachs 2010: 3). For example, the 2009 US Open Government Dialogue created 4,205 ideas, but less than 1 % of the US citizenry participated (Parycek & Sachs 2010). Similarly, in the Citizen's Briefing Book, 44,000 ideas were submitted by 125,000 citizens, and 1.4 million votes were casted (Nam 2011).

[49] This is especially the case if public agencies collect feedback from citizens without taking it into account in the decision-making process. For example, in a 2009 Ideascale initiative, President Obama asked US citizens to submit what they believed were the most important challenges to the United States. Marijuana users and lobbyists mobilized a significant crowd and got the issue voted in as number one (Hansell 2009). The idea has yet to be translated into policy.

[50] Although the internet is increasingly adapted "systematic differences in technology access and proficiency by age, gender, race, and economic status surely still exist." (Farina 2011: 433f). Digital divides are concerned with social-economic factors, not personality traits (Cullen and Morse 2011).

digital divides are increasingly problematic when empowering citizens further through open government. The increased online participation will give participants influence, likely on the expense of offline participation. People on the "wrong side" of the digital divide are likely to be even more neglected, as they are unable or unwilling to take part in the online political discourse. [51]

Cammaerts (2008) and Pollitt (2012) argue that many of the current power structures are reproduced and reinforced online. This is supported by a PEW (2009) study that finds that civic engagement, whether or not it takes place online or offline, continues to be an activity reserved for high income and highly educated citizens (Smith et al. 2009). Thus, there is a significant risk that open government will benefit the people on the right side of the digital divide, while harming the others (Jaeger & Bertot 2010).

This creates a new aspect to the have/have-not dimension (Jaeger & Bertot 2010; Schuler 2010). Farina et al. (2011) note that such differences exist across all channels of public participation. Not everyone has the same abilities to participate in open government, but the same problem applies to the vast majority of democratic processes. Open government should not be treated differently than other of these processes.

2.4.2. Improved Decisions: Participation as Outcome

Although this benefit might come at the expense of the improved democracy benefit described above, increased participation can lead to better decisions (Coglianese 2009; Armstrong 2010; McDermott 2010; Lathrop & Ruma 2010; Janssen 2011). Netherland & McCroskey (2010) highlight outcome efficiency rather than democracy as the main benefit of increased involvement: "Participation means more than just voting. Participation means applying your unique set of talents to improve the government that works for you" (Netherland and McCroskey 2010: 177).

---

[51] Yu (2006) argues "it has become one of the most acute developmental problems both in a global context and within individual countries" (Yu 2006: 237).

Decisions and rulemaking can be improved by including two kinds of knowledge: expert and situated knowledge (Farina et al. 2011). Expert knowledge is knowledge coming from domain experts who often work outside of government. Situated knowledge[52] comes from people on the ground. Rather than expert credentials, these people have experiential credentials that qualify their contributions.[53]

2.4.2.1. Pro: Expert Knowledge

Participation can help improve decisions in different ways. Public servants are often experts but not oracles that know everything about a subject matter. Harrison et al. (2011) ask "Who declared that government had to have all the answers? Citizens are smart, too" (Harrison et al. 2011: 126). Experts exist inside and outside government agencies (Harrison et al. 2011). A significant benefit is to get inputs from experts from related fields.

2.4.2.2. Pro: Situated Knowledge

Including ideas from the diverse citizenry can lead to better decisions than those based solely on ideas from bureaucratic experts and consultants (Raymond 1999; Surowiecki 2005; Brabham 2008; Chun et al. 2010; Noveck 2010; Farina et al. 2011). Harrison et al. (2011) explain that "citizens may know as well as bureaucrats, or perhaps even better because they deal with such problems more frequently, what options constitute desirable policy" (Harrison et al. 2011: 4).

The argument is not necessarily to let everyone have a say, but rather to use the knowledge of the people who want to have a say (Eaves 2010). As argued by Eaves (2010), few people are concerned with public policy at large, but the ma-

---

[52] Sometimes referred to as local knowledge and first-hand experience.

[53] Citizens with situated knowledge cannot provide the same kind of factual knowledge, such as research and citations, as experts can. But they can contribute with something different. The knowledge they do have is more of anecdotal character and revolves around personal stories and experiences. This knowledge differs from expert knowledge, but is not necessarily inferior.

jority of people are concerned with some - or just one - policy area. This is the long tail of public policy.[54]

2.4.2.3. Con: Unsteady Supply

Salomon (1987) argues that voluntary work causes unsteady supply. As there is no direct hierarchical control with the volunteers the productivity and availability is based on probability: "no one is 'scheduled' to be available, but someone will 'probably' be there to help" (Linders 2011: 172). The UN Department of Economic and Social Affairs (2010) raises the same concern.[55] It is still an open question if third-party suppliers will be able to address all the citizens' demand.

Relatedly, amateurs taking work away from professionals can be problematic as the overall standards are lowered due to insufficient training and organization.[56] Keen (2007) calls this the cult of the amateurs.

2.4.2.4. Con: Accountability

Using voluntary non-governmental service providers creates issues of accountability. When governments start to supply raw data instead of actual products such as websites or apps, they have to rely on third parties to do the job for them. But citizens cannot hold the third parties accountable in the same way as they can with public agencies. Brudney and England (1983) have previously argued against co-production on accountability grounds and called it "the ultimate excuse for bad government" (Brudney & England 1983: 59).

One mitigating circumstance to this critique is that government agencies can purchase the products developed by third parties once these have proven to be successful (Burton 2010). This helps government agencies save money on man-

---

[54] Due to the reduction of transaction costs, it is possible for networks around a single area to mobilize (Eaves 2010). There are *micro niche* experts who can make contributions to government agencies.

[55] The UN asks: "If governments free the data, will third parties come?" (UN 2010: 17).

[56] With article quality comparable to Britannica, Wikipedia serves as an example of successful broad participation (Giles 2005).

aging developing contracts with IT-development companies and developing products that do not work.[57]

2.4.2.5. Con: Internet Culture

Internet users are used to and encouraged to quickly vote on and rate everything on the web. This form of quick light-weight participation is now what most people want: "Drive-by participation is all that some users want, and they expect to be able to do so immediately, with minimal thought or effort" (Farina et al. 2011: 445). But this social media mindset does not translate well into complex policy issues where attention is needed to understand and comment on complex issues (Farina et al. 2011). There are several issues with light-weight participation such as voting: *First*, people have a bias toward popular items (Berns et al. 2010).[58] *Secondly*, one of the benefits of participation is civic education. But quickly voting up a comment is not very educational compared to writing a response. *Thirdly*, popularity rankings might not benefit the decision-making process.

2.4.3. Efficiency

Global recession and smaller generations in most Western countries causes financial strains and a need for increased efficiency. By opening up government and providing data free to all, more hands will be able to help operate and develop the public sector (Chun et al. 2010; Lakhani et al. 2010; Parycek and Sachs 2010; Chun et al. 2010). There are several ways efficiency gains can be realized.

2.4.3.1. Pro: From Retail to Wholesale

When non-governmental actors can participate in the development of digital services, the public sector can in some cases provide data rather than services to citizens. Specifically, this can reduce costs by letting non-governmental actors

---

[57] Burton (2010) argues that this is being successfully done in the BRIDGE (http://about.bridge-ic.net) program, although it no longer exists.

[58] This means that a comment with existing positive ratings will get disproportionate positive ratings, simply because it is already popular.

accept the risk and make the investments in software and web development (Parycek & Sachs 2010).[59]

2.4.3.2. Pro: Crowdsourcing

When jobs can be broken down into standardized and specific tasks, they can be handled by many different volunteers. This is often possible as the web 2.0 technologies help reduce transaction costs and thus leverage the assistance of a large number volunteers (Parycek & Sachs 2010; Eaves 2010).[60] Thus, when citizens are used in the right places and tasks are broken down appropriately, crowdsourcing can be an efficient way of organizing government.

2.4.3.3. Con: Moving Costs from Problem-Solving to Management

Although volunteers are working for free, it does take resources to recruit, train, and manage them. There is a risk of not actually reducing costs, but just moving them from problem-solving to managerial activities (Brudney & Duncombe 1992). This critique is certainly valid in a historical perspective within voluntarism. However, web 2.0 technologies vastly reduce, or potentially eliminate, transaction costs, thereby altering the calculations in favor of voluntarism (Eaves 2010).

2.4.3.4. Con: Opportunity Costs

Open government communities are associated with opportunity costs. Focusing on open government and participation might entail that attention and investments shifts from a more important area. When looking at traditional egovern-

---

[59] One could argue that the total sum of costs is constant, and that the main change is that they are shifted from the government to individual citizens.

[60] A good example of this is Peer to Patent (http://peertopatent.org), which is a joint project between the U.S. Patent and Trademark Office and the New York Law School. Parycek and Sachs (2010) explain Peer to Patent this way: "Pre-processing the patent applications by incorporating a great number of volunteers … helps accelerate the work process. Processing takes only half as long and the implementation of the peer- review concept assures quality. This has increased efficiency and the competitive position of the USA…" (Parycek and Sachs 2010: 4).

ment services in Western countries today, it is clear that there is a tremendous amount of challenges. The self-service solutions are often hard to find on the web (Høgenhaven og Andreasen 2010; Mitchell 2010) and hard to use (Andersen 2011). Consequently, the adaptation is scarce (Andersen 2011). It might be more efficient to focus on these solutions than on open government.[61]

2.4.4. Privatization

As described above, open government makes the government a data warehouse rather than a retail store (op cit.). When the government provides data to let third parties deliver the end services, a *de facto* privatization and commodification occurs (UN DESA 2010: 17).[62] Some open government proponents see this as an advantage as the market is considered to be more effective than government (e.g. O'Reilly 2010). Critics argue that co-production is a gimmick used to cut government expenses and programs without being a viable alternative to current government programs and services (Linders 2011).

2.4.4.1. Con: Alienation

Through open government, the public sector becomes more of a data provider and less of a service provider. This means that the public sector will be less visible to the citizens, as they no longer act as the last link of service provision (Provan & Milward 2001; Anttiroiko 2010). This entails a risk of alienation as the citizens will lose direct contact to the public sector (Pollitt 2012). The public sector might end up invisible to the citizens,[63] except when something does not work (Harrison et al. 2011). In short, it might be harder to demonstrate the value created by the public sector and thereby making it harder to legitimize taxation.

---

[61] Some scholars consider open government to be an elaborate stage of traditional egovernment (op cit.). Assuming such stage models are correct, one could argue that we at the present time not are ready to launch the open government stage.

[62] Intended or not, this model of service provision "has an inevitable resemblance to economic models that privilege market efficiency with the least government intervention." (UN DESA 2010: 17).

[63] Harrison et al. (2011) argue that "the reliance of governments on third party actors has generated the image of the hollow state to describe governments that become distanced from the services they deliver to citizens" (Harrison et al. 2011: 5).

2.4.5. Transparency

Transparency is an important aspect of democracy because it potentially helps citizens make informed choices, increases trust in government and reduces corruption (Hood 2007).[64] In the open government context, transparency is considered to be an aim to a better and fairer government, rather than a goal in itself (White House 2011). Today, transparency[65] almost equals online availability (Miller 2010). Online availability is important because the citizenry's expectations and perceptions have changed.

2.4.5.1. Pro: Trust and Reduced Corruption

Transparency creates trust in government (Bertot et al. 2010) and reduces corruption (Coglianese 2009).[66] Via open data, citizens get access to comprehensive, real-time data about public spending, contracts, and lobbyists' access to government (Chun et al. 2010; Lakhani et al. 2010). By making government data accessible in machine-readable formats, it is possible for citizens, organizations, and companies with decent tech skills to explore, mash-up, and visualize the data. This can vastly increase transparency, and thereby help facilitate accountability.

Furthermore, the collaborative elements of open government reduce the distance between citizens and public servants. This makes it easier for citizens to ask questions and for the public servants to communicate the at scale.

---

[64] Although transparency is usually celebrated, Schumann (2007) is concerned about the nearly universal belief in transparency and asks: "Who could question the value of transparency in politics, after all?" (Schumann 2007: 838). Hood (2009) considers transparency to be a 'banal idea', meaning it is so commonly accepted that it goes unexamined and no reflections about the desirability is made.

[65] Harrison et al. (2011) defines transparency as "the extent to which government actors make available the data and documents the public needs in order to assess government action and exercise voice in decision making" (Harrison et al. 2011: 3).

[66] The more information there is available about public administration the harder it is to engage in corrupt ventures (Christopher Hood 2007; White House 2011). As Bentham argues, "the more closely we are watched, the better we behave" (Bentham quoted in Hood 2009).

2.4.5.2. Con: The Scope of Transparency

According to Fung and Weil (2010), one of the most critical limitations of the current transparency agenda is the focus on public sector transparency vis-à-vis private sector transparency. The lack of transparency is more problematic in the private sector than in the public sector since private companies are playing increasingly important roles in the citizens' life. In short, creating transparency through open government is good but unambitious. Rather the goal should be to pursue open society, which includes creating openness in the private sector.

2.4.5.3. Con: Effect of Transparency

An important assumption about open data is that it increases transparency. But an international study could not find any significant correlation between countries' electronic access to government and transparency ranking (Kimball 2011). Furthermore, transparency does *not* equal action, change or improvements (Swartz 2010). Transparency "simply shifts the work from the government to the average citizen, who has neither the time nor the ability to investigate these questions in any detail, let alone do anything about it" (Swartz 2010: 269). Transparency does not solve problems, it merely expose them.[67]

2.4.5.4. Con: Need for Secrecy

Secrecy is a big part of the culture in government and public administration (Lakhani et al. 2010). Although transparency is generally desirable, some secrecy is necessary to make government run (Coglianese 2009; Hood 2007; White House 2011). Transparency in the wrong places might hurt internal discussion and dissent, and reduce inputs from non-governmental actors[68] (Coglianese 2009; White House 2011).

---

[67] As noted by Miller (2010), "Making information available was the first big step: being able to connect the dots is the next crucial one" (Miller 2010: 199).

[68] E.g. due to disclosure of sensitive information such as trade secrets.

Increased transparency makes it easier to constantly second-guess the decisions made by public agencies and mobilize resistance. That can benefit small groups of society, but tradeoffs in society do have to be made. Too much transparency can lead to paralysis.

2.4.5.5. Con: Misunderstanding Data

The data that is being made transparent can be misunderstood by the citizens not trained to read it (Lakhani et al. 2010; Parycek & Sachs 2010; Janssen 2011). According to the US Bureau of Labor Statistics this is a real concern: "Without people understanding some of the context of that data, there is potential for problems that might undermine confidence in our data" (Kryger cited in Lakhani et al. 2010: 7).

2.4.6. Government Culture

Open government is about more than technology. It is about a new mindset and culture that reduces barriers between public servants and non-governmental actors. The increased transparency and extroversion is challenging the Weberian tradition of government[69] and bureaucracy dominated by hierarchical decision-making, insiders and outsiders delineation, and risk aversion (Eaves 2010).

Making policies about open government might be relatively easy, but changing government culture is not (Parycek & Sachs 2010). For open government communities to be truly effective, public servants need to adapt to the transparency and citizen inputs. However, this requires a second-level culture change, entailing a transformation of core values and basic elements of operation (Hood 2007). Such transformations are hard to achieve. Failure to change government culture can yield two potential outcomes: open government communities have little effect, or open government communities create perverse reactions such as blame avoidance behavior (Hood 2007).

---

[69] The Weberian notion of an ideal bureaucracy advocates rational, legitimate, and efficient government (Weber 1978).

2.4.6.1. Con: Risk Aversion

Many public servants are quite concerned with avoiding blame and critique (Hood 2007; Lakhani et al. 2010; Kimball 2011). This behavior is partly caused by a negativity bias, meaning that negative information generates more activity than positive information (Hood 2007)[70]. Fung and Weil (2010) echo this bias and argue that the current discourse "produces policies and platforms that are particularly sensitive to government's mistakes but often are blind to its accomplishments" (Fung & Weil 2010:107). Increased transparency helps create narratives of failures, because the press prefers to report failure rather than success. [71]

One solution to this problem is to ask users explicitly to provide negative *and* positive feedback (Fung & Weil 2010). This might sound difficult to attain, as unhappy users are more likely to respond to surveys than happy ones (Crook et al. 2009). But it is nonetheless possible, and it is already happening on many ecommerce and travel websites.[72]

2.4.7. Security and Privacy

As noted above, transparency has many advantages. But improving transparency will inevitably bring out concerns about privacy and surveillance. When it comes to privacy, there are basically four areas of concerns: data collection, data publication, data linkage, and data storage. So far, data collection has remained constant and relatively unaffected by open government communities. Vast amounts

---

[70] This pervasive bias goes "all the way down the food chain of executive government" (Hood 2009: 201).

[71] An example is the Danish tax authorities' use of Twitter. The tax authorities are often praised by bloggers for their use of Twitter to communicate complex legislation and deadlines. However, when they follow users on Twitter, the users get an email saying "You are now being followed by the Danish tax authorities". Thus, their mere use of Twitter was turned into an undocumented story about surveillance on Twitter (http://www.comon.dk/art/208612/quot-skattefar-is-following-you-on-twitter-quot).

[72] If websites such as Expedia, Amazon, and IMDB only collect and publish negative feedback about hotels and products, they will not give a comprehensive portrayal of the product.

of data were collected electronically before open government, as it is also the case today.

2.4.7.1. Con: Data Collection

In some open government initiatives, citizens' private social media accounts are linked to governmental websites. This data source can potentially be used for policy-making, law-enforcement, decision-making (Citron 2010), and surveillance.

2.4.7.2. Con: Data Publication

As more data sets are made public, there is an increased risk of eroding the privacy of citizens, companies, organizations, and employees. The more data that is made publicly available, the higher is the risk of publishing some data that ought not to have been published. There is also a risk that data identifiers are not sufficiently removed from the data sets before publication.

2.4.7.3. Con: Data Linkage

An important aspect of open government is the use of common databases, structured data, and APIs. Different systems need to be able to communicate to create value for the citizens. Thus, each object in the database needs to have a unique identifier (Bass & Moulton 2010). [73] By linking various systems with unique identifiers, different data about citizens will be combined (Anttiroiko 2010; Chou et al. 2010). Thus, the problem is not so much the storage of individual data points, but the linkage of data that can potentially enable severe surveillance. [74] Data linkages makes it possible to create valuable open government services

---

[73] For example, if the citizens can report potholes in the roads, it is necessary to be able to identify each unique road, so the right reports are linked to the right potholes. The same is true when it comes to other entities such as citizens and companies.

[74] Chun (2010) explains the surveillance problem this way: "Government organizations collect and maintain different kinds of data concerning the same entity (citizen) … In order to share and have an integrated and unified view of an entity, data from different sources need to be reconciled to be integrated." (Chun et al. 2010: 7).

such as health and travel alerts, but it also makes it possible to engage in severe surveillance and ID theft. Parycek and Sachs (2010) note that unique database identifiers can also help protect the individuals: "Having a unique official ID in the virtual reality empowers citizens to monitor and eventually edit personal data" (Parycek & Sachs 2010: 9). This has to be seen as a contrast to today where no one really knows what data different government agencies (and private companies) have collected.

Data linkages also cause concerns in terms of national security. Although individual datasets are harmless, Lakhani et al. (2010) note that there is "the opportunity that seemingly harmless data might be combined with other seemingly harmless data to create harmful data" (Lakhani et al. 2010:.13). Thus, one serious disadvantage is that the data can be used for good as well as harmful purposes (Janssen 2011).

There are currently over 100,000 datasets available at data.gov. The datasets are all unclassified and deemed harmless. But the declassification is based on an individual assessment and unforeseen combinations of the datasets might generate knowledge that would otherwise be classified. The problem lies in unforeseen synergistic effects. Harmless data plus harmless data might equal harmful data.

2.4.7.4. Con: Data Storage

The open government movement is closely associated with cloud technology (Wyld 2009). Data is often stored in big server parks operated by private companies such as Google, Amazon, Apple or Microsoft. Placing data on these servers – rather than Governments' own servers – might cause privacy concerns, as nongovernmental actors suddenly get access to sensitive information (Wyld 2009; Bertot et al. 2011).

Using cloud hosting might be more secure than hosting on own servers, as the cloud hosting companies can spend more resources on security and privacy due

to scalability (Wyld 2009). Furthermore, some government agencies already rely on third-party hosting and data storage (Foley 2012; FedRAMP 2013).

2.4.8. Innovation

Innovation in an open government context refers to two things: (1) innovation in government, and (2) innovation outside government with the means of government data.

2.4.8.1. Pro: Innovation in Government

Bill Joy, founder of Sun Microsystems, famously said: "No matter who you are, most of the smartest people work for someone else" (Joy quoted in Lakhani & Panetta 2007: 97). By making it possible and easy to volunteer, the public sector gains access to talented people who work for someone else without having to pay for their services (Jeppesen & Frederiksen 2006; Nam 2011). This means that entrepreneurs at own will can develop a product relying on Government data.[75]

One reason often suggested for the failure of government IT-projects is the development process. Complete system specifications are written down in a comprehensive contract and awarded to an outside company that has to deliver the code. The problem with this approach is insufficient user feedback (O'Reilly 2010; Burton 2010). Most successful IT innovations happen through agile developing processes and quick iterations (Ries 2011). By letting non-governmental agencies do some of the innovation without any contracts, the likelihood of success will increase simply because there is room for repeated trial and error itera-

---

[75] In the first years of egovernment, public agencies held a monopoly on egovernment services. The subsequent innovation has occurred through closed, governmental-led innovation processes relying heavily on contractors and single-iteration development (O'Reilly 2010; Burton 2010). In most countries, the results have been poor (Heeks 2003; Burton 2010) entailing that the majority of services are scarcely adapted by the citizenry. Open government has the potential to open up many of the innovation processes.

tions by different developers (Burton 2010). This has the potential to create a collective learning process (Robinson et al. 2010).

2.4.8.2. Pro: Innovation outside Government

The link between open data and innovations is explained by (Malamoud 2010): '"Government information - patents, corporate filings, agriculture research, maps, weather, medical research - is the raw material of innovation, creating a wealth of business opportunities." (Malamoud 2010: 44). Thus, it is not necessarily altruism or public service ideals that drive the innovation.[76] There are different motives for different kinds of third part developers: nonprofit public service, volunteer enthusiasm, political advocacy, and business objectives (Robinson et al. 2010).[77]

Both European and North American countries are selling data back to citizens (Tauberer 2010) through exclusive contracts with companies or through their own systems. Some data is sold directly on governmental websites. Some data is sold through private companies. The problem with selling data sets is that it hinders competition, and thereby innovation (Parycek & Sachs 2010). Open data is a hindrance to monopolies as everyone is free to use the data, thereby essentially eliminating the barrier of entry (O'Reilly 2010; Robinson et al. 2010; Janssen 2011).

To Janssen (2011), the outside-government innovation is the most important reason for opening up government. One calculation estimates the commercial value of open government data in the EU to be € 27 billion (Janssen 2011). Add to this the value to non-commercial actors.

---

[76] It is not only non-governmental actors that can participate in open innovation. An entrepreneurial civil servant got frustrated with bureaucratic processes and developed a social network for government called GovLoop (http://govloop.com) (Drapeau 2010).

[77] Facebook serves as an example of the success of open innovation. Most of the popular Facebook games are developed within the Facebook ecosystem by third-party companies (Jung 2011).

2.5. Open Government Value: To Whom?

As noted by Moore (1997), government ought to create value for citizens and companies. This applies offline as well as online government (Hui and Hayllar 2010). It is therefore relevant to discuss to whom and how open government adds value. In the 2003 EU PSI initiative, the aim is to create value to the information industry (Janssen 2011). But the examination of the arguments for and against open government shows that value potentially is created for numerous actors: citizens, government, and private companies. The following sections describe how open governments add value, if implemented correctly.

2.5.1. Citizens

Open government adds value to citizens in various ways. *First*, citizens get increased influence in government. Depending on the perspective, this either helps citizens get a more democratic deliberative process, or it helps citizens use their expert or experiential knowledge to create better policies and decisions. *Secondly*, increased participations can increase civic competencies. *Thirdly*, open government can increase efficiency.

2.5.2. Government

Government agencies can also benefit from open government. *First*, they get access to more expert and experiential knowledge, which can help them make better decisions and policies. *Secondly*, they can get help with tedious work through crowdsourcing. *Thirdly*, government agencies can use outreach and user feedback to communicate positive stories and successes, and thus mitigate the existing negativity bias. *Fourthly*, the increased transparency created by open government increases accountability and might help mitigate corruption.

2.5.3. Private Companies

Open government creates values to companies in two ways. *First*, private companies are obvious winners of increased privatization. When government goes from retail to wholesale a new market is created for private companies. Although some companies, currently holding monopolistic contracts selling government

data, will lose on open government, the private sector at large will win on data liberation. *Secondly*, a more efficient government will help companies get better return on tax money.

2.6. Open Government: Ideal or Reality?

The arguments for and against open government clearly demonstrate that there are advantages and disadvantages to open governments. Democratic and economic progress can be made if open government is implemented correctly. However, it is not a guaranteed effect of open government. If implemented and designed the wrong way open government can have little – or even perverse – impact. Done right, it is possible to attain many of the advantages while keeping most of the disadvantages in check.

2.6.1. Is Open Government a Choice?

An interesting question to consider is whether open government is a choice at all this point. As technology and the internet are becoming ubiquitous, it becomes increasingly harder to ignore it. And once engaging with internet technologies, it is necessary to adapt to the premises of the internet (Høgenhaven & Andreasen 2011). In the words of (Sifry 2010):

> *"From Wikipedia to Craigslist to Amazon to Google, the Web keeps rewarding those actors who empower ordinary users, eliminate wasteful middlemen, share information openly, and shift power from the center to the edges. Applying those same principles to government will undoubtly be messy, but ... it is where technology is already taking us."*
>
> (Sifry 2010: 121)

The main question is whether the impact will be modest or radical (Anttiroiko 2010; Parycek & Sachs 2010; Sifry 2010). Parycek and Sachs (2010) argue "Governments must accept that the information flow within a society is changing and they should pro-actively adapt to the new situation." (Parycek & Sachs 2010: 1).

Moreover, open government becomes increasingly necessary over time as, "the principles of open government have already been internalized by the young generation" (Parycek & Sachs 2010: 7).

An increasing number of open government communities have been implemented in many countries over the last couple of years. According to the White House (2011), "federal agencies have done much to make information about how government works more accessible to the public and, beyond that, to solicit citizens' participation in government decision-making" (White House 2011: 4).

This is not to say, that it is the state of art today (Eaves 2010; Meijer & Thaens 2010). There is still a large gap between the ideals of open government technologies and the current realities of government (Meijer & Thaens 2010). There is a significant part of normative thinking in open government. The implementation of open government has been experimental and sporadic rather than consistent and exhaustive. Thus, open government has already had some effect, although it is far from fully implemented.

As described in Chapter 1, it is easier to find open government failures than successes. The biggest problem is gathering and maintaining sufficient participation to leverage the benefits of open government communities. In order to do this, more knowledge about why users participate to open government communities is needed. This is the **gap** that this dissertation seeks to address.

2.7. Chapter Summary

With over 55 countries currently implementing open government policies, it is an important and relevant field of study. Aiming to contribute to open government in research and practice, this chapter seeks to ground the field. In this chapter, open government is described and analyzed on many different parameters. In this chapter, open government is defined as seen as an evolvement of egovernment. Open government is concerned with participation and collaboration, while traditional egovernment is concerned with moving offline govern-

ment services online. The main part of the chapter examines the many claims for and against open government. The parameters analyzed are democracy, decision-making, efficiency, privatization, transparency, government culture, privacy, innovation, and security.

It is probably difficult to ignore the open government trend over the coming years and if implemented successfully, open government can create value in a variety of ways. Open government adds *value to citizens* (1) through increased influence in government, (2) increased civic competencies as a result of participation, and (3) more efficient government and value for taxes. Open government adds *value to companies* by (1) increased privatization, and (2) better return of tax money due to increased efficiency. Finally, Open government adds *value to government* by (1) gaining access to more expert and experiential knowledge, (2) get help with tedious work through crowdsourcing, (3) communicate positive narratives of success, and (4) increased transparency and accountability to fight corruption.

These value propositions rely on successful open government implementation. A necessary condition for this is participation. At this point, many open government communities fail as a result of under-participation. Therefore, more knowledge about why users participate in open government communities is needed. This dissertation addresses this gap.

Chapter 3: Case Description

3.1. Introduction

As illustrated in Figure 3 (in Chapter 1), the aim of this dissertation is to help solve the participation problem in open government communities. This task is undertaken by leveraging social science theories to generate hypotheses and design ideas. In order to examine the quality and effect sizes of these hypotheses, they are tested in an existing open government community called K10.

It is in this open government community the four experiments are conducted. K10 is called the case in lack of better word, for this is not a traditional case study trying to explain this particular case in-depth (Yin 2002). Rather, the aim is to examine if constructs of four different social psychology theories can be used to increase participation in an open government community. As described in this chapter, K10 is considered to be such a community and thus a suitable environment to conduct the experiments in. The aim is not to create knowledge that can be generalized to all open government communities. It is too early to tell if different open government communities are sufficiently alike for such extrapolation. This entails that the results conducted at K10 should not be empirically generalized to all open government communities, to the entire population in Denmark, or to all early retirement pensioners and flex job recipients. The aim is rather to create knowledge that allows analytical generalizations (Yin 2002) to the open government research community.

The purpose of this chapter is to thoroughly describe the open government community used as an experimental laboratory in this dissertation, and thereby document the environment in which the data is collected. Such a description is necessary in order to assess internal and especially external validity of the research conducted in the remainder of this dissertation. However, in order to select an appropriate open government community, the ideal type for open government communities is defined in the beginning of the chapter. This archetype might be useful in other contexts and thus serve as a contribution to the open government research community.

This chapter progresses the following way. Based on five parameters, the ideal case is described. Subsequently, the selected case, K10,[78] is evaluated on these five parameters. In the end of the chapter, the IT artifact is thoroughly described.

3.2. The Ideal Case

As discussed further in Chapter 5 and 10, the preferred methodology for this research is online field experiments. Conducting online field experiments is only possible when using an online community that already exists, limiting the number of available open government communities in Denmark drastically. The problem is a chicken-and-egg problem: To examine how participation can be increased, I need an existing open government community with some level of participation. As most communities fail due to lack of participation (op cit.), it is hard to create open government communities.

The ideal case for the experiments in this dissertation fulfills five criteria: (a) government owned, (b) many users, (c) able to conduct experiments, (d) willing to conduct experiments, and has a (e) representative sample population.

3.2.1. Government Owned

The definition of open government (op cit.) has three different traits: (1) participation and collaboration, (2) open data, and (3) technology enabled. These traits to not address ownership, entailing open government comprises interactive online entities that operate in the governmental sphere, regardless of the ownership (This is thoroughly described in Chapter 2). This means community ownership only plays a limited role.

It is important to understand the noteworthy differences between online communities in the government sphere and outside this sphere. There are at least five relevant parameters on which they differ: recruitment/retention balance,

---

[78] http://k10.dk

contribution type, cost of contribution, legal conditions, and organizational culture. Of these five differences, ownership matters in the latter two.[79] For this reason, the ideal case in this dissertation is a government owned community. Such a community will be accepted as a case of open government by all observers, thereby making it a clear-cut case.

3.2.1.1. Recruitment / Retention Balance

To build and sustain a successful online community, it is necessary to recruit and retain users. The balance between the two varies depending on the community type and desired contributions. Some communities thrive best with continuous participation, while other communities benefit from a high turnover.

Many, but not all, non-governmental communities benefit from repeated contributions. Early evidence from eRulemaking shows that this might not be the case in open government communities. According to Farina et al. (2011), the users of Regulation Room tend to only comment on one rule. This means that users often comment when they sign up, but then leave to never return. This is not necessarily seen as a problem, since different knowledge is needed for different rules. It is, however, a problem if too few people participate. A high turnover rate can therefore be acceptable and sometimes even desirable.

3.2.1.2. Contribution Type

Different communities rely on different kinds of contributions. As described in Chapter 1, typical forms of contributions include posts, comments, articles, photos, video, music, ratings, reviews, events, tagging, and geo-local information. Different types of contributions are associated with different contribution costs and knowledge requirement, which complicates direct transfer of knowledge

---

[79] This is a simplification that does not capture all the nuances and differences between online communities. For example, having communities such as Pinterest and Wikipedia in the same category can be problematic as the cost of contribution is significantly higher in Wikipedia than in Pinterest. Despite these limitations caused by high intra-group variation, it is a useful heuristic when comparing and contrasting government communities to non-government communities.

about one contribution type to another. It is unknown to which degree results from a community evolving around posts as contribution artifact extrapolate to special open government community artifacts such as patents reviews[80] and rule comments.[81]

3.2.1.3. Cost of Contribution And Required Knowledge

To leverage as many contributions as possible, most online communities seek to reduce the cost of contributions. On Facebook, users can press the like button instead of writing a post. On Reddit, users can vote an article up or down, rather than of posting something new. On Twitter, users can simply retweet instead of writing own thoughts. Social media users are used to and encouraged to vote on and rate everything. For in aggregate, such contributions add at least some value (Farina et al. 2011).

Farina et al. (2011) argue that such light-weight contributions are unsuitable for open government, in part because the aim is not always to vote for the best idea but also because it undermines some of the benefits of open government, e.g. better input to decision-makers and civic education (Farina et al. 2011). Voting an item up does not give any nuances, and the user is not required to truly reflect on the content and implications thereof. The costs of contributions are often higher in government than in non-government. Therefore, it is harder to define tasks for synthesizers, as they prefer to do the light-weight contributions that are unsuitable for at least some areas of open government.

In addition to higher contributions costs, the knowledge required to contribute is also greater in open government communities than in many non-government communities. For example, a superficial comment such as "Cute Kitten" can be useful and valuable on YouTube, but more substance is expected and a necessity in open government participation.

---

[80] E.g. http://www.peertopatent.org/
[81] E.g. http://www.regulations.gov/

3.2.1.4. Legal Conditions

In the three previous sections, ownership plays no role. But ownership does make a difference when it comes to legal conditions and organization culture. Approximately 150 countries have freedom of speech rights.[82] Consequently, governments cannot censor citizens offline or online. This leaves open government communities with significant challenges in terms of handling and responding to vandalism, trolling, and flaming.

In this area, ownership does matter. Whereas private community owners are free to delete whatever they find inappropriate, governmental community owners might not have the same degrees of freedom.

Moreover, government owned websites are often quite rigid. Reasons for this include legal aspects, bureaucratic procedures, and reliance on external contractors (op cit.). Thus, government owned websites are hard to use for research purposes due to the bureaucratic processes and potential legal issues (Farina et al. 2010). Farina et al. (2010) argue that working with open government projects not owned by government "allows flexibility to experiment with designs and methods that could be difficult on an 'official' government site" (Farina et al. 2010: 398). These legal conditions combined with bureaucratic processes further make open government communities complicated to use for research and experiment (Farina et al. 2010). The high-level point here is that although rules and laws vary across countries,[83] governmental actors face tougher regulation than non-governmental actors.

3.2.1.5. Organization Culture

According to Eaves (2010), the organizational cultures in most contemporary public agencies share a number of characteristics: "centralized decision making, risk aversion, a strong delineation between insiders and outsiders, and deference

---

[82] http://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-4&chapter=4&lang=en

[83] For example, in the US it is challenging to collect much user data due to the Paperwork Reduction Act. http://www.archives.gov/federal-register/laws/paperwork-reduction/.

to authority and specialization" (Eaves 2010: 139). Open government is in many ways the counter culture to this Weberian notion of government.

Implementing open government communities requires increased transparency, reduced hierarchical oversight, and reduced delineation between insiders and outsiders. This means that the culture has to change for open government to truly succeed. Although some large non-governmental organizations share the same characteristics as government agencies, most non-governmental organizations have a better alignment with web 2.0 principles, and thus better preconditions for supporting online communities. Thus, online communities face more severe intra-organizational challenges when they are government owned.

3.2.2. Sizeable User Population

Conducting field experiments in online environments is only possible insofar there are active users in the environment. The more active users a community has, the more suitable it is for field experiments.

As described further in Chapter 5, statistical methods are improved when each experimental group has n > 30 subjects, as this fulfills the central limit theorem. This means that the optimal case has hundreds, if not thousands, of active contributors. A rule of thumb is that only 1% of all users are very active participants, and 10 % of all users engage in some participation. 90 % merely consumes other users' contributions (Ali-Hasan 2008; Cullen and Morse 2011). In an experiment with three groups, a minimum of 90 subjects are needed. This entails that the minimum number of users in the community must be 900. However, most data analyses are strengthened by segmenting groups further, entailing that the number of groups practically doubles. The experiment thus needs at least 180 subjects, which ought to be provided by an online community with 1,800 members. This assumes that the group sizes are created equally, which is rare.

To be on the safe side, the examined community has at least 2,000 members. The more users the better as it is conducive to conducting experiments quickly as

well as detecting statistical significance. As described in Chapter 1, only 75 % of all company sponsored communities fail to reach the 1,000 user mark.

It is problematic that the size criterion often conflicts with the first criterion: government owned. As described in Chapter 1, government owned communities have a hard time building and sustaining participation over time. Kbh.dk managed to gather 1,800 users over a nine-month period despite the big budget of DKK 7 million. There are so few successful open government communities – especially in Denmark – that it is almost impossible to use a government owned community as a case, when size is a requirement.

3.2.3. Representative User Population

The ideal case community has a population that reflects the general Danish population in terms of key demographics such as age, gender, and income. This assumes that open government is a singular concept in which all citizens take part of. But this assumption does not reflect reality very well. Most open government communities, as well as other online communities, operate within a niche.

This means that it is unlikely to find an open government community that accurately reflects the entire population. It is therefore crucial to examine which demographics the case represents. The main implication of not being able to study an online community that statistically represents the entire population is that the external validity is limited. This is discussed in depth in Chapter 5 and 10.

3.2.4. Ability to Conduct Experiments

In Denmark, there are only few open government communities with 2,000+ users.[84] This means that most open government communities in their current state have too few users to be able to conduct meaningful experiments. Even when

---

[84] Those known to the author include K10 (http://k10.dk), Sundhed.dk patient networks (https://patientnetvaerk.sundhed.dk), and digitaliser.dk (http://digitaliser.dk).

government agencies are willing to conduct experiments, the organizational set-up behind the community often makes it difficult.[85]

The current economic climate further limits government agencies' abilities to invest in new technology and experimentation. This means that many agencies will be unable to cover the costs of conducting the experiments. Conducting experiments in governments is further complicated by the fact that most government agencies rely on third party contractors and consultants to make changes. This means there has to be allocated budgets for such experiments, which might take quite some time to get through.

3.2.5. Willingness to Conduct Experiments

To conduct experiments on an online community, it is necessary for the owner to accept the premises of the experiments and cooperate. This means the owner as a minimum is ready to disrupt status quo and provide different content do different users. Experimentation also entails taking risks, since the outcome effects are unknown. After all, it is only due to this uncertainty that the experiment is conducted. Beyond disturbing the status quo, the experiments can go somewhat wrong or have unintended consequences and upset users. This is a risk the community owner must accept. As numerous experiments need to be conducted for this dissertation over a six months time span, the open government community owner needs to commit to collaborating on the project for at least this timeframe, even if the experiments upset users.

Furthermore, the ideal case is sufficiently transparent about its practices and results that they agree to publish data from experiments.

---

[85] For example, when selecting a case I talked to a government agency with a number of fairly successful online communities. They were interested in being used as case, but they only managed the community facilitation process. The software was delivered by an external IT-company, and the content and moderation of the communities was provided by several autonomous agencies. Although this agency is willing to engage in the work, it was unable to provide an appropriate test environment within the timeframe of this work.

Getting access to study online communities can be troublesome. As the literature review in Chapter 4 shows, the only online community included more than twice in the review is MovieLens. The reason this community is used several times is probably that it is owned by University of Minnesota, and thereby easily accessible to researchers.

One can speculate that it is harder to get approvals to conduct live field experiments on large, successful communities than small unsuccessful ones, as the former has more to lose in case something goes wrong. This makes it even more difficult to find an online community that has a sufficient number of users *and* is willing to let an external researcher conduct experiments.

3.3. K10 as a Case

The case used for the four experiments conducted in this dissertation is K10.[86] K10 is a Danish open government community for people involved with one of two public benefit programs called "early retirement pension"[87] and "flexjob".[88]

Two people, both of whom receive early retirement pension, created the community in 2005. They needed a place to find help to understand the rules and bureaucracy that accompany the early retirement program. In addition, they needed a place to express the frustrations they had experienced throughout the casework. Since they could not find a proper place to find help and express the injustice they felt, they took it upon themselves to create one. Today, only one of the founders is left and he maintains the community with help from long-term community members and administrators. As the K10 owner suffers from several illnesses, the time he can commit to the community varies quite a bit.

---

[86] The owner of K10 prefers to stay anonymous. His identity is known to me and confirmed as we have talked and met several times.

[87] Førtidspension

[88] Fleksjob

K10 sees itself as a place for people who are lost in the bureaucracy. The community members feel they help themselves and each other to navigate through complicated bureaucratic processes and incomprehensible laws and rules. But not everyone sees K10 this way. The Danish libertarian blog 180grader[89] is highly critical of K10 and calls it an extremist website ("Juul" 2011) for people who want to cheat the system (Anonymous blogger 2012). This highly critical perspective does not seem to be shared outside the 180grader blog.

Criteria	K10
Government owned	Owned by a citizen, but operates in government sphere.
Large user database	Almost 9.500 users, whereof 4,000 have written a post
Willing to conduct experiments	Yes
Able to conduct agile experiments	Yes. No long approval procedures or bureaucracy
Representative user database	Yes. The users are very different from the young and tech savvy generation. But not representative of the entire population.

Table 5: Evaluation of K10 as a case

Table 5 shows how K10 is evaluated on the five criteria used to evaluate the case fit. Each parameter is further described below.

3.3.1. Government Owned

The ideal open government case is a community or platform owned by the government. But as described in Chapter 1, many government owned communities have failed due to lack of participation. The absence of successful government-owned communities makes it difficult to find one with sufficient participation to conduct statistically valid field experiments.

K10 is not owned by a government agency, quite contrary. A single person is behind the community. This raises two concerns: is K10 an open government community, and if so, to which degree will the findings from K10 extend to government-owned communities?

---

[89] http://180grader.dk

3.3.1.1. Is K10 an Open Government Community?

Since K10 is owned by a single person, and not the government, it is worth discussing if K10 indeed is an open government community. As described in section 2.1, open government is characterized by three traits and neither address ownership: (1) participation and collaboration, (2) open data, and (3) technology enabled. When assessing to which degree K10 is an open government case, these three traits are relevant to address.

Participation and collaboration is a defining trait of K10. According to the *About* page on K10,[90] the community has the following purpose:

> *"K10 is created with the aim of sick people, who are trapped in the system during their case handling with the municipalities, have a place where they can get advice and support from other people who are in the same situation. This support can be of statutory or moral nature..."[91]*

Thus, one of the primary aims of K10 is to let users discuss current case handling, legislation, and statutory statuses.[92] Both retired and active government

---

[90] http://www.k10.dk/index.php?pageid=Om_K10

[91] http://www.ltk.dk/composite(528).htm, https://www.rk.dk/forum/ and http://www.rudersdal.dk/sitecore/service/notfound.aspx?item=%2fkommunen%2fdebat%2ffri_debat&user=extranet%5cAnonymous&site=website

[92] This purpose is rather similar to the purpose of government owned communities. For example, the municipalities of Lyngby-Taarbæk, Rudersdal, and Rødovre have open debate sections on their websites. The purpose of these fora is to let citizens "give your opinion on exactly what concerns you as a citizen of the municipality". These government-owned communities are not limited to any special topics, but are open to everything relevant to the municipality. Although the use of these debate sections is scarce, some of the topics are concerned with case handling and statutory statuses, just like they are on K10. See for example https://www.rk.dk/forum/?tt_board_uid=157&cHash=e4e6c6e729 (title: P*ublished social casework deadlines*) and https://www.rk.dk/forum/?tt_board_uid=141&cHash=def7549306 (title: *re: award of the apartment as a disabled pensioner*).

case handlers are following, and sometimes participating in, the discussions at K10.[93] This means the discussions are potentially used as input into policy processes, although this has not been examined at this point.

All the content and data on K10 is **open** to anyone interested. The content on K10 is mainly comprised by discussions about the two government benefit programs, flex job and early retirement pension. Many of the discussions are about case handling. Much of the questions posted on K10 could just as well have been sent to the relevant authorities as they involve questions about legislation, requirements, rates, terminology, rulings, rights to repeal, rights to complain, etc. Since much of the content on K10 could have been sent directly to government authorities, it seems logical to conclude that the content on K10 is within the government sphere. Posting the issues on K10 instead of contacting government agencies directly create transparency about government case handling.

Although some members occasionally meet in person, K10 is primarily an online community and thus **enabled by technology**. As described in Section 3.3, K10 relies on a simple technological concept: it is a vBulletin discussion forum where users can start discussions and answer them. This concept is identical to many government-owned communities, e.g. the forums offered by the municipalities of Lyngby-Taarbæk, Rudersdal, and Rødovre. Moreover, The Danish Agency for Digitization has an online community evolving around e-government called digitaliser.dk. This community comprises different forms of discussion sections as well as data sharing sections. The first part is very similar to K10.

As K10 fulfills all three traits of open government, it is considered to be an open government community despite its lack of government ownership.

---

[93] Case handlers are often contacting the owner of K10 with comments or questions about specific discussions, thereby strongly indicating that they monitor the community,

3.3.1.2. Extrapolation

The question of potential extrapolation is a complicated one, as no cross-community study covers both government-owned and non-government owned online communities. The biggest limitation with K10 as case is that it is owned by an individual rather than a public agency. The concern is not so much the degree to which it affects user behavior, but rather that the research process circumvents legal conditions and bureaucratic processes.

As described above, online communities in governmental and non-governmental spheres differ in five ways: recruitment/retention balance, contribution type, cost of contribution, legal conditions, and organizational culture. Of these five differences, ownership matters in the latter two. This means the findings from K10 should not be extrapolated to government owned communities in terms of legal conditions and bureaucratic processes, as these are circumvented in this dissertation.

On the other hand, the scope of this dissertation is open government community design and not studies of law or bureaucracy. The two things are related and interdependent, but it is not feasible to study both things comprehensively in one dissertation. Furthermore, circumventing juridical issues and bureaucratic processes makes it possible to conduct multiple experiments over a shorter time span. This opportunity strengthens this dissertation work.

3.3.2. Sizeable User Population

The community was launched in 2005. At the time of the first experiment, the community had approximately 8,800 users. This number increased to 9,500 users at the time of the fourth experiment.

K10 is an active community. Table 6 shows the community stats for 2011. In this year, a total of 4,268 new threads were started. The activity fluctuates quite a bit over time: the total number of new threads started per month ranges from 295 (in June 2011) to 501 (in January 2011).

Month (2011)	New Threads	New posts	New Users
January	501	5,444	259
February	454	5,179	266
March	491	5,456	267
April	300	3,440	147
May	369	4,026	188
June	295	3,390	126
July	318	4,689	168
August	355	4,634	183
September	408	4,929	175
October	392	5,264	182
November	477	6,588	206
December	409	5,373	157
Total	4,769	58,412	2,324

Table 6: Community stats for K10, January - December 2011

58,412 posts were written in 2011. The lowest number of posts per month is 3,390 (June 2011) and 6,588 is the highest number of posts (November 2011). In 2011, 2,324 people signed up to K10. The lowest number of new user signup per month is 126 (June 2011) and the highest number is 267 new users (March 2011).

3.3.3. Representative User Population

To know if an online community population is representative of a larger population it is necessary to understand this larger population. Therefore, the general flex job and early retirement pension recipient population is described, followed by a description of the K10 user population.

3.3.3.1. Demographics of Flex Job and Early Retirement Pension Recipients

The flex job benefit program is for Danish citizens aged 18-64 who (a) are unable to maintain a job at normal conditions, and (b) suffer from permanent and substantially reduced ability to work. As the ability to work is reduced substantially and permanently, the program is for citizens with serious illnesses. The program can be seen as a government-subsidized job in which the government typically pays 1/3 to 2/3 of the citizen's salary (Statistics Denmark 2012).

In 2012, 52,679 people were enrolled in the flex job program (Statistics Denmark 2012). The number of people enrolled in this program has increased substantially over the last decade, cf. Figure 4. As Figure 4 also shows, the majority of the flex jobbers are aged 40-59. Women are overrepresented in the flex job program as 61 % are women and 39 % are men. Citizens supported by the flex job program tend to be between the age of 40-59 and have a low level of education.
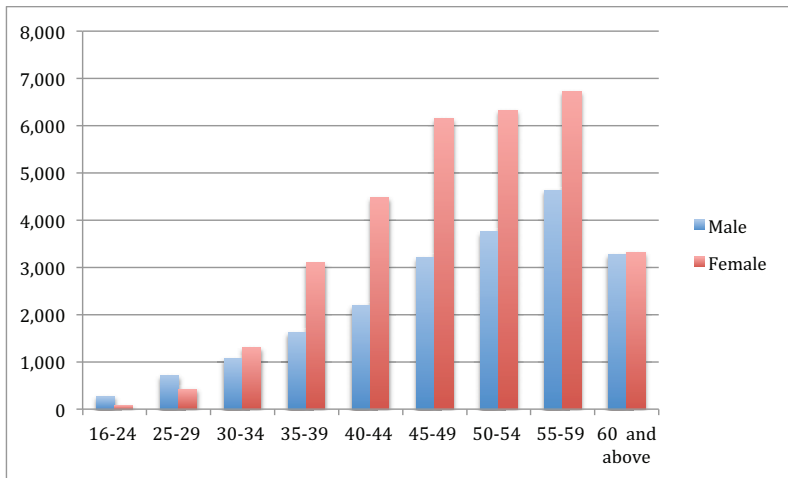
Figure 4: Number of people enrolled in the flex job program in 2012. Source: Statistics Denmark (2012)

Early retirement pension is a program for Danish citizens aged 18-64, who cannot work due to permanent disabilities. In 2012, 237,337 Danish citizens received the welfare benefit (Statistics Denmark 2012). This benefit program is a last resort for citizens unable to provide a living for themselves, as they are unable to maintain a fulltime job or a flex job. Early retirement pension is more often granted from psychological illnesses than physical illnesses (Statistics Denmark 2012).

It is possible to lose the early retirement pension due to improved health, but this is very rare in practice (Statistics Denmark 2012). Receiving early retirement benefits entails that citizens are leaving the job market for good, thereby

making them a marginalized group in the Danish society. According to Weatherall (2001), the risk of ending up on the early retirement pension program is greatest for people who share one or more of the following characteristics: low-income jobs, long-time unemployment, low education level, has been in an accident, single and no children, and being more than 50 years old (Weatherall 2001). Similar to the flex job program, women are overrepresented with respect to early retirement pension. Of the 243.630 people on early retirement pension, 54 % are woman, and 46 % male.

Figure 5 shows that the vast majority of the people in the program are over the age of 40. There is a negative correlation between education and early retirement pension, as 12 % of people with only a 9th grade school education are on early retirement pension, while 2 % of people with a graduate degree are on early retirement pension (Danmarks Statistik 2011).
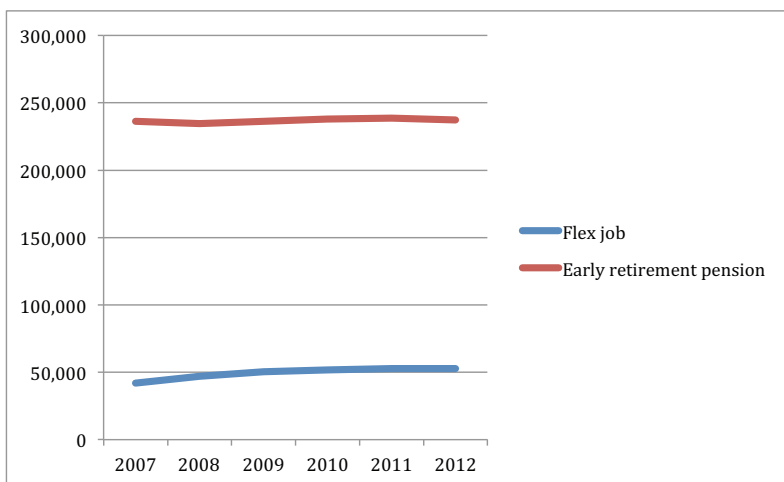
Figure 6 shows the development of flex job and early retirement recipients 2007-2012. The number of flex job recipients has increased from 41,875 in 2007 to 52,679 in 2012. The number of early retirement pension recipients has been steady: 236,238 people in 2007 and 237,337 people in 2012.

It is also worth noting that although citizens on the two benefit programs share some characteristics, there are fundamental differences. An important difference is whether or not the citizen is considered to be part of the workforce. People on flex job still bear a relation to the labor market while people on early retirement pension do not. People on early retirement pension are not unemployed; they are completely out of the workforce. The latter are, thus, more marginalized and stigmatized than the former. This means that the K10 users constitute a more heterogenic group than one could assume at first glance.

3.3.3.2. Demographics of K10 Users

K10 users can opt to enter their date and year of birth. Per April 1st, 2012, 1,752 out of 9,356 users have entered this data. Based on self-reported data from 18 % of the K10 users, it is possible to calculate the K10 subjects' age. The youngest

K10 member is 16 years old and the oldest 81 years old. The mean age is 44 years old and the median age 45 years old.

Figure 7 shows the age as reported by K10 users. It is impossible to know if these self-reported data are reliable and/or can be used to extrapolate to the users who did not fill out this information.
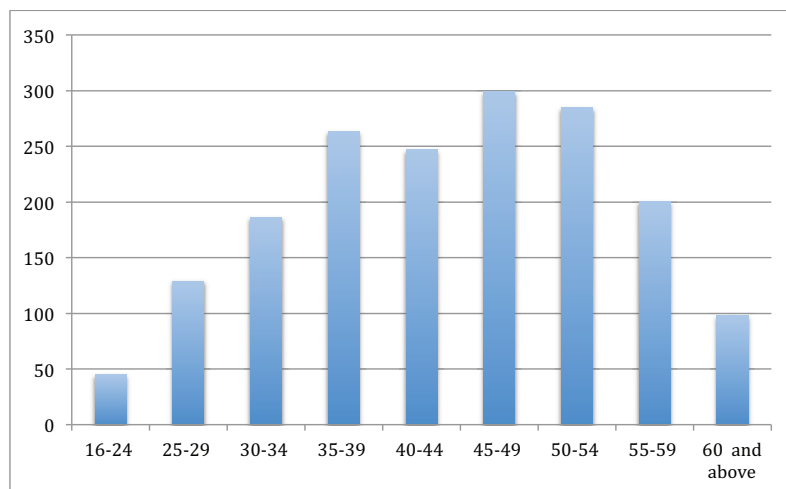
Figure 7: K10 users age. Source: self-reported data in K10 accounts

By reading discussions on K10, it is clear that the vast majority of the users are receiving, or trying to receive, flex job benefits or early retirement pension. This means that most of the users likely suffer from long-term illnesses. Furthermore, this segment of the Danish citizenry tends to be less educated than the population at large. This makes it an interesting case, as the users are different from those belonging to many other online communities and social networks, which tend to be dominated by younger, well-educated people with high income (PEW 2012).

Figure 8 compares K10 users' age (n = 1,752) with the age reported by Statistics Denmark (op cit.). Based on 2012 data, the chart shows that the K10 users are younger than the average flex job and early pension recipients. K10 has an

overrepresentation of young people and an underrepresentation of older people. This is not surprising as internet users tend to be younger than the average population (PEW 2012).
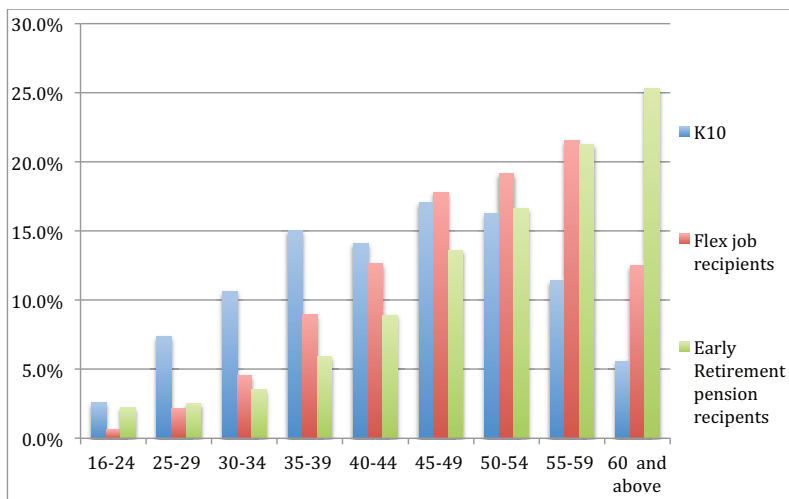
Figure 8: Comparison of K10 users and Statistics Denmark's data per 2012 (Source: Statistics Denmark 2011 and K10.dk)

Figure 9 shows the Danish population's use of social media based on age. This data shows that over 90 % of the citizens between the age of 16-24 have social media accounts while merely 12 % of the citizens between 65-89 have social media accounts (Statistics Denmark 2011). In short, the average social media population is younger than the general Danish population. At the same time, K10 users mean age is 44 years, entailing that K10 users tend to be older than Facebook and Twitter users.

Moreover, the citizens using social media tend to be better educated than those not using social media. Of those citizens attending primary and secondary school (year 1-10), the social media penetration is 51 %. The number is 52 % for Danes with a short higher education and 62 % for citizens with a long higher education (Statistics Denmark 2011). At the same time, flex job and early retirement recip-

ients tend to be lesser educated than the average Danish citizen. K10 users are thus less educated than the average social media user.
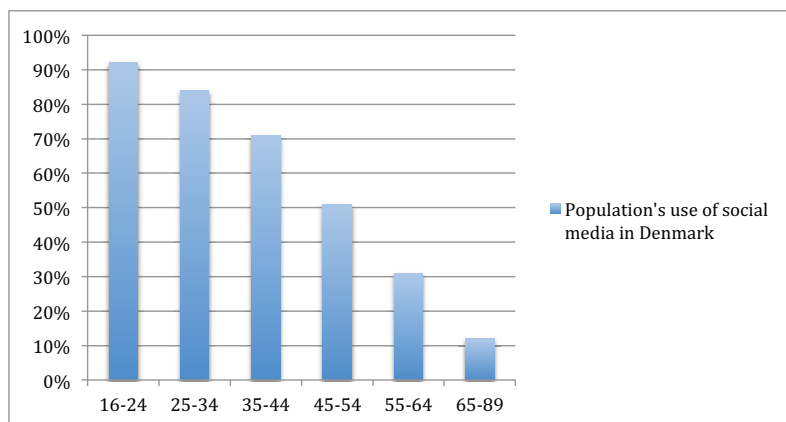
Figure 9: Danish population's use of social media (Source: Statistics Denmark 2011)

In sum, K10 is unlikely to be representative of the Danish citizenry at large. Based on age data alone, K10 users are younger than the average flex job recipients and especially early retirement pension recipients. This means that generalizations to the larger group must be made with caution.

3.3.4. Ability to Conduct Experiments

As described above, K10 is owned and managed by a single person. Due to various illnesses, the owner is an early retirement pension recipient himself. He is only able to work for a maximum of a few hours a day. Therefore, the owner has a very limited ability to assist the research, and K10 does not have any money to hire developers to do any coding. As there are few resources available to assist the research, it is done as lean as possible. This necessity has inspired lean experimentation, discussed in detail in chapter 10.

3.3.5. Willingness to Conduct Experiments

Although K10 does not provide resources to assist the experiments, the owner is willing to let the experiments be conducted under few constraints. The owner is

being informed about the experiments and provides feedback on the experimental design before the experiments are conducted.

3.4. The IT Artifact behind K10

In order to understand K10, it is necessary to understand the underlying platform and technology. As noted, the community software behind K10 is called vBulletin. In this section, both the technical aspects and user interfaces of vBulletin are described. It is not an exhaustive list of all features, as this is quite long.[94] It is, however, a description of the features and interface elements that are central at K10. Please see Screenshot 1 below to get an impression of K10.

vBulletin is a forum software developed by Jelsoft Enterprises and vBulletin Solutions. It is written in PHP and stores information in a MySQL database. Having data in a MySQL database makes it fairly easy to export and analyze data, as it is already structured in a database.

vBulletin is proprietary software sold via licenses costing between $249 - $399. The expensive alternative includes a mobile suite that lets the customer provide native iOS and Android apps. vBulletin was released by Jelsoft in 2000 and acquired by Internet Brands in 2007 (Limm 2007). According to vbulletin.com, over 100,000 websites use the vBulletin as forum software.[95] K10, thus, licenses its technology along more than 100,000 other online communities.

---

[94] For an exhaustive description of vBulletin features and functions, please refer to Kingsley-Hughes & Kingsley-Hughes (2006). Moreover, a full list of vBulletin features can be seen at http://vbulletin.com/features/

[95] Some of the major brands highlighted include Zynga, Sony, EA, and NASA.

Screenshot 1: K10 homepage (this screenshot is mean to give an overview of the homepage. Please go to K10.dk if you want to see a more readable version of the homepage or see some of the screenshots below)

3.4.1. Pros and Cons

As described above, the vBulletin software provides the basic features necessary to most online communities. vBulletin was one of the first forum software, but is by no means unique today. According to the forum software comparison service, ForumMatrix,[96] there are now more than 65 forum software alternatives to vBulletin, including successful ones such as phpBB and Vanilla Forums. These 65 different software solutions largely offer the same set of basic functionalities. The experiments conducted here are therefore fairly technologically agnostic as they can be conducted on most online communities. The findings of these experiments are therefore not limited to vBulletin communities.

In this context, it is worth briefly describing the pros and cons of vBulletin. The *pros* comprise (a) efficient code that scales well and can handle large volume of users and posts. (b) Supports multiple languages (including Danish). (c) Many useful community features, which are described further below. (d) Highly customizable in terms of both design and features (described further below) (e) The HTML is optimized for search engine access, thereby increasing the findability through search engines such as Google, Yahoo!, and Bing (Høgenhaven and Andreasen 2011). (f) Descriptive community statistics that provide an overview of the community. (g) Moderation tools to efficiently remove spam.

The *cons* of vBulletin include (a) a rather expensive license that has to be upgraded every time a new version is released. Although $249 - $399 might not sound overly expensive, it is much more expensive than some of the free alternatives. Moreover, (b) the interface is rather cluttered due to a large number of features. (c) The CSS is not coded to be responsive to different devices and screen sizes. vBulletin has sought to address this critique by selling native iOS and Android apps. But these are to be seen as supplements, not alternatives to a responsive website that is easily accessible through the HTTP protocol (Kadlec 2012). (d) Basic PHP knowledge is necessary to install vBulletin on a server as the config file has to be edited manually on the server.

---

[96] http://forummatrix.org, retrieved on April 7th 2013

3.4.2. How vBulletin Works

vBulletin is a self-hosted forum software that allows users to store information and let other users retrieve it. The vbulletin core is written in PHP and stores the data in a MySQL database located on the same server. The technical infrastructure can be explained in seven steps (vBulletin 2010):

Step 1: A visitor's browser sends a request to the web server requesting data through a URL (e.g. a specific thread such as http://www.k10.dk/showthread.php?t=26495).

Step 2: This request for a URL is routed to the web server running Apache, MySQL and PHP. The request calls the file showthread.php on the server along with the posted variables (t=26495 in this example)

Step 3: The Apache web server checks the request to the .php file and creates a data package to return to the visitor. As the apache server does not understand PHP code, the request is forwarded to a pre-compiled PHP module.

Step 4: The PHP module executes the commands. In order to get specific data such as user information, post content etc., the PHP script sends multiple database queries to the MySQL server, requesting the data.

Step 5: When receiving the queries, the MySQL server returns the data to the PHP module so it can finish the executed PHP script.

Step 6: When the PHP module is finished executing the script written in the PHP file, the Apache server can build the requested page and return it to the visitor.

Step 7: The visitor can see the requested page in her browser.

This process is illustrated in Figure 10, which is an adaption of Ullman (2011).
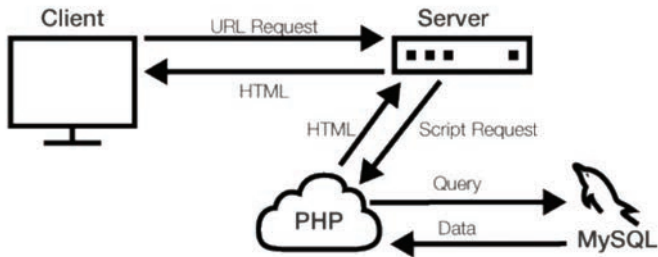
3.4.2.1. User Interface

As mentioned above, over 100,000 online communities use vBulletin software. In order to function in so many different contexts, the standard software template can be customized to fit various needs. The customization opportunities include visual elements, (e.g. different default templates and full access to change cascading style sheets (CSS)) as well as functions (e.g. possibility to install add-ons and plug-ins and language/phrase managers as well as develop own add-ons in PHP to further customize the functionalities in vBulletin).

3.4.2.2. Threads and Posts

A forum is categorized into threads, created by users and moderators. Moderators can edit, delete, move, and lock threads, and users can subscribe to new posts via email.

Once a thread is created, users and moderators can write replies. This is done by writing posts. All posts are listed within a thread, although the page is paginated when 10 or more posts are written within a single thread.

The user has a number of options when writing a post. As seen in Screenshot 4, the user is presented with a *What You See Is What You Get* (WYSIWYG) editor. This editor supports rich HTML formatting and smileys/emoticons. It is possible

to attach images (in popular formats such as .gif, .jpg., etc.) stored on the server that hosts the forum. The user can preview and edit the post before making it public. Moderators can edit, delete, and move posts.
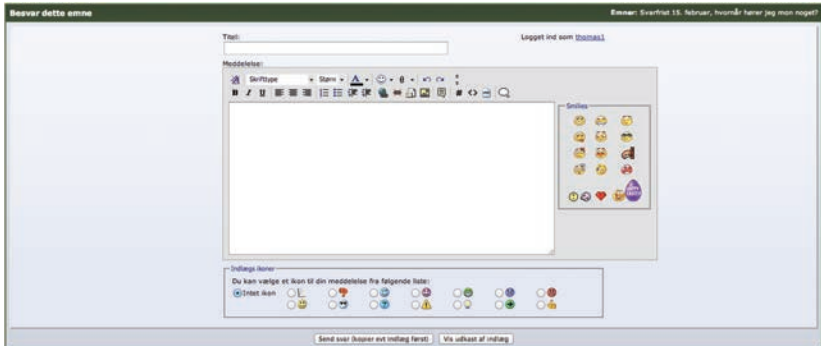
Screenshot 4: New post editor in vBulletin
(http://www.k10.dk/newreply.php?do=newreply&noquote=1&p=257942)

3.4.2.3. User Profiles

Everyone with access to the internet can read threads and posts on K10. In order to post something the user must create a profile with a unique username. This means that the users in a vBulletin system have a consistent username over time, although their real identity can remain anonymous if they prefer. As described further in Chapter 5, the vast majority of K10 users do not use their real identity.
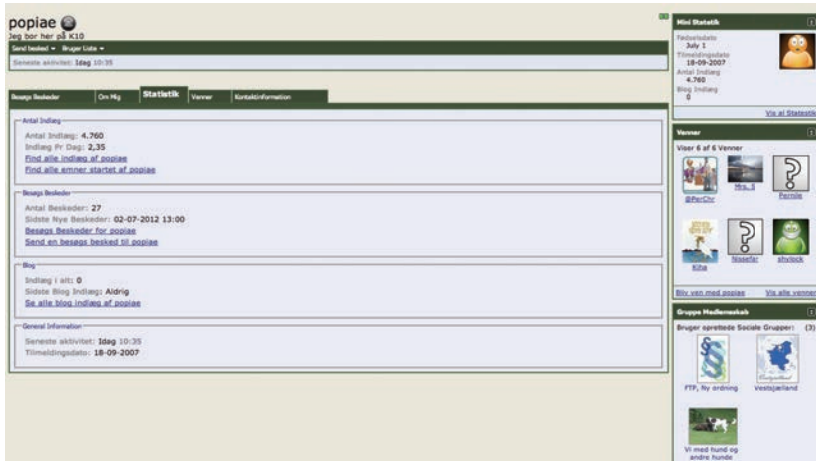
As shown in Screenshot 5, the user profiles contain self-reported data (including avatar, date of birth, location, occupation, and interests) and usage data (including number of posts, number of posts per day, registration date, and most recent login date). Finally, users with profiles can send messages to each other at K10 through the vBulletin software. A message can either be a public message that is shown at the recipient's profile page or a private message that is only visible to the recipient. An example of the latter is shown in Screenshot 6.
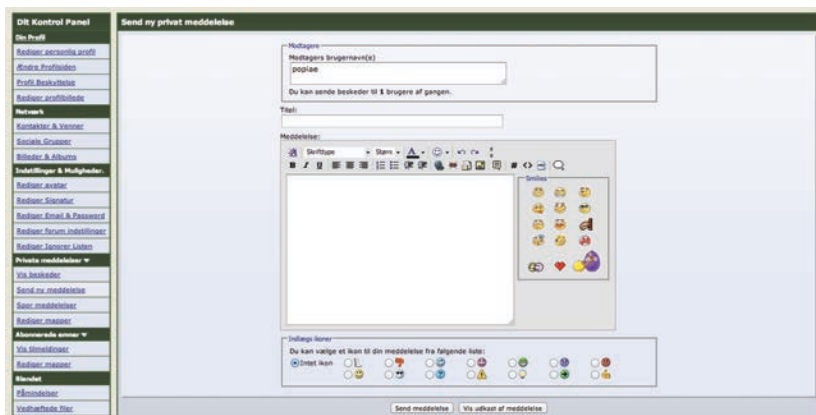
3.4.3. Security

vBulletin is generally considered to be secure (Kingsley-Hughes and Kingsley-Hughes 2006). The greatest security risks are caused by the underlying technologies: PHP and MySQL. These technologies are vulnerable toward denial of service (DDOS) attacks (Vaidyanathan and Mautone 2009), unauthorized access to the database (Davis and Phillips 2009), unencrypted data in the database (Curioso, Bradford, and Galbraith 2010), storing passwords in scripts, shared hosting vulnerabilities (Davis and Phillips 2009), not updating to newest versions (Curioso, Bradford, and Galbraith 2010), erroneous security permissions (Curioso, Bradford, and Galbraith 2010), MySQL injection (Curioso, Bradford, and Galbraith 2010), cookie theft (Shiflett 2009), filename manipulation (Shiflett 2009), and password sniffing (Shiflett 2009).

These security concerns are present for millions of websites, and K10 is no exception. However, K10 has little or no sensitive information such as credit cards and personal number (CPR numbers). The only non-public information is user login credentials and private messages sent directly between the users. Due to the lack of sensitive information, it is unlikely that anyone will try to hack the website. Only one security problems have been detected so far in K10's lifetime Turkish hacker group initiated a deface attack, in which the K10 site was replaced another website. Such website defacements are usually made with MySQL injections. The site was quickly recovered after the defacement attack. The users' privacy and data management during the experiments are discussed further in section 5.7.2

3.5. Chapter Summary

K10 is used as the case in the four experiments of this dissertation. K10 is a Danish online community for people receiving or applying for early retirement pension or flex job benefits. The community was created in 2005 and has 8,800-9,500 members during the period when the four experiments are conducted.

K10 is a good fit but not an ideal choice. The biggest limitation of the case choice is that K10 is privately owned. A compromise was made since there are very few government owned communities that also have a sufficiently large user population. It is not expected that this will have any impact on the user behavior in the experiments. It does, however, mean that no insights are yielded into bureaucratic processes behind open government communities.

The major benefit of using K10 as the field setting for the experiments is that the owner is willing to test different design constructs of social psychology theories. Furthermore, the case aligns well with the ontological, epistemological, and methodological choices described in the next chapter, since the owner has few resources to enact in the experiment. Finally, K10 is a good case because the population is different from the online communities usually studied. K10 users are suffering from long-term illnesses and are generally lesser educated than the general public.

Part 2

Theory and Methodology

Chapter 4: Social Psychology Theory

4.1. Introduction

As documented thus far in this dissertation, open government communities face severe participation challenges in terms of recruitment and retention of members. One way to seek to overcome this challenge is to use social science theory to generate design ideas and, thus, serve as a means to increase participation in open government communities (Kraut and Resnick 2012).

As illustrated in Figure 11, the theory selection process in this dissertation follows a funnel. The selection process begins broadly with a choice between different social science disciplines. After choosing a social science discipline, the literature in this field is reviewed more in depth. The outcome of the literature review is, among other things, a list of the theories used within the social science discipline. Based on this review, the theories can be classified and an appropriate number of theories selected and subsequently operationalized into hypotheses and experimental designs.
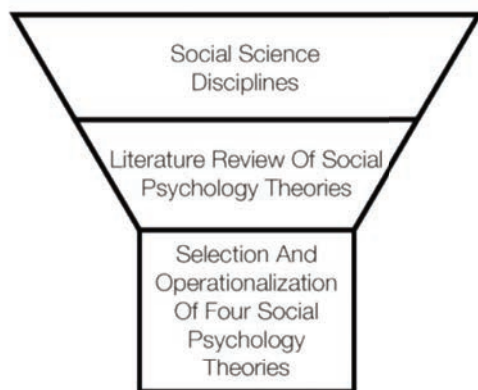
Figure 11: Funnel approach used for theory selection in this dissertation

In this chapter, four relevant different social science sub-disciplines are examined: economics, political science, sociology, and social psychology. Ultimately,

social psychology is chosen as the theoretical perspective for this dissertation because it is highly compatible is terms of methodology and granularity.

The main contribution of the chapter is a literature review on how 12 different social psychology theories are being used to examine and guide online community design. The literature review is primarily a contribution to the HCI community where the reviewed research is published. As the included reviews focus on improving and increasing participation in online communities, the review is also likely to be contribute to the open government community.

The aim of the literature review is to identify literature that helps set up experiments and thus answer the overarching research question. The literature review comprises 12 social psychology theories used in 28 papers across 34 named cases and six undisclosed cases. The reviewed papers rely on five different research designs and six different methodologies. Through this review, it is possible to find theories that successfully can guide design. In other words, social psychology theories are examined as a mean to inspire and guide online community design. Thus, it is not the social psychology theories *per se* that are described, but the appliance of the theories to online communities.

This entails that this chapter's aim is *not* to give an exhaustive overview of social psychology theory in general. Although useful, this task is too comprehensive for this dissertation. This chapter focuses, more modestly, on how social psychology theory is used to examine and increase participation in the online community literature in 2000-2010. Moreover, this review does not examine which theories are right, and which are wrong. After reviewing the literature, four theories are selected and operationalized for the four experiments conducted in this dissertation.

4.2. Social Science and Design

Many notable HCI researchers (e.g. Carroll 1997; Kraut 2003; Olson & Olson 2003; Carroll 2003; Arazy, Kumar, & Shapira 2010; Kraut & Resnick 2012) call

for increased utilization of social science theories in the online community field.[97] Kraut and Resnick (2012) argue that "social science findings can and should inform more directly the choices that online community designers make" (Kraut &Resnick 2012: 10). Social science theories are helpful in information science and open government community design for at least five reasons: theories offer rich descriptions (Kraut & Resnick 2012; Eriksson 2010); theories help avoid previous mistakes (Eriksson 2010); theories can generate ideas and solutions (Kraut & Resnick 2012); theories help designers be more explicit about hypothesis, assumptions, and goals (Eriksson 2010); and theories help social scientists make an impact (Carroll 2003).

Based on the calls for increased leverage of social science theories in online community research and the five benefits listed above, this dissertation relies on social science theory to inform the hypotheses and design.[98] That the social sciences can and ought to inform online community design is the premise and point of departure of this dissertation.

In the social sciences, theory is usually related to empirics (King, Keohane, and Verba 2001). This relationship between theory and empirics can take various forms. In some research, the purpose is to use empirics as a means to falsify or verify a theory. In other research the theory is used as a means to understand an empirical field (Bordens & Abbott 2011). In this dissertation, I take the second approach, and use theories to examine and potentially improve the empirical field, i.e. participation in open government communities. Thus, my primary contribution is to open government communities, not social psychology theory. As far as the experiments can give new insights into theory, the value of the research increases. But it is not the primary purpose.

---

[97] The English Oxford dictionary defines social sciences as "The scientific study of human society and social relationships." (cited in Vonhof 2010).

[98] In addition to these reasons, it is worth mentioning that the author has a background in the social science with a B.A. and M.Sc. in Political Science.

4.2.1. The Challenges of Using Social Science in Design

Despite the many benefits of using social science disciplines in design tasks such as online communities, there are a number of challenges. It is important to address them before selecting a social science perspective. Arazy, Kumar and Shapira (2010) highlight four challenges in selecting and using social science theories to guide design: (a) the theory's scope is too narrow, entailing that a single theory cannot solve an entire design problem; (b) kernel theories inform about the direction of the effect, but not the magnitude; (c) it is hard to find kernel theories that are relevant for a design task; and (d) granularity challenge as the theory's level of abstraction is too high to solve a specific task. Add to this the challenge of (e) knowing if a theory has been applied (Dourish 2006). These five challenges are discussed in the following sections.

4.2.1.1. Scope of Existing Theories

Designing online communities is a complex task, in which a plethora of factors have to been taken into consideration. As social psychology comprises many mid-level theories and not merely a grand theory, it can only solve part of a design problem (Carroll and Kellogg 1989; Beenen et al. 2004; Arazy, Kumar, and Shapira 2010). In other words, it is not assumed that one social psychology theory can create the perfect online community. More modestly, it is believed that a theory can improve the online communities on a few parameters.

There are at least two possible strategies that can be employed to leverage theories in this dissertation: *First*, one theory, e.g. social comparison theory, can be used across all experiments. This will provide an opportunity to examine the chosen theory very thoroughly. It can be tested via different constructs, e.g. qualitative and quantitative. The benefit to this approach is that it maximizes the insight into the possibilities and limitations of the theory. The disadvantage is that the limited fit between theory and empirics only will give one of many possible insights into online community design.

The *second* approach is to use different theories in the four experiments. This will make it possible to examine the effects of different social psychology theo-

ries on open government communities. The examination may, however, be less thorough than in the first approach. But most importantly, the likelihood of improving participation in open government communities increases when one relies on several theories. Furthermore, an important aim in science is to understand a diversity of needs, goals, and dynamics (Benkler 2006). Using more theories vis-à-vis one theory helps this diversity. To answer the research question in the best possibly way, I consider it more fruitful to explore several theories. As a consequence, contribution to practice is prioritized over contribution to theory.

4.2.1.2. Effect and Magnitude

Theories usually make claims about causality. Knowing causal directions is important, but not sufficient to make design decisions. Using a theory to change design is not very helpful if the effect size is small. This problem with low – and unknown – effect sizes is present in the social psychology literature. For example, a high effect size detected in a laboratory study does not necessarily entail that the effect has a similar size once implemented in real settings (Sadri & Robertson 1993). Synergistic effects and interaction effects often make the effect size vary across different settings.

Only few of the included studies in the review measure the effect sizes. This means the effect sizes in the online community domain is virtually unknown. To help overcome this issue, the experiments conducted in this dissertation measure intervention effect sizes.

4.2.1.3. Congruence between Theory and Design Task

It is challenging to find theories that can solve a specific design task (Walls, Widermeyer, and El Sawy 2004; Arazy, Kumar, and Shapira 2010). The problem is that social science theories and design theory have distinct, sometimes even incompatible, purposes, and foundations (Eriksson 2010). Walls, Widermeyer, and El Sawy (2004) argue that the "goal of a scientific theory is to understand or predict natural phenomenon…. While the purpose of a design theory is to guide artifact creation" (Walls, Widermeyer, and El Sawy 2004: 45).

Although there are different purposes and foundations of theory and design, it does not follow that there is no common ground at all: "Science and design have different principal objectives. It may be simplistic to imagine deductive relations from science to design, but it would be bizarre if there were no relation at all" (Carroll and Kellogg 1989: 13). Thus, using theory in design is not as simple as one might wish. . It takes effort to apply a theory successfully (Eriksson 2010),[99] and it is important to select a social science theory that is compatible with the design task.

4.2.1.4. Level of Abstraction Match

The level of abstraction and granularity in a theory ideally matches the design problem that is being explored (Arazy, Kumar, and Shapira 2010). As the four theories used in this dissertation have been developed well before online communities, there is a potential level of abstraction disparity. The theories are concerned with general human behavior while design tasks are concerned with behavior and problem-solving in a very specific context (Newell and Card 1985; Beenen et al. 2005).

To mitigate this issue, it is crucial to select theories with a good level of abstraction match in terms of the specific open government community context. This makes the theory more relevant, but might reduce the generality of the theory. Again, theoretical claims are sacrificed in order to offer as many insights into open government communities as possible.

4.2.1.5. How is it *Known* that Theory is Being Applied

Whether or not a theory is indeed applied is the most challenging critique to address. There are two challenges: whether or not a theory is being applied, and whether or not we can *know* if theory is being applied.

---

[99] This theory application challenge is not as big to online community design as many other design tasks. For as argued by Ma and Agarwal (2007) and Eastwick and Gardner (2009), the offline and online community settings are sufficiently identical to apply "old" theories online.

According to Dourish (2006), theory is often used in the HCI domain without a proper understanding of the theory. The underlying cause is that HCI research – and thereby online community research – is overly concerned with implications for design. Although many researchers and designers in HCI and IS refer to theories, they only apply it superficially. Simply put, designers often fail to use theories appropriately (Fedorowicz and Dias 2010). Some research (e.g. Li and Chignell 2010) claims to rely on social psychology, but does not specify which theories are used, how it is translated into online community design, or which constructs are used (Arazy, Kumar, and Shapira 2010). This approach does not further the usage of social psychology in online communities. It merely uses theories as a way to legitimize research, rather than informing it. In this dissertation, the applied theories are studied thoroughly and carefully operationalized into hypothesis and design. Thus, theories are not merely used as a mean of legitimization, as all hypotheses are derived from theory.

Even though theory is being applied carefully and rigorously in an online community experiment, it can be very difficult to know if the experimental intervention really is a construct of the theory. Single behavior can be interpreted in multiple ways and caused by different motives, drives, and psychological means (Carroll and Kellogg 1989; Dunning 2010). It is complicated to create certainty about the relationship between observations and theory, and thus to know if the theory is indeed able to explain the observed behavior. It is difficult to establish firm causality about the experimental construct of a theory. The same theory can be measured through multiple constructs, and the same experiment can be a construction of multiple theories.

Social psychologists have developed indexes and scales (op cit.) to measure theoretical constructs. These are usually measured through comprehensive survey instruments. Although they would help determine whether or not a theory is applied, these indexes are *not* used in this dissertation. Partly because their fit with online field experiments is awkward as it increases the likelihood of the Hawthorne effect. And partly because the users at K10 suffer from various disa-

bilities and may have a hard time focusing on and understanding the comprehensive indexes and scales.[100]

4.2.2. Social Science Disciplines

The social sciences comprise a number of sub-disciplines, most notably anthropology, economics, political science, political psychology, psychology, and sociology. Several of these disciplines can be useful to answer the research question of this dissertation, but only one will be applied. This choice is made in order to maintain a clear scope, and thus be able to use one sub-discipline more thoroughly instead of using several superficially. Under the circumstances of infinite resources, it would be optimal to include all the sub-disciplines. Alas, resources are always limited. This does not mean the discarded theories are not useful for the research, but that the selected sub-discipline is slightly more useful and fruitful than the others.

Of the many social science disciplines listed above, five are highly relevant when examining and improving social entities like open government communities: behavioral economics,[101] sociology, political science, anthropology, and social psychology.

4.2.3. Criteria for Selecting a Social Science Discipline

The philosophy of science literature offers different approaches to theory selection. Poincaré argues one should choose the most simple theory (Keuzenkamp and McAleer 1995). One obvious caveat of this approach is that the simplest theory might not have the sufficient level of granularity and/or grasp the complexity of the subject matter. Popper (2002) suggests selecting the theory with the strongest empirical support. However, in an emerging field, it is almost impossible to know which of many theories has the strongest empirical support *a priori*

---

[100] Please refer to Chapter 3 for a full case description.

[101] Behavioural economics is only a sub-discipline in economics, but the one usually used to understand social interactions outside the strictly economic sphere.

the research is conducted. This makes Popper's argument hard to apply in this dissertation.

Flyvbjerg (2001) has another perspective on theory selection in the social sciences. Based on Aristotle's *phronesis* concept,[102] Flyvbjerg argues that no theories or concepts can be context-independent.[103] In other words, "The context cannot be excluded because, as Bourdieu shows, context defines the type of phenomenon which the theory encompasses" (Flyvbjerg 2001: 47). Although context-independent theories are the ideal of social sciences, it is impossible (ibid.). Boellstorff et al. (2012) have a similar line of argument to the study of virtual worlds where the specific contexts ought to guide the decisions. Moreover, social science based on the *phronesis* concept is also more likely to be relevant to people outside academia. *Phronesis*-based science has critics (e.g. Laitin 2003) as well as proponents (e.g. Schram and Caterino 2006).

Flyvbjerg's argument entails that social science disciplines have to be evaluated and selected with the specific context in mind. This will both increase the social science discipline's correspondence with the case and make the research more applicable in practice. Although social science discipline selection ought to be context dependent, it does not entail that the selection is reduced to arbitrariness or gut-feelings. Instead of trying to find the most "correct" or the simplest social psychology theories, the aim is to find the most relevant ones to examine the case, K10.

	Poor methodology match	Good methodology match
Poor level of abstraction match	Political Science	Sociology
Good level of abstraction match	Anthropology	Social Psychology Economics

Table 7: Methodology match and level of abstraction match of the four social science sub-disciplines

---

[102] *Phronesis* refers to the need for practical wisdom. That wisdom cannot be reduced to general truths (Flyvbjerg 2001).

[103] This is the main difference between social sciences and natural sciences.

Table 7 shows how the five included social science disciplines perform across two parameters that are very relevant to the case examined in the dissertation: methodology match and granularity level match. In the following sections, the four social science disciplines are briefly described. Examining different theoretical perspectives is helpful in order to choose the right theory and help understand the comparative strengths and benefits of it. It is worth noting that this is not a general assessment of the sub-disciplines, but strictly an assessment related to open government communities.

4.2.3.1. Behavioral Economics

Behavioral economics is primarily concerned with using economic theory to explain and predict human behavior in non-economic domains (Lowenstein 1999).[104] One common denominator in the application of economic theory is that it rests on an assumption about rational actors (Green and Shapiro 1994). This simplicity makes the theory appealing, and helps explain why it is applied to so many subject matters outside the economic field.

Laboratory experiments are often used as the applied methodology in behavioral economics. This methodology is well aligned with online community research, thus making behavioral economics a potentially good match for generating hypotheses for open government communities.

In fact, behavioral economics is often used to examine and improve online communities. In the perspective of behavioral economics, online communities are treated as a market with a public goods problem. Users are considered to be selfish free-riders, who prefer to consume the value of online communities without producing it (e.g. Kollock and Smith 1996).[105] Thus, the mere existence of online

---

[104] Economic theory is being used in many non-economic fields such as political science (Downs 1957), psychology (Kahneman 2003), law (Posner 1998), and even football (Kuper and Szymanski 2009).

[105] This free-rider problem is sometimes referred to as the tragedy of the commons (Hardin 1968; Ostrom 1991; Wasko and Faraj 2005)

communities is a paradox in the sense that fully rational people, *homo economicus*, will not contribute to the community themselves, but only consume information.

The simplicity of behavioral economics is both its strength and weakness.[106] Due to the few, simple assertions about human behavior, only a quite limited number of hypotheses can be derived from this theory. However, the theory has already been used quite extensively to study online communities (see for example Lampel and Bhalla 2007; Ba, Whinston, and Zhang 2003; Ba, Whinston, and Zhang 1999; Ciffolilli 2007; Tedjamulia et al. 2005; Jeon, Kim, and Chen 2010; Rafaeli, Raban, and Ravid 2005; Chen, Ho, and Kim 2010; Harper et al. 2008; Dellarocas 2001). This strand of theory will only offer limited novel insights into the empirical subject matter of this dissertation. Behavioral economical theory is therefore not the optimal theoretical perspective to employ when examining open government communities as the findings are less likely to be novel and, thus, fruitful.

4.2.3.2. Political Science

As the subject matter is open government, political science is highly relevant to the empirical subject. Political science can help examine how open government can transform public administration and how open government policies are being adapted around the world (e.g. Janssen 2011). Especially one area within

---

[106] Behavioral economics can explain some of the behavior in online communities. However, it can also be critiqued on two grounds. The *first* critique is that online communities are not in the economic sphere. Although online communities can be treated as a solely economic matter, this approach neglects complex social relationships and dynamics. As argued elsewhere, behavioral economics serve as a prolific theoretical perspective within the economic domain, but a vague perspective in non-economic domains (Nielsen and Høgenhaven 2010). The *second* important critique is that the assumption about rationality and self-interest is often critiqued. The assumption about selfishness holds little empirical validity in the domain of online communities. People only consume instead of contributing for many reasons such as poor group fit, lack of motivation, lack of knowledge, lack of learning, usability errors, lack of requirement etc. Selfishness is only one of many reasons for not contributing (Wasko and Faraj 2000; Preece, Nonecke, and Andrews 2004; Shen et al. 2010; Luther and Bruckman 2010).

political science is potentially beneficial in open government research: public governance.

Public governance can be defined as "self-organizing, interoganizational networks" (Rhodes 1996: 652). This theory strand is closely related to the open government literature as both of them aim to open up governance to non-governmental actors via formal and informal networks. New public governance is, moreover, concerned with "policy and spatial domains where multiple public, private and non-profit actors join together to shape, make and implement public policy" (Skelcher, Mathur, and Smith 2005: 573). This strand of political science is concerned with democratic performance and policy output (ibid.).

For the purpose of this dissertation, new public governance and political science in general, have one significant caveat. More often than not, political science is focused on structures and institutions rather than the individual citizen.[107] New public governance theory asserts that it is desirable to include networks and non-governmental actors in the policy process (Peters and Pierre 1998) and defines the institutions and structures for this. However, there is little or no direction on how to get the citizens to interact and contribute to these institutions. In other words, there is a granularity level disparity between most political science theory and the aim of this dissertation. Due to this discrepancy, it is difficult to generate design hypotheses from the political science literature, which limits its value for this dissertation. Therefore, political science is not used as the theoretical perspective.

4.2.3.3. Anthropology

Broadly speaking, anthropology is "the study of other cultures, employing the technique of participant observation and collecting qualitative (not quantitative)

---

[107] For example, David Easton's canonical definition of political science is "to seek to understand political life is to address oneself to the study of the authoritative allocation of values (valued things) for a society." (Easton 1985: 134).

data" (Barrett 2009, 3f).[108] Anthropologists often locate themselves close to the subject matter and embed themselves in the culture they study in order to understand the meanings of actions. Such field research often takes months or years (Boellstorff et al. 2012).

Anthropology is often used in HCI (Boellstorff et al. 2012). In the study of online communities, anthropological methodologies have been used to examine "cross-cultural, multileveled, and multi-sited phenomena; emerging constructions of individual and collective identity; and the culturally embedded nature of emerging communicative and social practice" (Wilson and Peterson 2002: 450). The studied topics include IT artifacts as cultural products, identity negotiation, and power dynamics (Wilson and Peterson 2002).

In order to examine these research topics, the methodologies of long term participant observation and ethnography are often used. One of the dominant methodologies in anthropology is ethnography, which "typically seek to produce detailed and situated knowledge accounts of specific cultures in a manner that reflects the perspective of those whose culture is under discussion" (Boellstorff et al. 2012: 14). For example Nardi (2010) has been an active participant in World of Warcraft for several years in order to gain an in-depth understanding of the cultural codes in the virtual world.

The strength of anthropology is thus to gain deep understandings of the cultural practices in online communities (Boellstorff et al. 2012). As anthropology relies on qualitative data, the social science discipline is not the perfect fit when it comes to examining the participation challenge. To examine and increase participation, it is suitable to rely on quantitative metrics in order to measure the effect of different design patterns. This means that the anthropological perspective is not used due a methodology discrepancy with the specific research question. This is not to say that anthropology has nothing to offer the open government

---

[108] Anthropology has a large number of sub-disciplines, and the definition used above is not a precise definition for all these subfields. It is outside the scope to cover these various sub-disciplines.

community research field, merely that the methodologies are less suited for examining participation quantity.

4.2.3.4. Sociology

Sociology is "the scientific study of human society" (Odum 1952). The aim is explain general patterns of behavior and interactions, and often promote reflection and social change. The aim is often to create knowledge that can be applied to social policies and facilitate changes on the overarching societal level.

Sociology has a long history of studying communities in an offline context.[109] The theoretical perspective might therefore be appropriate when studying these communities' online counterparts. In terms of online communities, sociology-based research often focuses on qualitative relations between actors (Rheingold 2000; Reich 2010), how technology changes relations and behavior (Castells 2011; Sosik et al. 2011), and quantitative network analyses (Carrington, Scott, and Wasserman 2005; Easley and Kleinberg 2012).

Sociologists often rely on four research methods: surveys, experiments, participant observations, and secondary data analysis. As described in Section 4.2.3.1, experiments are a methodology that is well aligned with HCI and online community research. This makes sociology a potentially good fit with the subject matter of this dissertation.

As noted by (Thoits 1995), "Sociologists generally devote their efforts to identifying *which* social phenomena have effects on individuals" (Thoits 1995, 1231). Contrary to this, psychologist and social psychologist "generally specialize in identifying *the mechanisms or processes through which* social phenomena have their effects on individuals'" (Thoits 1995:1231). Although this difference might be subtle, it is crucial in terms of this dissertation. When seeking to improve participation in open government communities, it is critical to understand the specific mechanisms that help shape human behavior.

---

[109] See Weber 1905 for an example of early sociological work in this domain

Sociology is not used in this dissertation, primarily due to the mismatch between theory and design. Most sociological theories focus on a macro-level, and are hard to turn into specific micro-level design patterns and features. These analyses are helpful in understanding online communities, but not necessarily in generating specific design ideas and patterns. It does not follow that sociology has nothing or little to offer in this domain. Critical reflection over the impact of open government to wider society is necessary, and sociology can help facilitate this. This is, however, outside the scope of this dissertation.

Due to the granularity level discrepancy, sociology is not used as this dissertation's theoretical framework; at least not directly. For as described in the next section, this dissertation relies on social psychology, which can be seen as a sub-discipline of sociology. And it is sometimes difficult to distinguish between sociology and social psychology (Thoits 1995). Some theories, such as reciprocity theory, are categorized in both fields.

4.2.3.5. Social Psychology

Psychology serves as an important reference discipline in IS and HCI literature. As the web is becoming more social and the users increasingly interconnected, it is no longer sufficient to understand the individual psychology and cognition of the users. It is also necessary to understand social dynamics and how groups affect the individuals' thoughts and behavior (Ren, Kraut, and Kiesler 2007).

Social psychology can be traced back to 1898. The field is well developed, and serves as a sound foundation for research offline and online.[110] Theories from

---

[110] The discipline is starting to focus on the internet, but the internet is often used as a mean to collect data rather than studied as a subject matter per se. Reis and Gosling (2010) argue that "The online world is a legitimate venue in which to examine a plethora of social psychological behavior" (Reis and Gosling 2010: 92). Reis and Gosling further point to a number of such internet venues, including online support groups, social networks message boards, and blogs. Moreover, although social psychology is mostly aimed at physical relations, there is little doubt most of the findings are

this field can be used to examine and explain current levels of motivations, as well as enlighten design strategies and tactics leading to increased user contributions. Social psychology offers several lenses as the field comprises many mid-level theories, rather than one coherent theory (Kraut & Resnick 2012).

In the words of Clemmensen (2006), the use of social psychology "may help illuminate the social and cognitive consequences of system development and computer use" (Clemmensen 2006: 145). To understand how to design social systems, it is necessary to understand social dynamics and how groups affect users' thoughts and behavior (Porter 2008; Goldstone and Gureckis 2009; Consolvo et al. 2009).

An increasing amount of social psychology-based research on online communities is being published. Some of the first researchers to explicitly apply social psychology theories to online community settings were Beenen et al. (2004). They argued that "unlike theories in cognitive psychology, this [social psychology] theoretical base has been inadequately mined in the HCI and CSCW literatures" (Beenen et al. 2004: 220). Applying social psychology to the social web is currently a growing research area (Ma and Agarwal 2007), but there are still calls for more social psychology-based research (Li and Chignell 2010; Kraut and Resnick 2012). Furthermore, there is a need for more thorough theory application as:

> *"Attitudinal theories from social psychology have been quite extensively applied to the study of user intentions and behavior, [but] these theories have basically provided checklists or rules of thumb rather than systematic design methods or methodologies to develop software solutions."*
>
> (Kukkonen and Harjumaa 2008: 165)

---

valid in online settings (Zadro, Williams and Richardson 2004; Beenen et al. 2005; Zanbaka et al. 2007; Jonson 2008; Reis and Gosling 2010; Toma 2010).

As this chapter shows, much research has since been conducted in this field. A body of social psychology founded online community research already exists, making it a worthwhile area to contribute to. All this work is conducted in non-governmental contexts, leaving a gap to be filled in terms of open government communities. Moreover, neither the magnitude nor the consistency of the results across studies has been examined. A systematic review of the literature can facilitate accumulation of scientific knowledge.

Social psychology theory has the potential to improve participation in open government communities. But along this potential comes certain challenges: the field is fragmented and it is difficult to navigate among many different mid-range theories (Ross, Lepper, and Ward 2010). The *Handbook of Theories of Social Psychology* (Kruglanski et al. 2011) lists over 50 different social psychology theories. Most of the theories have the same aim: study thoughts and behavior in social contexts (Ross, Lepper, and Ward 2010). On the plus side, this means there is a sizeable and active community around the social psychology theories. The challenge of such a big community with many mid-level theories is that a lot of filtering is required when using the theories.

Social psychology is not just fragmented across different mid-level theories but also by location. The American social psychology community focuses on the individual level while the European community is more preoccupied with the societal level and contextual factors (Smith 2005; Chryssochoou 2010). They still aim to explain the same phenomena, but weigh factors differently. The most important difference might be in the applied methodology: Europeans often seek to get a broad societal dimension though surveys whereas Americans seek to examine the behavior of a smaller group through experimentation (Chryssochoou 2010). As described further in Chapter 5, this dissertation relies on experimental methodology. This work is therefore closer to the American social psychology community than to the European.

4.2.4. Social Science Discipline Selection

The five social science traditions outlined above have the potential to explain social behavior and thereby help understand how to increase open government participation (Kraut & Resnick 2012). The four perspectives also have some limitations – some more severe than others. In sum, behavioral economics, political science, sociology, and social psychology all have their advantages when seeking to understand and increase participation in open government communities. Out of the four social science sub-disciplines, social psychology is chosen as the theoretical lens for this dissertation. This choice is made due to the sub-discipline's good granularity match and methodology match with the subject matter. Moreover, the theoretical body of social psychology is inadequately mined (op cit.), thereby making it fruitful to leverage it.

4.3. Literature Review: The Social Psychology of Online Communities[111]

The following literature review covers current research that uses social psychology theory to study online communities. The scope of this review is limited to research relevant to the research question asked in this dissertation.

Research into online communities has been conducted since the 1990s (Preece, Maloney-Krichmar, and Abras 2003). Early research was mainly empirical descriptions of early online communities, and theory-based research was a rare sight before the turn of the millennium. As shown in Figure 12 below, most of the included papers are publicized in the second half of the 2000s. Most of these studies examine real online communities. It is therefore logical that research is lacking behind practice, as the existence of online communities is a prerequisite for studying them. Therefore, the time period included in this review is limited to 2000-2010.

---

[111] This part of the chapter is currently pending peer review in *First Monday*.

4.3.1. Scope and Selection Strategy

This literature covers research that fulfills two basic criteria: *First*, the paper must include one or more social psychology theory. *Secondly*, the social psychology theory must be applied to an online community. In short, this review is concerned with the intersection of social psychology and online communities. Pure theory-building papers are excluded and so are empirical studies not relying on a social psychology theory.

A number of variables are coded for each article included:

- Bibliometrics (publication year and publication outlet)
- Theory
- Research Design
- Methodology
- Number of subjects (both n and N)
- Case

A full summary of these variables is available in Appendix 1.

There are two methods often used to perform literature reviews: journal-centric and concept-centric (Larsson and Hrastinski 2011). As social psychology-based research into online communities is quite multidisciplinary, it is hard to rely on a journal-centric approach, as research is published in very different fields, let alone journals. This literature review relies on a two-sponged approach, as recommended by Webster and Watson (2002). It begins with a journal-centric approach to identify work from the top IS journals and subsequently expands to a concept-centric approach based on searches in Web of Knowledge. The journal-centric part of this review is conducted by scanning table of contents in leading IS journals, and the concept-centric part is conducted by searching Web of Knowledge for relevant keywords. Please refer to Appendix 2 for the literature review process.

The selection process has a number of limitations. *Firstly*, the Web of Knowledge is not a complete database of all research. If other databases such as ProQuest or Google Scholar were used the results probably would have varied. *Secondly*, only

articles written in English are included. This might be problematic as social psychology research points to some potential differences between behavior in collective and individualistic cultures (Heine 2010). This limitation is mitigated by the fact that several included articles are conducted on communities in Asia (e.g. Chou et al. 2010; Shin and Kim 2010). *Thirdly*, only one researcher has conducted this review. This might lead to subjectivity in selection and classification. *Fourthly*, the vast majority of the papers retrieved through the Web of Knowledge search do not meet the criteria of containing both social psychology theory and an empirical test.[112] The review therefore has a low inclusion rate.[113]

4.3.2. Literature Overview

Whereas many literature reviews only quantify the research into different categories, a more qualitative approach is used here. This qualitative approach is deemed appropriate due to the high degree of diversity and fragmentation of the field. As many different theories and methodologies are being used, a short description of each paper is included in the review. Before getting to these qualitative descriptions, a quantitative overview is provided.

---

[112] Please refer to Appendix 2 for a complete breakdown of numbers.

[113] The excluded research focuses on many related topics. For example, some papers are concerned with identity and social dynamics in an online community (e.g. Perotta 2006; Szell and Turner 2010). However, these papers are not design-oriented, but merely concerned with social psychological dynamics. Another group of papers excluded from this review are papers examining online e-learning and learning communities (e.g. Tseng and Kupo 2010). This research is usually not relying on social psychology. Yet another strand of literature is concerned with the individual consequences of online participation, e.g. participating in online health forums, weight loss groups, etc. (e.g. Hwang et al. 2010). Another research area is knowledge *contribution in closed organizational settings*. This research is concerned with employee participation rather than volunteer participation (e.g. Kankanhalli, Tan and Wei 2005). Although there are many similarities between participation to volunteer communities and organizational communities. the latter are excluded from this paper is different as people act for different reasons at work and in their spare time. Finally, motivation to use websites without online communities is also being studied in much of the research (e.g. Sangwan, Siguaw and Guan 2010). As this research is not concerned with any social aspect of the usage, it is excluded from this review.

Figure 12 shows the year of publication of the included articles. It is worth noting that the field is growing rapidly, and that almost half the included research was published in 2010. One reasonable explanation for this growth is the high growth rate that social networks such as Facebook and Twitter have experienced at the end of the decade. This makes it an interesting and important topic for researchers. As noted above, research usually lack behind practice.
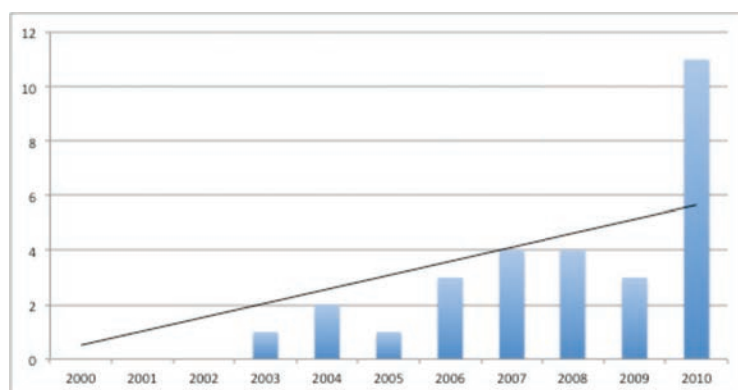
Figure 12: Number of papers published between 2000-2010 included in the review

4.3.2.1. Research Designs and Methodologies

Four different research designs are used in the included papers. The most frequently used design is a single case study, used in 16 papers. It is followed by comparative case studies, used in seven studies. Four papers are used to further theory development, and one paper uses longitudinal design. This means that the research community primarily is concerned with the short-term effect of behavior in online communities.

Research Design	Number Of Papers
Single case study	16
Comparative case study	7
Theory building	4
Longitudinal study	1
Total	28

Table 8: Applied research design in the included papers. See Appendix 1 for more details

Although this research strand is multidisciplinary and diverse, the data source is not. As Table 9 shows, 28 of the papers rely solely on quantitative data, and two papers combine quantitative and qualitative data. None of the included papers use qualitative data exclusively. One explanation for this is that the included papers often use contribution level as the dependent variable. Another explanation is that social psychology theory often is operationalized quantitatively.[114]

Data Type	Number Of Papers
Quantitative	26
Qualitative	0
Mixed	2

Table 9: Data type used in the included papers

Table 10 shows that the research community's preferred methodology is surveys. 19 of the included papers rely wholly or partly on surveys. Nine papers use field experiments. Data log analysis is used in five papers. Netnography, semi-structured interviews, and evaluation are each used once. Although the categories are not mutually exclusive the use of several methods in one study is rare in the research.

---

[114] It is worth noting that many studies (e.g. Li and Chignell 2010; Reich 2010) use qualitative methodology to study online communities. But these papers are not based on social psychology, and thus not included in this review.

Methodology	Number Of Papers
Survey	19
Field experiment	9
Data log analysis	5
Netnography	1
Evaluation	1
Semi-structured interviews	1
Total	36

Table 10: Applied methodology in the included papers. Categories are not mutually exclusive

When more than one methodology is applied, it is usually a combination of field experiment and survey. These two methodologies are combined in three papers.

4.3.2.2. Examined Cases

In the 28 papers included, 34 different cases and six undisclosed cases are examined, cf. Table 11 below. This shows that the research is scattered across many different online communities. Only two cases, Facebook and MovieLens, are appearing in more than one study. This provides a need to validate the research already conducted. MovieLens is by far the most studied online community in the research. MovieLens is a movie recommendation service created by University of Minnesota. Its origin means that researchers have access to data and can conduct experiments in the community without permission from third party owners.

Case	Number Of Papers
Flickr	1
Facebook	2
Quitnet	1
Lexus IS3000	1
Electronic Engineering Times	1
Everything2	1
IBM customer support community	1
Microsoft customer support community	1
Baidu	1
Yahoo! Knowledge+	1
MovieLens	7
Naver	1
Daum	1
Yahoo	1
http://www.es.lastminute.com/	1
http://www.minube.com	1
http://www. losviajeros.com	1
http://comunidad.muchoviaje.com/CS/	1
http://www.turismo20.com/	1
http://www.foro.geoplaneta.com	1
http://www. travelmarketing.biz/	1
http://www.escapadarural.com/	1
http://www.ruralon.com	1
http://www.viajaris.com	1
http://www.es.ulises.com	1
http://www.viajered.com	1
eetimes.com China	1
eetimes.com Taiwan	1
netscape.public.mozilla.ui	1
Newsgroup: alt. support.diet	1
Newsgroup: alt.support.cancer.breast	1
Newsgroup: alt.politics.usa.constitution.gun-rights	1
Newsgroup: alt. sports.hockey.nhl.ny-rangers	1
Newsgroup: alt.baldspot	1
Undisclosed	6

Table 11: Cases used in the included papers (34 named and six undisclosed cases)

As noted in Section 4.2.3.5, there are different approaches to social psychology in Europe and the US. Table 12 lists the included articles and the publication outlet. The table shows that the included literature is quite US-centric, as the vast ma-

jority of the papers are published in American journals and conference proceeding. Only real exception to this is the work of Utz (2003), which is published in Swiss Journal Of Psychology. Thus, this dissertation falls well within the American approach to social psychology.

Another interesting insight from Table 12 is that the majority of the included literature is published in the HCI community. Of the 28 articles, 16 are published in HCI journals or conference proceedings.

Article	Publication
Cosley et al. (2003)	ACM CHI Conference
Utz (2003)	Swiss Journal Of Psychology
Ludford et al. (2004)	ACM CHI Conference
Wang and Fesenmaier (2004)	Tourism Management
Beenen et al. (2005)	Journal Of Computer Mediated Communication
Cosley et al. (2006)	ACM CHI Conference
Joyce and Kraut (2006)	Journal Of Computer-Mediated Communication
Rashid et al. (2006)	ACM CHI Conference
Harper et al. (2007)	Int. Conference On Intelligent User Interfaces Poceedings
Ma and Agarwal (2007)	Information Systems Research
Chan and Li (2008)	Journal Of Business Research
Drenner et al. (2008)	ACM RecSys Conference
Nov, Naaman and Ye (2008)	ACM CHI Conference
Lin (2008)	Information & Management
Burke, Marlow and Lento (2009)	ACM CHI Conference
Shen and Chiou (2009)	Internet Research
Shen and Khalifa (2009)	AIS Transactions On Human-Computer Interaction
Carmagnola, Vernero, and Grillo (2009)	User Modeling, Adaptation, And Personalization, Proceedings
Nambisa and Baron (2009)	Organization Science
Yang and Lai (2010)	Computers In Human Behavior
Shin and Kim (2010)	Journal Of Information Science
Chen et al. (2010)	The American Economic Review
Lin (2010)	International Journal Of Human–Computer Interaction
Shen, Yu, and Khalifa (2010)	Behaviour And Information Technology
Lampe et al. (2010)	ACM CHI Conference
Chou (2010)	Online Information Review
Chou et al. (2010)	Behaviour & Information Technology
Casaló, Flavián, and Guinalíu (2010)	Tourism Management

Table 12: List of the 28 articles and publication included in the review

4.3.2.3. Unit of Analysis

As social psychology is an interdisciplinary and fragmented field, there is no consensus on the unit of analysis. Some scholars emphasize the *psychology* part and take a reductionist stance, while others emphasize the *social* part and take a holistic stance (Sarup 1975). Consequently, there is no single unit of analysis when it comes to research on online communities. In the research reviewed here the unit of analysis is a social artifact: the online community. What is being studied,

and sometimes manipulated, is the online community per se. However, the over-arching goal is to manipulate the behavior of the community members and thereby improve the community.

It is sensible to use a community, rather than the individual, as the unit of analysis because the aggregate behavior is more useful than the behavior of the individual group member (Hsieh, Kraut and Hudson 2010). A new design feature in an information system will have different impact on different people. To say something meaningful about the impact, it has to be studied in aggregate.[115]

4.3.2.4. Theoretical Measurements

One important inclusion criterion is the usage of social psychology theory. One common trait among the papers is that theory is operationalized into different levels of the independent variable. After the initial operationalization, the theory is rarely used. It is rarely measured to which degree the operationalization *per se* works. Contrary to the reference disciplines themselves, structured scales are not used to measure to which degree the theories actually work. Such an omission would be a critical error when researching within the reference disciplines. But in the online community research, the primary purpose is to examine online communities, not verify or falsify the theories *per se*.

4.3.2.5. Applied Theories

As illustrated in Table 13, the structure of the review is concept-centric. The research included in this review is classified on the basis of social psychology theory. The theories included are: social comparison theory, social learning theory; cognitive dissonance theory; self-concept theory; need to belong; the collective effort model; reciprocity theory; group conformity theory; goal-setting theory,

---

[115] In other words, the point of interest in this research reviewed here is not why John Smith writes a review of a book he bought on Amazon, but rather why millions of people in aggregate write reviews on Amazon. This interest in aggregate behavior explains why the majority of the research is quantitative with large samples.

social presence theory; perceived identity verification theory; and common identity and bond theory.

Theory	Origin	Origin's Web of Knowledge Citations[116]	Origin's Google Scholar Citations	Used In Included Papers
Social Comparison Theory	Festinger (1954)	n/a	9,326	Chen et al. (2009) Carmagnola, Vernero, and Grillo (2009)
Social Learning Theory	Bandura (1962) Bandura (1977)	n/a	661	Burke, Marlow and Lento (2009) Lin (2010)
Cognitive Dissonance Theory	Festinger (1957)	n/a	20,275	Drenner et al. (2008)
Social Identity Theory	Tajfel and Turner (1979)	n/a	6,475	Utz (2003) Shen and Chiou (2009) Lampe et al. (2010) Nambisa and Baron (2010) Shin and Kim (2010)
Self-Efficacy Theory	Bandura (1977)	8,914	22,523	Wang and Fesenmaier (2004)
Need to Belong Theory	Baumeister and Leary (1995)	2,164	5,233	Lin (2008) Burke, Marlow and Lento (2009)
Collective Effort Model	Karau and Williams (1993)	393	956	Beenen et al. (2004) Ludford et al. (2004) Cosley et al. (2006) Rashid et al. (2006) Harper et al. (2007) Faridani et al. (2010)
Reciprocity Theory	Gouldner (1960)	2,275	5,717	Joyce and Kraut (2006) Olivera, Goodman and Tan (2008) Burke, Marlow and Lento (2009) Chan and Li (2010)

[116] Many of the original works are published at books, and therefore not available in Web of Knowledge. To compensate for this, Google Scolar is also used.

Goal-Setting Theory	Locke (1968)	n/a	1,537	Beenen et al. (2005)
			5,199	
Group Conformity Theory	Asch (1956)	193	2,591	Cosley et al. (2003)
Social Presence Theory	Short, Williams, and Christie (1976)	n/a	3,416	Nov, Naaman and Ye (2008)
				Shen and Khalifa (2009)
Perceived Identity Verification Theory	Ma and Agarwal (2007)	63	238	Ma and Agarwal (2007)
				Chou (2010)
				Chou et al. (2010)

Table 13: Theory-centric classification of papers included in the review. Some papers use several theories, and are categorized multiple times. The articles assigned n/a in Web of Knowledge citations cannot be found in the database.

147

4.3.3. Social Comparison Theory

In the original theory, Festinger (1954) argues that people look to others who they consider to be better off for guidance and learning, and compare themselves to others who are worse off in order to feel good about themselves. Although some studies disputes these findings, others find that comparing yourself to more distinguished others can in fact serve as an inspiration and motivation. (Kruglanski and Mayseless 1990; Lockwood et al. 1997; Suls, Martin and Wheeler 2002). The effect is greatest when people compare themselves to others with similar perceived and actual status (Festinger 1954; Fogg 2003).

Fogg (2003) contends that people compare themselves to others when evaluating their capabilities, opinions, and actions people compare. An objective yardstick is rarely available, leaving people with little choice other than to compare to other people (Festinger 1954). Thus "social comparison theory suggests that if many of your friends contribute content, you would want to 'keep up with the Joneses' and contribute more as well" (Burke, Marlow, and Lento 2009: 951).

Chen et al. (2010) use social comparison theory to understand users' motivation and increase their contributions to MovieLens.[117] The authors conduct a field experiment (n = 398, N = 109,366) combined with a pre-experiment and post-experiment survey. The control group subjects receive information about their own past behavior, while the subjects in the two experimental groups receive comparative behavioral information. The effect is convergence around the median level of contribution with a 530 % increase in participation from users below the median, and a 62 % decrease by users above the median. Based on this single case experiment, the study finds that when knowing how the median community member performs, users tend to aim for the same participation level.

Carmagnola, Vernero, and Grillo (2009) use social comparison theory and group conformity theory to develop a social recommendation algorithm. Based on social comparison theory, the authors "hypothesize that individuals become inter-

---

[117] http://movielens.org

ested in topics or subjects that do not necessarily match their personal pre-existing preferences and tastes, but that reflect those of the network" (Carmagnola, Vernero, and Grillo 2009: 226). To test this hypothesis, the authors develop a recommender algorithm based on the network's preferences. An evaluation study is used to test the algorithm. In this study, the researchers recommend Facebook groups to the subjects included in the study (n = 45, N = N/A) to find out if the groups are relevant. Through this single case design, the authors find the algorithm to be satisfactory in terms of precision and accuracy, and inadequate in terms of recall.

4.3.4. Social Learning Theory

Social learning theory (Bandura 1962; Bandura 1977)[118] is concerned with behavior modeling. The basic axiom in the theory is that people are influenced by what they learn when observing social interactions and their consequences (Fogg 2002; Burke et al. 2009). Through mechanisms such as attention, retention, reproduction, and motivation, it is possible to manipulate motivation and change behavior (Bandura 1977). Like the mechanism in social comparison theory, people observe and learn the most "when behavior is modeled by others who are similar to themselves but somewhat older or more experienced" (Fogg 2002: 201). Bishop (2007) explains that "lurkers can be persuaded to change such a belief [that they are being ignored] if they see that [other] novices have their posts responded to in a constructive way" (Bishop 2007: 1888).

Burke, Marlow, and Lento (2009) conduct a single case study on Facebook to see how social learning, feedback, and distribution affect newcomer contributions to Facebook.[119] Social learning theory is operationalized by exposing users to other

---

[118] Social learning theory is sometimes referred to as social cognitive theory (Fogg 2003). Social cognitive theory can be defined as a theory "examining the reasons why individuals adopt certain behaviors. Particularly, SCT [social cognitive theory] explains psychological functioning in terms of triadic reciprocal causation in which behavior, personal, and environmental factors operate as interacting determinants to individuals' behavior" (C. P. Lin 2010: 346).

[119] http://facebook.com

users' desired behavior. Based on a field experiment on (n = 140,292, N = 150,000,000) newcomers to Facebook, the researchers find that when new users are exposed to a desired behavior displayed by their friends it "modestly impacts the newcomers' eventual sharing" (Burke, Marlow and Lento 2009: 950). The experiment uses log-analysis and semi-structured interviews as methodology.

Lin (2010) uses social learning theory[120] to study user loyalty in an interactive online gaming community. The aim of the paper is to further theory building by testing several constructs within a model. Based on a survey (n = 303, N = n/a), Lin finds support for two constructs within the model: affective commitment and social norms. Two other constructs, exchange ideology and social support, are not supported by the survey results.

4.3.5. Cognitive Dissonance Theory

Cognitive dissonance theory (Festinger 1957) is considered to be "one of social psychology's grandest theories" (Elliot and Devine 1994), and one of the most studied phenomena in psychology and social psychology (Egan et al. 2007). Cognitive dissonance "describes a psychological state in which an individual's cognitions - beliefs, attitudes, and behaviors - are at odds" (Egan et al. 2007: 978). The feeling is undesirable, and people feeling dissonance between discrepant cognitions, are motivated to reduce it (Jarcho et al. 2010).

Consistency over time is perceived as a good, and sometimes even a necessity. Consistency and consonance is valued in most cultures because it leads to predictability, trust, and rationality (Cialdini 2007). Therefore, individuals strive to reduce dissonance between two or more incompatible cognitions, and thereby increase consistency over time (Deutsch and Gerard 1955; Festinger 1957; Nowak, Vallacher and Miller 2003). Whenever a choice is made or a stand taken,

---

[120] The author refers to the theory as social cognitive theory, but the term is used interchangeably with social learning theory.

"we will encounter personal and interpersonal pressures to behave consistently with that commitment" (Cialdini 2007: 57).

Drenner et al. (2008) use cognitive dissonance theory in a field experiment that might seem counter-intuitive to most behavioral economics: make the barrier of entry for new users *higher*. The reasoning behind this approach is that "Research on 'cognitive dissonance'… predicts that requiring more work will make users, who do complete the entry barrier, value MovieLens more, and thus (we hope) contribute more" (Drenner et al. 2008: 191). Through a single case field experiment on MovieLens, the researchers use number of added tags as the dependent variable.[121]

The results of the study (n = 583, N = n/a) show that the completion rate for sign up process decreased when the users were asked to do more work. However, the number of tags increased dramatically - especially for the group that has to do the most work before becoming members. It is also worth noting that the tag quality was not reduced. More importantly, the high entry barrier has long-term consequences for behavior as 24.7% of the high-barrier variation entered tags during the post-entry phase, while only 5.6% of the control group did.

4.3.6. Self –Efficacy Theory

People like to have an effect on their environment (Bandura 1977). This is the basic premise of self-efficacy theory,[122] which states that people are more motivated to perform certain behaviors, for example share their knowledge, when they know these actions have an effect on other people.

---

[121] The variations in the experiment are: (a) current entry (control group): rate 5 movies; (b) Variation 1: rate 5 movies + view tags page; (c) Variation 2: rate 5 movies + add 5 tags; and (d) Variation 3: rate 5 movies + add 25 tags. TAG25 users entered an average of 18.47 tags while TAG0 users entered an average of 3.67 tags.

[122] Self-concept theory is an umbrella-theory including three sub-theories: social identity theory, self-presentation theory, and self-efficacy theory. Self-concept theory argues that humans are motivated to present and reaffirm their identities in everyday practices (Ma and Agarwal 2007).

In a single case study examining knowledge sharing behavior at Wikipedia, Yang and Lai (2010) use self-efficacy theory,[123] external self-concept theory, and intrinsic/extrinsic motivation factors in a model. Based on a survey (n = 219, N = n/a), the authors find a high positive correlation between self-efficacy and sharing behavior in Wikipedia. Thus, users with higher self-efficacy tend to contribute more.

Examining the reasons behind contributions to a large, undisclosed, travel community, Wang and Fesenmaier (2004) use self-efficacy as a construct. Self-efficacy is used in a multi-construct model also including instrumental and expectancy constructs. This model is tested through a single case study using a survey (n = 332 N = 150,000) with respondents from the community. The scholars find self-efficacy to be the construct best correlated with contribution level in the community.

4.3.7. Social Identity Theory

Social identity theory (Tajfel and Turner 1979) is concerned with the relationship between the individual and the group. To examine this relationship, the theory relies on the individual's self-concept, and examines how it is influenced by group identity (Hogg 2006). According to social identity theory, a person's self is partly developed by the groups he belongs to and identify with (Hogg and Terry 2000). Thus, when people identify with a given group, they will also change their self-concept and act accordingly.

Utz (2003) uses social identity theory and the overarching self-concept theory in a single case study of an undisclosed text-based multi-user dungeon. The aim of the research is to examine if users primarily are attracted to other *people* in the group (interpersonal relationships) or to the *group* per se (group identity). Based on a survey and field experiment (n = 237, N = n/a), the author concludes that both types of relationships exist although independently. Moreover, group

---

[123] The authors refer to the theory as internal self-concept motivation, but the meaning is the same.

identity exists independently of the member's time in the group, while interpersonal relationships are dependent on time invested in the group as these are formed over a long period of time.

Shen and Chiou (2009) examine why users are loyal to the same blog service for years. Based on social identity theory, the researchers hypothesize a positive correlation between community identification and loyalty. Based on a survey among (n = 243, N = n/a) Taiwanese bloggers, they conclude "higher community identification intensified greater consumer's willingness to invest in specific assets on a blogging community … and intention toward using the blogging community" (Shen and Chiou 2009: 403). Thus, an intergroup identity is correlated with loyalty and participation intention.

In a comparative case study comprising two company-hosted software support communities, Microsoft[124] and IBM[125], Nambisan and Baron (2009) examine two types of contributions: (a) contributions to the community and (b) contributions to the company. To examine the different reasons to contribute, the researchers draw on four theories: Social capital theory, social exchange theory, involvement, and social identity theory. Based on log data measuring actual contributions and a survey (n = 152, N = n/a) with 121 from Microsoft community and 31 from IBM community, the researchers find that users contribute to communities and companies for different reasons. Identification matters relatively more in community participation, while a sense of partnership and reward-expectancy matters relatively more regarding company contributions. The authors recommend "that members' identity with different sets of stakeholders should be taken into consideration to understand the impact of social identity in online communities…" (Nambisan and Baron 2009: 567). This entails that identity is more nuanced than sometimes assumed in the literature.

---

[124] http://microsoft.com

[125] http://ibm.com

153

Studying knowledge contributions in three online communities, Naver, [126] Daum,[127] and Yahoo!,[128] Shin and Kim (2010) use social identity theory along with three other theories: organizational citizenship behavior theory, cohesiveness, and affection similarity. Based on a survey (n = 192, N = n/a), the researchers find that organizational citizenship behavior theory, affection similarity and social identity theory are all positively correlated with knowledge contributions to the three online communities. Thus, the comparative case study shows that social identity and knowledge contributions are positively correlated.

In a single case study examining a community evolving around write-ups, Everything2,[129] Lampe et al. (2010) use three different theories to examine motivations to contribute: uses and gratification theory, social identity theory, and organizational commitment theory. Social identity theory is turned into two constructs: evaluative social identity and cognitive social identity. By studying actual behavior through log files and surveying 295 anonymous users and 304 registered users (n = 599, N = 1,850,000), the researchers find a correlation between the evaluative social identity construct and participation. However, they find no correlation between the cognitive social identity construct and participation.

In a comparative case study including 12 different Spanish company hosted online travel communities,[130] Casaló, Flavián, and Guinalíu (2010) use a mix of technology acceptance model, theory of planned behavior, and social identity theory to examine the intention to participate. Based on a survey (n = 456, N = n/a), the researchers find the three theories to be a reasonable predictor of in-

---

[126] http://naver.com

[127] http://daum.com

[128] http://yahoo.com

[129] http://everything2.com

[130] http://www.es.lastminute.com/, http://www.minube.com, http://www.losviajeros.com, http://comunidad.muchoviaje.com/CS/, http://www.turismo20.com/, http://www.foro.geoplaneta.com, http://www.travelmarketing.biz/, http://www.escapadarural.com/, http://www.ruralon.com, http://www.viajered.com, http://www.viajaris.com, http://www.es.ulises.com

tention to participate, explaining almost 60% of the subjects' intention to participate.

4.3.8. Need to Belong

According to Baumeister and Leary (1995), most people have a need to belong. This need causes people to form and maintain relations whenever an opportunity presents itself (Baumeister and Leary 1995). This need to belong applies both offline and online. For example, Zadro, Williams, and Richardson (2004) found that cyber-ostracism has the same emotional effects as ostracism. In other words, being ostracised by software triggers the same emotions as being ostracised by humans, although to a lesser degree.

In a theory building study aiming to expand the scope of the DeLone and McLean Model (DeLone and McLean 1992; Delone and McLean 2003) to encompass online communities, Lin (2008) includes sense of belonging as a central variable, similar to system satisfaction. The study examines "the impact of system characteristics (e.g., information and system quality) and social factors (e.g., trust and social usefulness) on success in virtual communities" (Lin 2008: 522). Based on a survey (n = 198, N = n/a) answered by students who have personal experiences with online communities, Lin finds a significant correlation between sense of belonging and member loyalty, as well as a positive correlation between member satisfaction and loyalty. The latter correlation is higher.

Furthermore, two studies placed in other categories in this review are partly relying on the need to belong. In research into online communities, the need to belong can be translated into feedback, as this is an expression of belonging. In their study on Facebook (op cit.), Burke, Marlow, and Lento (2009) find that getting feedback on one's contributions is well correlated with an increase in future contributions. Lampe et al. (2010) examine the motivation behind contributions to the everything2 community (op cit.). The researchers find that people with a "strong sense of belonging believed that they use the site more frequently, and

are more likely to both use and contribute to the site in the future" (Lampe et al. 2010: 1931).

4.3.9. Social Loafing Theory / Collective Effort Model

Social loafing theory (Karau and Williams 1993) suggests that many groups suffer from under-contributions because group members are contributing less when being part of a group vis-à-vis working alone. In a comprehensive review of the social loafing literature, Karau and Williams (1993) list five reasons for under-contributions in groups: (a) arousal reduction, (b) reduced evaluation potential about own work, (c) feeling that contributions are not essential to the group, (d) expects that other members are contributing well, and (e) reduced self-attention and therefore lower standards for performance and self-regulation. Based on these findings, Karau and Williams (1993) propose the Collective Effort Model that seeks to mitigate the causes for social loafing.

Ludford et al. (2004) seek to improve the level of contributions. The researchers rely on the collective effort model and seek to highlight the uniqueness of contributions. Based on a field experiment (n = 245, N = 8,500), the authors find a significant increase in the level of contribution from users who are in a group with dissimilar others *and* know the uniqueness of their contributions compared to the control group.

In a similar case study on MovieLens, Beenen et al. (2005) use the collective effort model and goal-setting theory to drive innovation in design. In a single case design, the researchers seek to improve the contribution level to MovieLens. To test the improvements, the researchers conduct two surveys and an email-based field experiment (n = 830, N = 80,000), where "different versions of an email message inviting existing MovieLens subscribers to rate movies [are sent]." (Beenen et al. 2005: 213). The researchers find that highlighting uniqueness in contributions is positively correlated to participation. No such correlation is established between benefit to self or group and participation.

Building on Ludford et al. (2004), Harper et al. (2007) aim to get more MovieLens users to contribute to the discussion boards by sending them invitations highlighting their uniqueness. Based on a single case field experiment (n = 2,021, N = 100,000), the researchers find a positive correlation between uniqueness and recruiting new members, although there is no correlation between uniqueness and contribution level post-recruitment.

Cosley et al. (2006) use the collected effort model in a field experiment (n = 2,723, N = n/a) on MovieLens. In a single case study, the researchers develop four algorithms that give users suggestions about movies to rate. The algorithm that selects relevant movies and highlights the distinctiveness in the user's contributions outperforms other algorithms by a high margin: "over 22% of Rare-Rated subjects edited at least one movie, compared to about 6% for the other groups" (Cosley et al. 2006: 1040). In other words, when a user knows his work is unique, he is much more likely to contribute.

Rashid et al. (2006) also rely on the collective effort model in a single case field experiment (n = 160, N = n/a) on MovieLens. In the experiment, four groups are created in which the value of a contribution is presented differently to the subjects: (1) value to self; (2) value to a small group the user has affinity with; (3) value to a small group the user does not have affinity with; and (4) value to the entire user community (Rashid et al. 2006). The results from this experiment demonstrate that showing a contribution is valuable to anyone increases contributions in general, and that showing that the contribution is valuable to a group with similar others has the largest positive impact on contributions.

4.3.10. Reciprocity Theory

One of the most important and pervasive mechanisms in society is that of reciprocity (Gouldner 1960). The rule of reciprocity states that we repay what others have provided us (Cialdini 2007). The rule of reciprocity is so fundamental to human behavior that basically all societies subscribe to the rule (Cialdini 2007). It influences human behavior in two ways: it drives actions as community mem-

bers expect other people to repay before they act, and it makes other people reciprocate actions already conducted.

Building on Gouldner's conception of reciprocity, Joyce and Kraut (2006) examine the role of feedback in six online communities: the Mozilla User Interface newsgroup (netscape.public.mozilla.ui) and five Usenet support groups: alt.baldspot., alt. support.diet, alt.support.cancer.breast, alt. sports.hockey.nhl. ny-rangers, and alt.politics.usa.constitution.gun-rights. Reciprocity theory predicts that a response to a comment will lead to further comments. To test this prediction, Joyce and Kraut use a longitudinal research design in which they collect six months worth of data for new users in each group (n = 2,777, N = n/a).

When receiving a reply to their first post, new users are much more likely to return for a second post. Likelihood increases from 44% - 54%. Moreover, getting basically any reply is better than getting no reply, as the researchers find no significant correlation between second contribution and the tone of the first reply (positive/negative) or length. In short, when a user receives a response to a post, he is more likely to post again.

Chan and Li (2008) examine OnlyLady,[131] a Chinese consumer community. In this single case study, reciprocity is used as the independent variable seeking to explain commitment to the online community and the level of co-shopping. Relying on a two-pronged methodology, netnography and a quantitative survey (n = 899, N = n/a), Chin and Li find a positive correlation between reciprocating behaviors and the commitment to the online community, as well as a positive correlation between reciprocating behaviors and co-shopping intentions.

4.3.11. Group Conformity Theory

Through a famous series of lab experiments, Asch (1956) revealed that most people conform to the group's opinion. In the lab experiment, the participants

---

[131] http://onlylady.com

were asked to compare the length of lines. The test was very simple, and the participants usually got the results right. When the subjects were confronted with a divergent – and obviously wrong – result from a peer group, most of them conformed to the group's judgment and reported a result that contradicted their own perception.

Based on a field experiment (n = 488, N = 70,000), Cosley et al. (2003) conclude that MovieLens users are likely to conform to group pressure. In this single case study, the authors find that users of MovieLens conform to the ratings suggested by the recommender system: "a recommender system can influence users to move from a negative to a positive rating" (Cosley et al. 2003: 590). That is, users who see that other users have rated a product positively are more likely to give a positive rating than they otherwise would be. Although group pressure can influence behavior, false and misleading predictions entail decreased satisfaction with the system.

4.3.12. Goal-Setting Theory

Goal-setting theory (Locke 1968) contends that goals affect action and effort level (Locke and Latham 2002). According to Bandura (1990), it is crucial that a goal is feasible in order for it to have a subsequent behavioral effect.

According to Beenen et al. (2005), "Abundant research since the 1960s shows that providing people with specific, high-challenge goals stimulates higher task performance than easy or 'do your best' goals" (Beenen et al. 2005: 216). For goal-setting theory to be applied to online communities, goals need to be defined for the community and community members.

Beenen et al. (2005) use goal-setting theory in a single case study on MovieLens. The researchers use emails to recruit subjects (n = 834, N = 80,000) who are then assigned to one of 10 groups based on the independent variables, group goals/individual goals and goal-specificity. The dependent variable is number of movies rated. The authors find that specific goals are better than do-your-best

goals but that too hard goals have a negative impact on performance. Furthermore, community goals are more beneficial than individual goals in terms of attaining increased productivity.[132]

4.3.13. Social Presence Theory

The basic premise of social presence theory (Short, Williams, and Christie 1976) is that the presence of others is influencing group members, whether or not this presence is actual, imagined, or implied (Allport 1968; Miranda and Saunders 2003; Nov, Naaman, and Ye 2008). Social presence is created when a "medium facilitates awareness of the other person and interpersonal relations during the interaction" (Fulk et al. 1990: 11).

Based on log data and a survey with Flickr users[133] (n = 237, N = n/a), Nov, Naaman, and Ye (2008) examine the effect of social presence on participation. In this single case study, the researchers find a significant positive correlation between perceived social presence and tagging participation. Highlighting social presence in online community design is thus expected to increase the level of participation.

Shen and Khalifa (2009) examine how social presence works in online settings. They introduce a multidimensional conception of social presence. The conception is tested in a comparative case study design. The applied methodology is a survey (n = 430, N= n/a) with members of four general interest communities in Hong Kong (Shen and Khalifa 2009). The researchers find that their conceptualization of social presence as an aggregate of three dimensions – awareness, affective social presence and cognitive social presence – is valid. This conceptualization can be used in further research to guide specific design.

---

[132] This is only significant at the p < 0.10 level. As explained in section 5.10.6.3, a p < .05 significance level is used in this dissertation.

[133] http://flickr.com

Shen, Yu, and Khalifa (2010) build on the research relying on social identity theory. Instead of examining how social identity theory affects participation, the researchers study how an online community can grow and strengthen identification through social presence. The authors argue that the two theories are logically connected, because one cannot identify with others without perceiving their social presence. Aiming to develop theory further, the researchers perform a data log analysis and a survey (n = 430, N = n/a). The authors find a significant positive correlation between social presence and social identity. Furthermore, a significant positive correlation between social identity and knowledge contribution is documented.

4.3.14. Perceived Identity Verification Theory

Social psychology is concerned with relationships between the individual and the group. An important aspect in this is identity. Taking their point of departure in Erickson (1968) and Swann (1983), Ma and Agarwal (2007) build the perceived identity verification theory defined as "the perceived confirmation from other community members of a focal person's belief about his identities" (Ma and Agarwal 2007: 46). Perceived identity verification is based on four constructs drawn from the social psychology literature: virtual co-presence, persistent labeling, self-presentation, and deep profiling. A positive correlation between these four constructs and knowledge contribution is hypothesized.

The authors of the perceived identity verification theory, Ma and Agarwal (2007), use the theory in a comparative case study. Based on a survey with community members from two online communities, Quitnet[134] and Lexus IS300 community[135](n = 666, N = 5,559), the authors find "strong empirical support for the proposed relationships" (Ma and Agarwal 2007: 58) between the four constructs in the theory.

---

[134] http://quitnet.com

[135] http://is300.net

Chou (2010) use perceived identity verification theory in a broad framework based on multiple theories and constructs including computer self-efficacy, computer anxiety, personal innovativeness in IT, and perceived identity verification. Through a comparative case study on Electronic Engineering Times[136] in China and Taiwan (n = 241, N = n/a), the researcher finds a positive correlation between perceived identity verification and knowledge contributions to the two online communities.

Chou et al. (2010) also apply perceived identity verification theory in a comparative study on Baidu[137] in China and Yahoo! knowledge+[138] in Taiwan. In their paper, the theory is used along with performance expectancy theory. Examining the continued intention of knowledge creation, the researchers find a correlation between perceived identity verification and continued intention of knowledge creation. They find no correlation between perceived identity verification and satisfaction with the community. The results are based on a survey (n = 429).

4.4. Theory Selection and Operationalization

The literature review above indicates that many theories can be used to examine the area of open government communities and improve participation. The challenge is not so much identifying *a* relevant theory but rather setting up criteria for finding *the most* relevant theories for the four experiments in this dissertation. Most of the theories described thus far in the chapter are highly relevant to the research question of this dissertation.

The social psychology theories described above are explanatory and predictive.[139] But in this dissertation, the theories are used for design and action. For example, social comparison theory is usually concerned with explaining how

---

[136] http://eetimes.com

[137] http://zhidao.baidu.com

[138] http://tw.knowledge.yahoo.com

[139] Besides being explanatory and predictive, theory can be analytical, predictive, or aimed at design and action (Gregor 2006).

people compare themselves to others for guidance. But in this dissertation the theory is translated into design hypothesis seeking to increase participation in an open government community.

4.4.1. Selecting Social Psychology Theories

As described above in Section 4.2.3 this dissertation follows Flyvbjerg's (2001) recommendation to use theories relevant to the specific subject matter rather than (presumed) universally applicable theories.

As in Section 4.2.3, a classification will serve as a helpful heuristics to make theory selection more systematical.[140] To the extent of the author's knowledge, no canonical social psychology theory classification exists. In order to select theories for the four experiments conducted in this dissertation, a classification of the 12 aforementioned theories is suggested in Table 14. This evaluation is not of the theories per se, but of the theories in the context of this dissertation's research question. The theories are evaluated on the basis of the literature reviewed in section 4.3.[141] Thus, the assessment and interpretation made below is conducted with the research question and literature review in mind, and is not a general, context-independent assessment.

As described above, Arazy, Kumar, and Shapira (2010) list potential challenges when using social science theory in information science: (a) scope; (b) effect size, (c) relevancy to the design task, and (d) level of abstraction. (a) Scope and (b) effect size are not particularly problematic in terms of the research question asked here (op cit.). All the social psychology theories reviewed here have a scope that matches online community studies. And the effect size can be calculated for the theoretical constructs. However, (c) relevancy to the design task

---

[140] As argued by Wagner (1963), "It cannot be claimed that any classificatory system, as such, is more 'logical' and more 'systematic' than another. The decisive question is whether it is heuristically more useful." (Wagner 1963: 736).

[141] If the theories are evaluated in terms of other research question and literature they are likely to be evaluated differently.

and (d) level of abstraction match are important criteria when seeking to answer the research question. [142]

	Poor Case Fit	Good Case Fit
Poor level of abstraction match	- Collective effort model - Need to belong - Group conformity	- Reciprocity - Social presence - Social learning
Good level of abstraction match	- Cognitive dissonance - Perceived identity verification	- Social comparison - Goal-setting - Self-efficacy theory - Social identity

Table 14: Theory classification based on case fit and level of abstraction match

As Table 14 shows, relevancy to the design task and level of abstraction match are therefore guiding the theory choice.[143]

Four theories are categorized as both a good case fit and a good match in term of level of abstraction: social comparison theory, goal-setting theory, self-efficacy theory, and social identity theory. Consequently, four theories are used to guide the empirical field experiments in this dissertation. This complements the intellectual style of social psychology, which is "to build and test a large number of medium-level theories, each attempting to account for an interesting social phenomenon in a limited domain" (Kraut 2003, 329).[144]

The four included theories operationalize at the same level: they seek to explain aggregated individual behavior. This corresponds very well with the research question asked in this dissertation.

---

[142] The second criteria, case fit, is different than the criterion applied earlier in this chapter when choosing between different social science disciplines. Case fit replaces methodology fit because most social psychology theories rely on the same methodologies: experimentation and surveys.

[143] Using these criteria for theory selection entails that the case guides the selection, which is recommended by Flyvbjerg (2001). It is, however, worth noting that it is only one of many ways to select theories.

[144] For further details on theory of science and methodological considerations, please refer to chapter 5.

4.4.2. Social Comparison Theory for Experiment 1

The fundamental assumption in social comparison theory is simply: the relative absence of objective parameters for evaluating human behavior leads people look to other people for clues regarding appropriate social behavior and performance level (Festinger 1954). This serves as a way to gain self-awareness, which can be a source of motivation (Swann and Bosson 2010).[145]

Social comparison theory is well tested, and many empirical studies find it to be accurate (see Taylor and Lobel 1989; Suls and Wilis, 1991 for comprehensive literature reviews). The theory has proven useful in various contexts including management (Blanz et al. 1998), compensation (O'Reilly III, Main, and Crystal 1988), and advertisement (Richins 1995). As shown in the literature review above, social comparison theory has been used in two online community studies (Beenen et al. 2004; Chen et al. 2010) and zero open government community studies.

The first experiment in this dissertation helps verify the findings in Chen et al. (2010), but in a different context.[146] These studies indicate that the theory can be used to increase participation in non-governmental communities. Although the theory has yet to be used in open government communities, it is likely that the similar positive effects can be attained. Despite the scarce use of the theory in online communities, the theory is fairly accurate when applied.

There is a reasonably good match between the theory and available data sources at K10. Social comparisons are usually made through quantitative measure-

---

[145] People develop the cognitive ability to explicitly compare to others when they reach middle school. After that, it happens effortlessly and sometimes unintentionally (Swann and Bosson 2010).

[146] This experiment does not explicitly recruit users, but includes all existing K10 users. This study seeks to identify unintended consequences by providing social comparison information to regular users. It is worth noting that this increases the value of the experiment, but reduces the possibilities to compare the results to Chen et al. (2010).

ments, which is the most readily accessible form of data in this case. The includ-
ed variables and measurements are described further below.

4.4.2.1. Limitations

Like all other online communities, K10 comprises many complicated and inter-
dependent interactions, social cues, and incentives. Social comparison theory is
unlikely to explain all mechanisms in an online community. No middle-level the-
ory can do this.

Due to a divide in the social comparison literature, the consistency is only mod-
erate. In the original theory, Festinger (1954) argues that people compare to
others who are better off for guidance and learning, and compare to others who
are worse off in order to feel good about themselves. Several scholars find that
people also make upward comparisons in order to be inspired and motivated
(Kruglanski and Mayseless 1990; Lockwood et al. 1997; Suls, Martin and
Wheeler 2002).

Other scholars argue that people might not necessarily perform better after mak-
ing social comparisons. Stapel and Suls (2004) argue that social comparisons
"produce assimilation, that is, self-evaluations displaced toward the comparison
referent" (Stapel and Suls 2004: 860). According to this second strand of social
comparison theory, it is expected that users will regress toward the mean in
terms of contributions. Thus, the users who typically contribute a lot will con-
tribute less, while the users who usually do not contribute much will contribute
more. In an online community context, Chen et al. (2010) found a regression to-
ward the mean.

The point is that although this theory has been widely used, it is still not entirely
consistent as different interpretations exist. In the operationalization of this the-
ory, the second strand is used. This version of social comparison theory is pre-
ferred as the granularity is considered to be higher.

4.4.2.2. Operationalization of Social Comparison Theory

Social comparison theory can be operationalized through a research question and a number of hypotheses. The research question asked is: *How does exposure to social comparison information affect users' participation in K10?* Derived from social comparison theory, a null hypothesis and six alternative hypotheses are proposed in order to answer the research question.

H_0	Receiving social comparison information has *no* effect on subsequent participation.
H_a	Receiving social comparison information has an effect on subsequent participation
H_{a1}	Receiving social comparison information reduces the contribution level among above-means contributors
H_{a2}	Receiving social comparison information increases the contribution level among below-means contributors
H_b	Receiving non-comparative information has an effect on subsequent participation, but smaller than comparative information.
H_{b1}	Receiving non-comparative information increases the contribution level among above-means contributors
H_{b2}	Receiving non-comparative information increases the contribution level among below-means contributors

To operationalize social comparison theory into research, variables and measurements are defined. The selected variables and corresponding measurements are shown in Table 15.

Variables	Measurements
Social comparison yardstick	Number of posts written 30 days prior to the experiment
Behavior before and after the intervention	Number of posts written 14 days before and after the intervention

Table 15: Variables and measurements for social comparison theory

4.4.2.2.1. Variables

Social comparison theory is operationalized with two variables. The *first* variable is a yardstick created to enable social comparison. This yardstick is a quantitative metric that shows how much the subject contributes compared to the group mean. The *second* variable is the contribution level before and after the intervention.[147] This variable seeks to detect if the social comparison yardstick has an effect on the subjects' behavior.

4.4.2.2.2. Measurements

Group 1 receives social comparison information by showing the subject's participation during the previous 30 days compared to community mean of 6.3 posts. *Group 2* is exposed to non-comparative information by showing the subject's participation during the previous 30 days but without comparing it to other users' participation. *Group 3* receives no information at all.

As noted above, social comparison theory would predict a regression toward the mean: those who contribute a lot are expected to contribute less after the experimental intervention and those who contribute less before are expected to contribute more after. Therefore, *group 1* subjects are divided into two different groups: above and below mean. The mean that separates the above-mean contributors from below-mean contributors is 6.3 posts in the previous 30 days. Thus, all the subjects that write 1-6 posts before the intervention are categorized as below-mean contributors, and subjects that write 7 or more posts are categorized as above-mean contributors. This segmentation helps operationalize the social comparison theory and add further granularity to the study.

The dependent variable in this experiment is number of contributions. It is possible to make two kinds of contributions at K10: start a new thread and write a reply to an existing thread. No distinction is made between these two types of

---

[147] The first experiment is a split-plot design. This entails that there is a between-subject element and a within-subject element to the experiment.

contributions as they are treated as a single variable, posts, regardless of the character of the post.

This variable, number of posts, is measured through a database query made directly in K10's MySQL database. The query asks for posts per user in the period defined in the experiment. Thus, the measurement is number of posts per subject in the experimental time period. The dependent variable is measured through number of posts written at K10 in a 14-day period post-intervention.

Both the dependent and independent variable rely on quantitative metrics: number of posts the last 30 days. This choice is made for several reasons. It is a fairly objective unit of measurement. It is easy to compute, and it is easy for the subjects to understand. However, social comparison theory can be translated in many ways, and the translation used in this study might not match the users' most salient comparison metrics. They could, for example, be concerned with other quantitative metrics such as how many people read their replies, and/or how many users respond to their posts. It is also possible that they are more preoccupied with qualitative metrics such as the quality of the posts, how the feedback of their posts compare to the feedback given to other users etc.

4.4.3. Goal-Setting Theory for Experiment 2

Goal-setting theory (Locke 1968) asserts that goals tends to increase the motivation, effect, and performance for those who are subject to the goal (Locke and Latham 2002).[148] The effects of goal-setting are thoroughly documented and considered reliable (Locke and Latham 2002). According to Beenen et al. (2005), goal-setting theory is "among the most robust psychological findings on human motivation" (Beenen et al. 2005: 216) and considered one of the best cognitive determinants of behavior (Stajkovic and Luthans 1998).[149]

---

[148] In this context, a goal is defined as "the object or aim of an action… usually within a specified time limit" (Locke and Latham 2002: 705).

[149] Replication errors occur frequently, but they are usually caused by methodological errors (Locke and Latham 2002).

According to a meta-analysis performed by Locke and Latham (2002), goal-setting theory has been verified across many contexts and countries. Goal-setting theory is most widely examined in organizational and work-related environments. Other areas of application include education and health (Locke and Latham 2002). Despite the specific focus, the theory aims to account for motivational drives in all aspects of human function (Bargh, Gollwitzter, and Oettingen 2010). This means that the theory holds high external validity and has a broad scope.

As with most other social psychology theories, goal-setting theory was developed before open government communities. This means that the subject matter of this dissertation is outside the original theory scope. As shown in the literature review earlier in this chapter, goal-setting theory has been used in one online community study (Beenen et al. 2005). It is likely that similar positive effects can be attained in open government communities.

In most goal-setting studies the dependent variable is quantitative (Locke and Latham 2002). This corresponds well with the case of this dissertation, as it is more feasible to collect quantitative data than qualitative data on K10.[150]

4.4.3.1. Limitations

Despite the simplicity of the theory, it has many qualifications. Not all goals have the same level on action and effort as several factors are intervening (Bargh, Gollwitzter, and Oettingen 2010). Feedback, commitment type, goal complexity, availability, goal difficulty, and self-regulation are all important factors in successful goal-setting (Bargh, Gollwitzter, and Oettingen 2010). Add to this context variables such as affective state, competing action tendencies, and power position (ibid).

---

[150] Please refer to Chapter 5 for full case description.

Some of these factors are highly relevant to the experiments conducted here. The highest level of effort occurs when the goal is *moderately difficult* and still feasible (Locke and Latham 2002).[151] Moreover, a *specific goal* leads to increased effort vis-à-vis a vague/unspecific goal (Bargh, Gollwitzter, and Oettingen 2010). The reason is the reduced ambiguity regarding accepted levels of performance (Locke and Latham 2002).[152] Ongoing *feedback* on the progress toward goal accomplishments makes the goals more effective (Locke and Latham 2002). Hollenbeck, Williams, and Klein (1989) find that more people tend to set difficult goals when the goals are *private*, not public. The researchers also find that a sense of self-control over the goal increases the goal difficulty. According to Garner (1996) people become more reflective over the desirability and/or feasibility of a goal when they *actively define the goals* instead of doing it passively or implicitly. Goal-setting does not need to be *social*, but it often is because of the context in which the goals are set.[153] According to Ajzen (1985), normative expectations from a community and other social cues often influence people's goal selection. These qualities and factors make the theory's accuracy questionable. This makes it fruitful to use the theory in specific contexts to help identify boundary conditions and effect size.

As it is the case with social comparison theory, goal-setting theory is not a fully consistent and unified theory. The cognitive processes behind both goal-setting and goal-striving are disputed and still not fully understood (Bargh, Gollwitzer, and Oettingen 2010). Despite these objections, there is wide agreement that the right goals in the right context can lead to increased motivation and performance.

---

[151] The function of goal difficulty and effort is curvilinear.

[152] Beenen et al. (2004) explain that "Abundant research since the 1960s shows that providing people with specific, high-challenge goals stimulates higher task performance than easy or "do your best" goals" (Beenen et al. 2004: 216).

[153] Goal-setting theory makes a distinction between goals set by the individual herself and goal set by others on her behalf (Bargh et al. 2010). In this dissertation the operationalization of the theory is only concerned with the former: how self-assigned goal influences the goal-setter.

4.4.3.2. Operationalization of Goal-Setting Theory

Based on goal-setting theory, a research question and two hypotheses are formed. The research question is: *How does a self-assigned goal to a number of contributions to K10 affect subsequent participation?* In order to answer this re-search question, one null hypothesis and one alternative hypothesis are proposed.

H_0 Setting a participation goal has no effect on subsequent participation

H_a Setting a participation goal makes subjects strive to achieve the goal

To operationalize social comparison theory into research, three variables and corresponding measurements are defined. The selected variables and corresponding measurements are shown in Table 16.[154]

Variables	Measurements
Goal assignment	The subjects get to define their own goal in this experiment through a non-public survey.
Goal content	The goals for this experiment are limited to a quantitative metric. Each subject selects how many posts she will write the following two weeks.
Goal fulfillment	The number of posts contributed after the intervention is measured. By comparing this measurement number to the goal, it is possible to see if the goal has an effect.

Table 16: Variables and measurements for goal-setting theory

4.4.3.2.1. Variables

Operationalization of goal-setting theory entails three important variables. The *first* variable is the goal assignment. Goal-setting theory can be operationalized in different ways. At one end of the spectrum the individual assigns the goal herself. At the other end of the spectrum a third party assigns the goal (Bargh, Gollwitzter, and Oettingen 2010). The case used in this dissertation, K10, mainly

---

[154] In the second experiment, a split-plot design is used. Like the first experiment, this entails between-group variation and within-group variation. The independent variable comprises three levels: Group 1 subjects answer a survey containing a goal-setting component. Group 2 subjects answer a survey without a goal-setting component. Group 3 subjects answer no survey at all.

comprises people with fewer capabilities than the average population.[155] Due to a fragile health, it is difficult for others to assign goals that seem high yet realistic.[156] Therefore, the theory operationalization in this experiment relies on self-assigned goals. To maximize the intervention effect, and thus discover if goal-setting can impact participation in K10, the goal-setting is kept private by having the intervention in a survey.

The *second* variable is goal content. It is often easier to follow a goal if the goal is quantitative (Locke and Latham 2002). To reduce goal complexity, the goal is expressed as a number and thus not in qualitative terms.

The *third* variable is goal fulfillment. In order to know if the defined goal has an effect the post goal-setting behavior must be measured.

4.4.3.2.2. Measurements

The subjects set a goal for the number of posts they should write on K10 during the following 14 days. To make the subjects assign themselves a goal, they are asked the following question in a survey: *How many posts do you expect to write the next two weeks?* The response to this question is used as the subject's goal.

The number of posts written at K10 in the 14 days following the experimental intervention is the variable used to examine the different levels of the independent variable. This variable is measured the exact same way as in the first experiment, so please refer to section 4.4.2.2 for details on variable measurement.

The final variable, goal-fulfillment, is measured through number of posts written during the 14 days after the goal is set. If the experiment goes as expected, there the subjects' behavior will reflect the assigned goal.

---

[155] Please refer to Chapter 3 for a complete description of the demographics of the K10 community.

[156] It would be possible to set the goal 30% higher than their participation during the previous 14 days, but a pilot study indicates that the subjects do not like to be pressured into participation.

4.4.4. Self-Efficacy Theory for Experiment 3

According to self-efficacy theory (Bandura 1977), perceptions of capabilities matter and influence behavior (Bandura 1977; Pajares 2003). Self-efficacy is defined as "the conviction that one can successfully execute the behavior required to produce the outcomes" (Bandura 1977: 193). The underlying premise of self-efficacy theory is that people like to have an effect on their environment and are more motivated to do certain tasks when they know these actions have an effect on other people.

If one believes she has the chance to succeed in a specific situation, she will approach the task differently than she will if she does not hold such beliefs (Bandura 1977). Bandura (1977) writes that "expectations of personal mastery affect both the initiation and persistence of coping behavior" (Bandura 1977: 193). Thus, a task will be met with increased effort when the person has high self-efficacy. According to Bandura (1977) self-efficacy is "derived from four principal sources of information: performance accomplishments, vicarious experience, verbal persuasion, and physiological states" (Bandura 1977: 191). Of these four information sources, performance accomplishment is the most important (Pajares 2010).

Self-efficacy theory is applied broadly to quite different contexts including education (Schunk 1991), management (Stajkovic and Lufthans 1998), health (Bandura 1990), computers (Compeau and Higgins 1995), and online environments (Wang and Fesenmaier 2004).

As with most other social psychology theories, self-efficacy theory was developed before the creation of open government communities. Therefore, the subject matter of this dissertation might be outside the original scope. Online community researchers find self-efficacy to be an important aspect of contributions to online communities. Yang and Lai (2010) find a high positive correlation between self-efficacy and sharing behavior in Wikipedia. Wang and Fesenmaier (2004) find self-efficacy to be the best correlated construct out of many (op cit.) to explain the level of contribution in the community. However, both studies are

correlation studies and cannot make any claims about causality.[157] Using self-efficacy in a controlled experiment will help shed light on the causality.

4.4.4.1.1. Limitations

Self-efficacy is believed to be an important predictor of effort and motivation (Pajares 2003). According to Stajkovich and Luthans (1998), "only a few cognitive determinants of behavior (e.g. goal-setting) have received as amble and consistent empirical support as the concept of self-efficacy" (Stajkovich and Luthans 1998: 240). But like other middle-range theories, self-efficacy is only to be considered one of several important elements in behavior and motivation (Bandura 1997; Stajkovich and Luthans 1998).

Meta-analyses from several domains including education (Multon, Brown and Lent 1991), performance (Sadri and Robertsen 1993), behavior choice (Sadri and Robertsen 1993), and work-related settings (Stajkovich and Luthans 1998) find the theory to be valid. Despite this broad verification, Sadri and Robertsen (1993) note that the link between self-efficacy and performance appears to be stronger in field studies than in laboratory studies.

Measuring self-efficacy is often a comprehensive task and includes several standardized scales such as the General Self-Efficacy Scale (Chen, Gully and Eden 2001) and Computer Self-Efficacy Scale (Murphy, Coover, and Owen 1989). Having K10 subjects complete such comprehensive surveys is challenging due to the subjects' health issues.[158] As noted several times in this dissertation, the aim of this dissertation is not to verify theories but to use them. The goal is to see if some of the mechanisms that usually increase self-efficacy will increase the subjects' post-intervention participation. Therefore, the subjects' experienced self-efficacy is not measured in the experiment. This increases the risk of theory operationalization errors and reduces the claims that can be made about the theo-

---

[157] Online community members might experience higher self-efficacy because they contribute more to the community instead of the assumed reverse causality.

[158] Please refer to Chapter 3 for a full description of the demographics and the case.

ry's validity in the open government community domain. This caveat is discussed thoroughly in Chapter 5.

As with the first two experiments described above, the main data used is participation behavior. This data is readily available and thus help secure an acceptable fit between theory and data.

4.4.4.2. Operationalization of Self-Efficacy Theory

If self-efficacy theory holds validity in the open government community domain, it might be possible to help solve the participation challenge by increasing the users' self-efficacy. This leads to the following research question: *How does knowledge of other users' gratitude for previous contributions affect future contributions?* In order to answer this research question, one null hypothesis and one alternative hypothesis are proposed:

H₀ **H_0** Knowledge about users gratitude for the subjects' previous contributions does *not* affect the subjects' future contributions

H_a Knowledge about users gratitude for the subjects' previous contributions affects the subjects' future contributions

According to Schunk (2010), people with high self-efficacy are trying harder and longer to complete tasks than people with low self-efficacy. In terms of K10, it is assumed that community members with high self-efficacy are more likely to be active in the community. The aim of the third experiment is to increase the subjects' self-efficacy in order to make them contribute more.

Variables	Measurements
Increasing self-efficacy	Performance accomplishment, verbal persuasion, and physiological states
Contributions when self-efficacy is sought to be increased	The number of posts contributed when self-efficacy is sought to be increased
Contributions when self-efficacy is not sought to be increased	The number of posts contributed when self-efficacy is not sought to be increased

Table 17: Variables and measurements for self-efficacy theory

Two variables are used to operationalize the theory: increasing self-efficacy, and measuring contribution level after the subjects are being acknowledged.

4.4.4.2.1. Variables

The *first* variable is increasing the subjects' self-efficacy. As noted above, self-efficacy is derived from four information sources: performance accomplishment, vicarious experience, verbal persuasion, and physiological states. Three of these information sources are used as variables in this experiment.[159]

Performance accomplishment is the most important information source in self-efficacy (op cit.). This variable is based on information about mastery of previously performed tasks (Bandura 1977). The email and message sent in this experiment highlights previous performance accomplishment by stating that other users are aware of the performance. *Verbal persuasion* can help convince a person that she has the capabilities to cope with a situation. However, this information is not as effective as performance accomplishments because they do not "provide an authentic experiential base" (Bandura 1977). The email and message sent in this experiment uses verbal persuasion by leveraging social proof (Cialdini 2001).[160] *Physiological state* is affected by emotional arousal (Bandura 1977). The email and message sent in this experiment seeks to stimulate emotional arousal through appraisal. In sum, the email and message sent to the subjects aims at encouraging them and showing that they have an impact on the K10 community. The wording in the email[161] and message aims at improving their self-efficacy:

---

[159] Vicarious experience is information learned by observing how others perform an activity. This variable is not included in this experiment

[160] This is further discussed in Chapter 8.

[161] Due to a small sample size, this experiment relies on within-subject design. This means that all subjects are used in both the control and experimental condition. Subjects in the experimental condition receive an email and message that shows they have an impact on the community. Subjects in the control condition receive no mail or message.

> Hi [username],
>
> Last month, we conducted a survey on K10. Among the questions was this: *Is there a K10 member who does something particularly good for other users?*
>
> Several users highlighted you for your contributions in the community.
>
> So we just want to let you know that other users are aware of you and appreciate your contributions to K10.
>
> If you have any questions, you can just reply to this message.
>
> Best regards,
> K10
>
> Ps. Many users also wrote that K10 pull together – and that all users contribute in their own way. We agree. We do not wish to highlight some members over others, but rather express our gratitude for your contributions to K10. It means a lot to other people.

The *second* variable is the contribution level for the subjects after the experimental intervention. Here the self-efficacy is sought to be increased for the *experimental group*. The contribution level is measured for both the control group and the experimental group.

4.4.4.2.2. Measurements

As described above, the subjects' self-efficacy level is not measured. However, as this study is a controlled experiment, it is likely that any statistically significant change is caused by the intervention as other variables are held constant.

The other two variables focus on the number of contributions. This variable is measured over a two-week period. For details about measurement, please refer to section 4.4.2.2.

4.4.5. Social Identity Theory for Experiment 4

According to social identity theory (Tajfel and Turner 1979), human behavior in social contexts is derived from two things: personal identity and group identity. Human interactions are therefore always placed on a continuum ranging from purely interpersonal interactions to purely intergroup interactions (Hornsey 2008).[162] Intergroup interactions are furthered by making a distinction between *us* and *them* salient. This distinction makes the person enhance similarities to the in group and adopt some of the group's goals and motives as his own (Hornsey 2008). Thus, people act differently when they share a group identity (Hornsey 2008).

Social identity theory is hailed as "one of social psychology's pre-eminent theo-retical perspectives" (Brown 2000). The theory has undergone a high number of empirical tests using different methodologies, including laboratory and field ex-periments (Hornsey 2008). Social identity theory was originally developed as a theory about intergroup relations, but is now broadly applied in areas such as group cohesion (Brewer 1993), justice (Tyler 2011), management (Ashforth and Mael 1989), and social influence (Brown 2000; Hornsey 2008). [163] Although

---

[162] "A purely interpersonal interaction … involves people relating entirely as individuals, with no awareness of social categories. A purely intergroup interaction is one in which people relate entirely as representatives of their groups, and where one's idiosyncratic, individualizing qualities are overwhelmed by the salience of one's group memberships." (Hornsey 2008: 206).

[163] The scope of the theory has increased significantly from the original theoretical intentions. This means that the theory now is so broad that it is hard to falsify, which is a potential threat to its accuracy (Hogg and Williams 2000). Furthermore, the nomenclature around social identity theory is not as consistent as one could desire. One of the authors of the social identity theory, Turner, later produced a closely related theory, named self-categorization theory. The combination of these two theories is sometimes referred to as *social identity perspective* and *social identity approach*

some studies have only shown low correlations between identity perception and actions, the theory is still considered accurate in many contexts (Hornsey 2008).

As described in the literature review earlier in this chapter, social identity theory has successfully been applied to online communities in six different studies (Utz 2003; Shen and Chiou 2009; Nambisan and Baron 2009; Shin and Kim 2010; Lampe et al. 2010; Casaló, Flavián, and Guinalíu 2010). The theory can explain some of the factors that motivate participation in online communities. It is therefore likely that the theory can help improve participation in K10, and thus ensure a fruitful outcome of the experiment.

4.4.5.1.1. Limitations

Group identification is often measured through surveys and scales (e.g. (Luhtanen and Crocker 1992). As described further in Chapter 5, the work in this dissertation is primarily based on observed, not self-reported, behavior. This makes it challenging to measure if the subjects do indeed feel a social identity at K10. As noted extensively, the aim of the research conducted here is not to verify/falsify theories, but to examine if a particular operationalization of these can lead to increased participation in K10. The dependent variables used to measure different levels of social identity theory are behavioral data from K10.

4.4.5.2. Operationalization of Social Identity Theory

In order to operationalize social identity theory, the following research question is proposed: *How does benefit of contribution affect subsequent behavior on K10?* Based on social identity, one null hypothesis and two alternative hypotheses are tested:

H_0 The group constructs have no effect in terms of engagement and contribution behavior.

(Hornsey 2008). It is this combined meaning of the term that is used in this dissertation, although the term social identity theory is used.

H$_a$ Subjects assigned to the *benefit other* groups will engage and contribute more than subjects assigned to *benefit self* and control group.

H$_b$ Subjects assigned to the *small benefit other* group will engage and contribute more than subjects assigned to *large benefit other* group.

As described above, social behavior is determined by a mix of interpersonal and intergroup drives. It is unlikely that behavior is caused exclusively by one of these drives. When Social Identity Theory is operationalized in the fourth experiment, the aim is to prime the subjects with different levels of interpersonal/intergroup drives. When subjects are primed with a message appealing to their interpersonal identity, it is expected that they will act differently than when the message appeals to their intergroup identity.

One challenge in this experiment is that it comprises both new and existing K10 users. Existing users are more likely to be part of a K10 group-identity than new users. However, the K10 group identity is not the only relevant identity at play here: the group identity associated with early retirement pension and flexjob is equally relevant. K10 can in fact be seen as a smaller community embedded in a larger early retirement pension and flexjob community. It is thus assumed that it is sufficient to appeal to the overall pension and flexjob identity in order to make the in-group identity salient.

This approach is the same as used in several classic social identity theory experiments. Tajfel et al. (1971) and Billig and Tajfel (1973) form groups immediately before the experiment, and still detect a social identity bond during the experiment (Hornsey 2008). Moreover, as reported above, Utz (2003) finds that group identity exists independently of a member's time in an online group. While this group construct is not ideal, it is still justifiable.

Variables	Measurements
Identity priming	Group constructs
Effect of identity priming	Click through rate, new accounts, new posts, new threads, time on site, pageviews and bounces.

4.4.5.2.1. Variables

The *first* variable in the theory operationalization is identity priming. To find out if intergroup identity affects participation, four different groups are created.

The *second* variable, effect of identity priming, seeks to detect how the members of different groups behave after the identity priming. Seven dependent metrics are tracked: click through rate, new accounts, new posts, new threads, time on site, pageviews, and bounces. These levels of measurement build on the findings from Shen and Chiou (2009), who found that increased group identity is positively correlated with investment in the community and loyalty. In addition, they build on Lampe et al. (2010), Shin and Kim (2010), and Casaló, Flavián, and Guinalíu (2010) who find a shared identity to be correlated with intention to contribute to online communities.

Based on these findings, it is expected that the subjects primed with a group identity are more likely to create new accounts, contribute posts to the site, and spend time in the community.

4.4.5.2.2. Measurements

Each group is *primed with a different identity frame*. Group 1: Priming of interpersonal identity: participating at K10 benefits the subject herself. Group 2: Priming of intergroup identity: participating at K10 benefits a small in-group. Group 3: Priming of intergroup identity: participating at K10 benefits a large and abstract in-group. Group 4: No identity primed. This priming is done through advertisements on Google.com.

It is assumed that these primes work, but their effects are not measured directly. As described above K10 is a poor fit with this theory due to long and complex standardized scales. This weakens the operationalization of the theory, and few claims can be made about the accuracy of the theory in this domain.

Click through rate (CTR): Different ad copies are shown to different users. The first dependent variable is CTR, i.e. how many percent of the users react on the advertisement they are exposed to. This data is provided through the Google AdWords platform. If a Google user clicks on the advertisement he counts as a subject in the experimental group corresponding to the advertisement. If a Google user does not click on the advertisement, he counts as a "no clicker" and is not used as a subject.

Account creations: As the advertisements mainly target new users, it is relevant to see how many of the subjects that create an account on K10. Account creation is an important metric for a community as it is a prerequisite to participation. If one of the experimental subjects creates an account while the experiment runs, it is noted as an account creation.

Posts written and threads started: As in the other three experiments, an important dependent variable is the number of contributions the subjects make. Due to technical limitations in vBulletin, it is impossible to connect individual user IDs to the visitors from the Google AdWords campaigns. Thus, it is not possible to track number of new posts directly. A workaround gives a very close proxy to number of posts. Instead of tracking number of posts contributed by each user, it is possible to track the number of times a user clicks on respectively "Write reply" and "Start new thread" buttons. This number will likely overestimate the number of posts slightly, as some users start a post without submitting it. However, this bias is not systematic and is equally likely to happen in all four conditions. Thus, if one of the experimental subjects tries to create a new thread or write a post, it counts as a new post or thread.

Pageviews: Number of pageviews is a metric indicating how interested a user is in a website. A high number of pageviews usually indicates high engagement, although it occasionally indicates usability issues as users have to go through much content before finding what she is looking for (Høgenhaven and Andreasen 2011). Therefore, this metric is not as reliable as the number of posts metric.

Despite these limitations, a high number of pageviews is interpreted as a positive sign of engagement. Insofar this metric varies across groups it means that some users are either more engaged or more persistent in using K10. Potential usability problems are constant across groups, thereby making this interpretation of the metric less relevant in this particular context. Number of pageviews is also interpreted as a sign of engagement by Dow et al. (2010). Moreover, it is often used as the primary metric on media websites because many websites earn advertisement income for each generated page view. Technically, a page view is logged in Google Analytics "each time the tracking code is executed on a web page" (Google 2013b: 1).

Time on site: According to Google Analytics, time on site "indicates how long a visitor spent on a particular page or set of pages. It is calculated by subtracting the initial view time for a particular page from the initial view time for a subsequent page. Thus, this metric does not apply to exit pages for your site" (Google 2013c: 1). This means that this metric underestimates the actual time on a page as users only seeing one page are noted as zero seconds, regardless how long they stay on the site.

This metric has the same interpretive challenge as number of pageviews: a long time on a site can be a sign of both highly engaged users and of usability problems. Nonetheless, long time on a site is interpreted as a positive sign of involvement and persistence in this experiment. Time on site is also interpreted as a sign of engagement by Dow et al. (2010).

Bounce rate: In Google Analytics, bounce rate is defines as "the percentage of visitors that see only one page during a visit to your site" (Google 2013a). In other

words, users are considered to bounce when they come to K10 and do not click on any active links on the site before leaving it. A bounce indicates low interest and low engagement as the users do not click anywhere on the website.

In sum, the following dependent variables are measured in this experiment: (a) CTR, (b) number of account creations, (c) number of new posts and threads, (d) time on site, (e) number of pageviews, (f) bounce rate.

4.5. Chapter Summary

Existing research shows that social science theories are able to improve participation in online communities. In this dissertation, it is examined if the same holds true for communities in the government sphere. Out of many relevant social science sub-disciplines – most notably economics, political science, sociology, and social psychology – social psychology is chosen as the theoretical lens in this dissertation. Social psychology is preferred due to a good level of abstraction match, methodology match, and broad set of mid-level theories that can be used to generate many different design ideas.

The main contribution of this chapter is a literature review of social psychology applied to online communities. It shows that the theoretical application is a growing research area. The literature review, covering 2000-2010, illustrate that 12 theories are used to study online communities: social comparison theory, social learning theory; cognitive dissonance theory; self-concept theory; need to belong; the collective effort model; reciprocity theory; group conformity theory; goal-setting theory, social presence theory; perceived identity verification theory; and common identity and bond theory. The data used in the research is overwhelmingly quantitative, primarily relying on survey and field experimental methodology. The most commonly used research design is a single case study.

Four out of the 12 theories are selected based on case relevance and granularity level. These will guide the dissertation's four experiments: social comparison theory, goal-setting theory, self-efficacy theory, and social identity. Each theory is

operationalized to accompany an experiment. This important link makes it possible to generate research questions and hypotheses, and thus use theory to improve the participation in K10. Before the four experiments are conducted and reported in chapter 6-9, the research design and methodology is described in the next chapter.

Chapter 5: Research Design and Methodology

5.1. Introduction

Open government faces a participation challenge. To help solve this challenge, four social psychology theory-based experiments are conducted on K10. This chapter outlines the research positioning, methodologies, and design used to perform these experiments.

In this process, the following steps are taken: First is a description of the ontology, epistemology, methodology, and axiology used in this dissertation. These fundamental considerations are followed by a discussion of ethics and privacy concerns associated with online experiments. The chapter is concluded with a description of experimental design and analysis in online settings. This is the main contribution of this chapter. The research designs described in this chapter are at a general level. The specific research design and analysis for each experiment is described in depth in chapter 6-9.

5.2. Research Foundations and Positioning

Although this dissertation is submitted to Copenhagen Business School, Denmark, much of the research has been done at Cornell University in USA. This makes it relevant to describe respectively the Scandinavian and American IS traditions.

The Scandinavian approach arose in the 1960ies and has roots in Marxism, critical theory, and trade unions (Nygaard and Bergo 1975; Bansler 1988; Spinuzzi 2002). This research tradition is often concerned with increased user participation in information systems design (Bjerknes and Bratteteig 1995; Spinuzzi 2002). As explained by Bjerknes and Bratteteig (1995), user involvement has three benefits: (1) improved knowledge foundation under the system; (2) managing expectations and reducing change aversion; (3) increasing work-place democracy by giving users influence in the design process. It is the third benefit

of user-involvement that is most crucial to the Scandinavian IS tradition (Bjerknes and Bratteteig 1995).

This democratic aim is pursued through empowerment of the workers in the design process, thus ensuring that the information systems are an expression of the workers', and not just the organizations', needs and interests. Topics in the Scandinavian tradition include user-centered design (Olson and Olson 1991), participatory design (Bjerknes, Ehn, and Kyng 1987; Ehn 1993; Kyng 2010), distributed participatory design (Danielsson et al. 2008), co-creation (Molka-Danielsen 2011), emancipatory design (Holmlid 2009) cooperative prototyping (Bødker and Grønbæk 1992), and online communities (Brincker and Gundelach 2010). The knowledge collected in the Scandinavian approach is often used to make broader reflections and theorizing over the societal changes caused by technology (Markus and Bjørn-Andersen 1987; Mathiassen 1998).

The Scandinavian IS literature usually relies on a qualitative approach (what Fitzgerald and Howcroft (1998) call "soft" research) such as ethnographic studies, user studies, usability testing, scenarios, prototyping, design games, and interviews (Spinuzzi 2005).

Spinuzzi (2002) argues that the American IS tradition is based on Capitalism rather than Marxism. The aim is profit maximizing, not democracy (Bansler 1988). This entails that the company, not employees, have the major say in the development of information systems. In this tradition, regular employees are not hired to use systems, not design them. As described further below, the American IS tradition has a quantitative approach (what Fitzgerald and Howcroft (1998) call "hard" research), in which the aim is to identify the best possible information systems in terms of selected metrics. Often used methodologies include experiments, log analysis, and surveys.

Although the two traditions are quite different in origins, they have some common ground as both focus on the users. In this research, I follow the American tradition rather than the Scandinavian. There are four reasons behind this

choice: *First*, much of the literature reported in Chapter 4 relies on quantitative analyses of large sample data sets rather than qualitative analyses of small sample interviews and user studies. Since the research reported in this dissertation follows the literature reported in Chapter 4, the "hard" American research tradition is more appropriate than the "soft" Scandinavian one. It is not to say that the American tradition is superior to the Scandinavian. They are considered supplemental, not competitive.

Secondly, as written in the preface, much of the research has been conducted at the Information Science department at Cornell University. In order to maximize learning during the stay as a visiting research fellow, it was logical to follow the local research tradition.

Thirdly, this dissertation is sponsored by the industrial PhD program in Denmark. In order to gain useful knowledge to be implemented in the sponsoring company, it is arguably logical to take an American company-centric approach instead of the Scandinavian employee-centric one.

Fourthly, K10 is the IT artifact examined in this dissertation. As described further in Chapter 3, a single person owns and moderates the platform, making it impossible to involve any employees in the design process. On the other hand, it would be possible to involve some of the many thousands of users on the site in the design process. The vast majority of the k10 users are on public welfare programs due to varies disabilities, and it is therefore difficult to involve them heavily in the design process. This arguably makes the American approach more suited for this case.

In order to advance information science research by being more explicit about research foundations, Fitzgerald and Howcroft (1998) propose a scheme comprising of a set of dichotomies covering ontology, epistemology, methodology, and axiology. This scheme is used to structure the first part of this chapter. Table 19 gives an overview of the stances taken in this dissertation.

Level	Parameter
Ontological Level	Realist
Epistemological Level	Positivist
	Objectivist
	Outsider
Methodological Level	Quantitative
	Exploratory
	Deduction
	Field
	Nomothetic
Axiological Level	Rigor

Table 19: Dissertation positioning based on Fitzgerald and Howcroft's (1998) soft vs hard research dichotomies

5.3. Ontology

Ontology is concerned with "what is there" (Hollis 1994: 8) and is important to research because it defines the observed entities. The ontology in this dissertation is based on realism. This ontology rests on the assumption that there is a real, tangible world out there, which can be examined and tested independently by a detached researcher (Walsham 1993; Hollis 1994). The aim of research within ontological realism is to identify causal relationships between the examined elements. Research within this realism ontology focuses on the elements that can be measured quantitatively.

In realism, the ontology is seen as precedent to epistemology. The ontological reality exists out there and it can be examined through epistemological inquiry (Wendt 1999). In other words, the world is already existing and it is the researcher's role to find out what exist and identify the causal mechanisms at play (Hollis 1994).

This entails that open government communities are treated as objective/brute facts (Searle 1995). Through experimental design, it is possible to gain real, independent, and objective knowledge about an online community and the people using/creating it. Through theory-derived hypothesis testing and experimental design, objective knowledge about an open government community can be derived and described.

From a realist ontological stance, an open government community comprises people, technology and artifacts: through a technological system, people write posts. These are the directly observable elements that can be studied within this ontology.

From a relativist perspective (Harré and Krausz 1996), the realist ontology fails to grasp how entities and facts are socially constructed and highly context dependent. The same people, technological system, and entities can have highly different meanings in different context. Although elements are treated as brute facts in the realist ontology, they are indeed socially constructed. Consequently, realists tend to view elements as static and objective when they indeed can be changed. Although this critique has plenty of merit, it cannot be incorporated into this dissertation, as there is only room for one ontological stance due to incommensurability between multiple ontologies (Kuhn 1996; Jackson and Carter 1991; Fitzgerald and Howcroft 1998).

5.4. Epistemology

Ontology and epistemology are closely attached, as epistemology is concerned with how knowledge about *what is there* is obtained. Three different dimensions make up the epistemological stance: positivism, objectivism, and being an outsider.

5.4.1. Positivism

Positivism is the epistemological position most suited to accompany a realist ontology (Hollis 1994). A positivistic epistemology entails that the researcher remains as detached and objective as possible, and that hypotheses are tested through direct observation (Hollis 1994). Through hypothetical-deduction and experimental design and analysis, it is possible to examine causality between design and behavior.

Taking a positivistic epistemological position aligns well with much of the research used in this dissertation. The majority of the research used in Chapter 4 is published in HCI journals and conferences, which, according to Clemmensen (2006), is dominated by positivism and quantitative experiments, often conducted in laboratory environments. Sharing this positivistic epistemological position makes it easier to communicate with the HCI field. One obvious caveat of this is that any research field benefits from applying different perspectives and paradigms to the same problem (McGrath 1995).

5.4.2. Objectivism

When conducting the experiments, the aim is to collect data about the subjects' behavior that is objectively true. These observations need to have as few biases as possible (Fitzgerald and Howcroft 1998). The researcher needs to be detached in order to avoid personal biases and instead rely on objective data metrics. The data points collected are quantitative metrics such as number of posts written or pageviews within a defined time frame. Such metrics are meaningful, because they measure actual behavior that took place.

The four experiments conducted and reported in this dissertation rely on in-depth analysis of one data source rather than superficial data analysis of many sources. This methodological choice follows the tradition of the majority of the online community research reported in Chapter 4.

Quantitative metrics do not tell the full story of behavior in online communities. But objective metrics give insights into actual behavior. In many ways, the total number of posts per user speaks for itself. A different insight and possible broader understanding would be gained by conducting qualitative interviews with members of the online community and/or through discourse analysis of the posts in the community. This is, however, outside the scope of this dissertation. One caveat of this process is that not having such qualitative and subjective data increases the likelihood of misunderstanding the quantitative data collected.

5.4.3. Outsider

Operating within the environment that is being examined might increase the understanding of the situation. But it also risks infusing preconceptions and biases into the research. When taking a realist ontological stance, it is therefore necessary to be a detached outsider (Hollis 1994).

The aim of this dissertation is to examine open government communities without contaminating the subject matter with own biases and values. The case used in this dissertation, K10, has existed since 2005. I have not been a user of this community, and I only use it as experimental environment. Thus, I am not part of the community or pretending to be. Going into the research, I have no *a priori* relations to anyone in the community, except the owner, and thus no particular interests or need to take inappropriate considerations to users. This means that I, the researcher, am an outsider rather than insider. The benefit of being an outsider is that I am unbiased towards the community.

Being an outsider comes at a price. I do not have any firsthand experience with the community, and thus not a complete understanding of the norms, values, and behavior within the community. This might entail that the social psychology theory based experiments are suboptimal or incongruent to the community. Being an outsider is likely to increase the risk of failed experiments, simply due to lack of firsthand knowledge about optimal fit between theory and practice.

5.5. Methodology

A methodology is a scientific process used to examine and identify the social world (Hollis 1994). A realist ontology and positivist epistemology narrows the methodological choices.[164] Methodology used in a positivistic epistemology is concerned with explanatory outcomes and causality (Hollis 1994). Five different

---

[164] It is necessary with congruent ontology and epistemology, due to incommensurability (Fitzgerald and Howcroft 1998). This incommensurability does not apply to methodology and axiology. Adhering to either a hard or soft approach helps guide the research, but it is not a strict requirement.

dimensions make up the methodological stance (Fitzgerald and Howcroft 1998): quantitative, exploratory, deductive, field based, and nomothetic.

5.5.1. Quantitative

The type of data that is being collected is largely a consequence of the theoretical, ontological, and epistemological choices made. In terms of data type collected, social psychology theory has stronger quantitative than qualitative traditions (cf. the literature review in Chapter 4). Due to the focus on objectivity, direct observations, and causality, quantitative data aligns well with the realist ontology and positivist epistemology.

As stated in the initial research question driving this dissertation, the aim is to examine how participation can be improved in open government communities. The primary dependent variable measured is number of contributions. As described in section 2.3.3, participation is measured through number of posts written by users within a given time frame. Thus, the data collected in this dissertation is quantitative. The nature of the data is observational: it is overt traits of actual behavior.

Measuring behavior through number of posts is also a limiting factor. For overt behavior cannot account for everything (McGrath 1995). The most important limitation is that it does not generate the same interpretive depth as qualitative data can. It would therefore be preferable to not only measure the number of posts, but also the quality. But as explained in section 2.3.3, there are no scalable metrics that measure the quality of online participation at this time (Farina et al. 2010).[165]

According to Fedorowicz and Dias (2010), quantitative IS research is usually focused on optimizing an existent system toward a given goal. Qualitative research

---

[165] In the case of K10, basic quality is ensured in the comments as low quality comments such as spam are deleted by administrators, and thus not included in the dataset. The participation quality therefore meets some low-threshold minimum standards.

is, on the other hand, more fundamental and focuses on how design is done. This critique has some validity to it. The experiments in this dissertation seek to improve participation in an existing online community within the current design framework of the community. This entails that few, if any, fundamental changes are considered.

In this dissertation, large-scale experiments are used as the only methodology. It is outside the scope of this dissertation to elaborate thoroughly on alternative research methods such as statistical analysis, surveys, diaries, case study, interviews, focus groups, ethnography, usability testing, and participatory design (Lazar et al. 2010). It is fully acknowledged that the use of experimental methodology only gives particular insights into the examined open government community, K10. The research would be improved if the experimental methodology is supplemented with other methodologies - especially qualitative methodologies such as diaries, interviews, and ethnography. These offer substantial descriptions and depth to the quantitative observations generated through the experiments (Boellstorff et al. 2012). But as noted by Boellstorff et al. (2012), it is the research corpus in aggregate that ought to rely on multiple methodologies, not necessarily the individual piece of research. This is discussed further in section 11.5.

5.5.2. Exploratory

The research in this dissertation is exploratory in nature. By operationalizing social psychology theory into specific hypotheses, it is examined to which degree the theory can be used to improve participation in K10. It is exploratory in the sense that it is applying and developing existing theoretical body to a new subject matter.

As stated in Chapter 4, the primary aim is to use theory to examine and improve existing open government communities. This entails primacy to exploration over confirmation. It is not the aim to verify or falsify the theories, as this requires

multiple constructs of each theory applied to multiple cases. This work is outside the scope of this dissertation.

5.5.3. Deductive

The process in each experiment is hypothetical-deductive. Research design and hypotheses for statistical design is generated from social psychology theories. Consequently, this dissertation is clearly positioned as deductive.

5.5.4. Field Based

This dissertation relies solely on online field experiments. As described earlier in this chapter, this methodological choice is closer to the American than the Scandinavian research tradition. Three reasons account for this reliance on experimental methodology. *First*, a significant part of this dissertation work has been conducted at Cornell University. This makes it relevant to follow the local research methodology in order to maximize learning.

Secondly, as documented in Table 10 (in Chapter 4), field experiments are much more common than laboratory experiments in online community studies. As I am new to this research community, it makes more sense to join the existing research paradigm rather than attempt to change it.

Thirdly, there are more advantages with online field experiments than there is with online laboratory experiments. Comparative advantages and disadvantages thoroughly examined in section 5.8, later in this chapter.

5.5.5. Nomothetic

Idiographic research is individual-centered and nomothetic research is group-centered (Fitzgerald and Howcroft 1998). The underlying assumption of this dissertation is that social psychology theory can improve open government communities. The focus is on group outcome rather than the individual subject. Consequently, the methodological perspective will be nomothetic rather than ideographic. The level of analysis is further described in section 4.3.

5.6. Axiology

On the axiological level Fitzgerald and Howcroft (1998) proposes a dichotomy between relevance and rigor. The two parameters are somewhat contradictory as highly rigorous research is very time-consuming, causing research to be less relevant at the publication time.[166] Although the two dimensions in some ways are contradictory, "relevance should come first and drive rigour" (Fitzgerald and Howcroft 1998, 12). Rigor is often built over time when a research corpus is built (Fitzgerald and Howcroft 1998).

In an emerging domain such as open government, relevance becomes even more important. In order to keep up with practice and be relevant, open government research needs to be conducted and published within a reasonably short time frame. Some rigor is thus sacrificed to benefit relevance. This does not exempt the work from careful research design and statistical analysis. However, the priority given to relevance over rigor means that the research is of minimal use to the social psychology theory literature it relies on. The construction of variables is not done sufficiently rigorously that alternative theoretical explanations can be ruled out.[167]

Finally, source credibility is an important part of rigorous research. Please refer to Appendix 3 for an assessment of the source credibility of the literature used in this dissertation.

After defining the fundamental aspects in terms of ontology, epistemology, methodology, and axiology, the next sections describe and discuss concerns regarding ethics and privacy in the experimental research hitherto described.

---

[166] Long double-blinded peer review processes serves as an example of this.

[167] As described in chapter 1 and 0, this is not the aim of the research, as the focus is on design practices rather than theory.

5.7. Ethics and Privacy Concerns

Although the internet has been mainstream for at least a decade, research practices and ethical guidelines are still not as clearly defined as they are in offline environments (Wilson and Peterson 2002). Current ethical research guidelines such as the APA ethical principles (American Psychological Association 1992), do not cover the specifics of internet research (Bordens and Abbott 2010).

According to Bordens and Abbott (2010), experimentation with human subjects ought to fulfill three fundamental criteria: human, respectful, and ethical. Of these three terms, ethical is the unclear. Two types of ethical and privacy concerns are relevant to discuss in the context of this dissertation: informed consent and privacy. The first issue is concerned with research ethics and the second one with legal issues in terms of privacy and data management.

5.7.1. Research Ethics

As described in section 5.9, the experiments in this dissertation are conducted directly in the field. In online field experiments, the users are not explicitly recruited for the experiment. They are already acting in a context, often unaware that they are part of an experiment. This leads to the principle of informed consent.

5.7.1.1. Informed Consent

Asking for subjects' consent has been an important part of research since the Nazi medical experiments on coerced subjects in the early 1940ies (Boellstorff et al. 2012). This need has been reinforced by psychological experiments (e.g. Haney, Banks, and Zimbardo 1973; Milgram 1974) that potentially hurt subjects (Boellstorff et al. 2012). But as noted by Boellstorff et al. (2012), such experimentation is often very different from online research. This makes it relevant to assess the informed consent criteria in the specific research context.

To increase the realism and internal validity of the experiments, the experimental subjects are not informed of their part in the experiment. This entailed

that the subjects do *not* give an explicit, informed consent to participate in the experiments. There are three important concerns in this informed consent discussion: the risk for the subjects, the research context, and the beneficence of not asking for consent.

5.7.1.2. Risks for the Subjects

The four experiments conducted in this dissertation evolve around relatively small changes. Design patterns similar to those used in the experiments are already used in other online communities. The fact that such design patterns are already adapted in other contexts makes it reasonable to assume that they are not harmful. Furthermore, similar experiments have been exempted from Institutional Review Board evaluations on Cornell University.[168] This strongly suggests that the risks and ethical concerns in such experiments are virtually non-existing. If an experiment causes a risk to the subjects, such exemptions are not granted. In sum, as the experiments come with very low risk, it is considered ethical justifiable to omit informed consent.

5.7.1.3. The Context

The omission of informed consent cannot be understood detached from the specific context: the internet. Boellstorff et al. (2012), find it

> *"legitimate to see subscription-based virtual worlds as having public areas where it is not necessary to have every person in an interaction sign an informed consent form – just as there is nothing inherently unethical about taking a picture of a tourist in an open, general area…"*

> (Boellstorff et al. 2012: 134f)

The same argument applies to online communities (Wilson and Peterson 2002). This interpretation is also the *de facto* moral code on the internet. Each day a

---

[168] A copy of the IRB exemption is attached in Appendix 4.

vast number of experiments are conducted around the internet. For example, Google wants to "run so many experiments, we can't afford to put you in any one group, or we'd run out of people… On most Google queries, you're actually in multiple control or experimental groups simultaneously" (Levy 2011: 61). In other words, each time someone uses Google, she is part of one or more experiment. This means that millions of people are part of similar experiments every minute without giving prior consent.

Another important aspect of the research context is that millions of *natural experiments* are conducted on the internet each day. Websites are changed, emails are sent, advertisement bought, and surveys conducted. Most of these are not treated as true experiments due to lack of control groups. But internet users around the world are daily exposed to conditions equal to the experimental manipulations in these experiments without giving their consent. And even when internet users have to consent to terms and conditions by ticking of an *I agree* box, less than 1 % of the users actually read the terms and conditions before giving their consent (Luger, Moran, and Rodden; 2013).

This practice is also seen in much published research. For example, the following articles from the literature reported in Chapter 4 do *not* report having attained any informed consent prior to their data collection: Cosley et al. (2006); Harper et al. (2007); Drenner, Sen, and Terveen (2008); Burke, Marlow, and Lento (2009).[169] In short, the norm in the research context is not to ask for consent both in academia and practice.

5.7.1.4. The Beneficence of Not Asking for Consent

An important part of the ethics discussion is to which degree the research yield results that are beneficial to society at large (Boellstorff et al. 2012). As noted by Keller and Lee (2003), "The concept of beneficence (weighing the amount of good the research will do against the risks, as well as making sure the research is

---

[169] This is not an exhaustive list of online experiments omitting informed consent, but merely a small sample from the papers included in the literature review in Chapter 3.

not purposefully selective about who benefits) must always be considered" (Keller and Lee 2003: 214).

Asking subjects for consent is a serious threat to the validity of the research, and thus to the findings. As websites normally are changed without asking for the users' consent, the most realistic setting is to not ask them in experiments. Telling users about experiments is likely to confuse the subjects as well as cause a Hawthorne effect. The internal validity of such experiments is thus very questionable, potentially undermining the purpose and value of the experiment. Not asking for subjects' consent is justified by very low risk for the subjects and likelihood of more valid research findings. In sum, the benefits of not asking for consent are considered to outweigh the negative and thus be favorable.

5.7.2. Privacy and Data Management

According to Keller and Lee (2003), the primary ethical principles internet researchers ought to be aware of are "protecting the anonymity and privacy of the participants and being cognizant of cultural issues such as socioeconomic status, gender, race, and disability, among others" (Keller and Lee 2003: 213). This section describes how subject privacy, including cultural issues, is addressed in this dissertation.

The first important point to make in terms of privacy is that this research is conducted on a public dataset. Since the content written on K10 can be accessed by everyone with an internet access, the data cannot be considered confidential (Boellstorff et al. 2012). Furthermore, K10 users interact through aliases, and are thus not identifiable.[170] The only data point that *potentially* links the recorded data to real people is their email addresses. This information is not public, but stored in the K10 database. This data point was not collected for this dissertation but was in the database before the experiments were conducted. As explained

---

[170] It is possible that users use their full real name in their alias and thus can be recognized. Using a real name as an alias is, however, a rare exception to the rule. But the overarching point is that the aliases are public information, entailing that I, the researcher, do not have extra knowledge.

further below, the email addresses are used in two experiments but immediately removed from the dataset, thereby making it impossible for me or others to identify any real person based on the dataset.

Whereas the previous section on consent is based on ethics, privacy is strictly regulated by both Danish and European law. In terms of privacy and data management, the research is primarily subject to the EU *Data Protection Directive* and the Danish *Act on Processing of Personal Data*.[171] In the following sections, it is described how this research is regulated by these two directives and how data is managed in order to respect the legislation.

5.7.2.1. The EU Data Protection Directive

According to the *EU Data Protection Directive*, officially Directive 95/46/EC (The European Parliament And The Council 1995), personal data are defined as "any information relating to an identified or identifiable natural person".[172] As the datasets used in this dissertation do not relate to identifiable or already identified persons, the dataset is *not* subject to the personal data regulations in the EU 95/46/EC Directive. Although some aliases and email addresses *potentially* makes it possible to link a user to a real person, no such linkage is made, and the critical data point, email addresses, is deleted from the datasets used in the analyses.

The K10 website as such is subject to the 2011 revision of the 95/46/EC Directive,[173] mandating that users now need to consent that cookies are tracking their behavior. This is described later in section 5.7.2.3.2.

---

[171] In Danish: Persondataloven

[172] The full definition of personal data is "any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" (Directive 95/46/EC art. 2).

[173] This revision was implemented in Denmark on May 25th 2011

5.7.2.2. Act on Processing of Personal Data

Generally speaking, *The Act on Processing of Personal Data* allows researchers to collect and examine personal data in Denmark.[174] One has to be especially careful when dealing with sensitive data. According to *The Act on Processing of Personal Data*, data is considered sensitive if it concerns peoples (1) race or ethnicity, (2) political, philosophical, or religious beliefs, (3) associations to unions, or (4) health and sexual relations.[175]

As described above, data cannot be considered sensitive when they are published by the data owner/creator (Datatilsynet 2001). By writing the information on a public forum that can be accessed and read by everyone, K10 users have made their information public.[176] The main dependent variable used in this dissertation is number of contributions per user. This data is already publicly available on each user profile at K10.[177] Such public information includes number of posts, average posts per day, and links to all posts per user.

In the fourth experiment additional dependent variables are used, including time on site, number of pageviews, and bounce rate. Contrary to the number of posts, this data is not publicly available. However, the data is automatically made anonymous by the software at the time of collection. This entails that neither I nor anyone else can link the collected data to particular individuals. Furthermore, basic usage data are not characterized as sensitive data, and are thus excluded from the requirements of *The Act on Processing of Personal Data.*

---

[174] The wording in Danish is "Hvis behandlingen er nødvendig for at kunne udføre en opgave i samfundets interesse. Dermed vil behandling af oplysninger til gavn for en bredere kreds af personer, eksempelvis i statistisk, historisk, informativt eller videnskabeligt øjemed kunne ske." (Datatilsynet 2000).

[175] Although some posts on K10 are health related, they are not relevant to the data collected and stored for the research in this dissertation. The content of each post is stored in the MySQL database (described further below), but is not accessed or used in this research.

[176] This entails that the data is *not* considered to be sensitive information and, thus, not subject to the rules concerning sensitive information. A similar argument is found in Pittenger (2003), who argues that the internet by default is a very public medium (cited in Bordens & Abbott 2010).

[177] See for example http://www.k10.dk/member.php?u=573.

The Act on Processing of Personal Data is monitored by the Danish Data Protection Agency. Through telephone correspondence, [178] I have verified that the research conducted in this dissertation is happening in concordance with the law.

5.7.2.3. Data Management

To ensure privacy and confidentiality of subjects, the data collected in this dissertation has to be stored securely. The data is stored several places: K10's MySQL database, Google Analytics, YMLP, Google AdWords, Survey Monkey, and my work computer. These six places are described below.

5.7.2.3.1. K10 MySQL Database

The primary data repository for the data used in this dissertation is the K10 MySQL database. This database is necessary for K10 to function as a website, as user generated data has to be stored in a database to be retrieved by web browsers. Thus, the data stored in this server has little to do with the research conducted in this dissertation.

Although the MySQL database is unaffected by these experiments, it is relevant to summarize the content of the database.

- The database does *not* contain information about the users' social security numbers.[179]
- Users have to submit an email address when signing up. This information is the only information that *might* be able to link a K10 user to a real person. As described above, this information is used in two of the experiments to contact users but is immediately deleted in order to preserve privacy and confidentiality.

Please refer to section 3.4.3 for more details on MySQL security.

---

[178] Conducted on August 6, 2012 at 9:59 am.

[179] Generally referred to as *CPR numre* in Danish.

5.7.2.3.2. Google Analytics

Google Analytics is used to track user behavior in one of the experiments. Google Analytics uses web browser cookies to track users across pages and sessions. Using this or similar analytics software is common on both governmental and non-governmental websites as it gives data points concerning number of visits, time of site, bounce rate etc.

Until recently, such cookie tracking has not been regulated. But the 2011 revision of the 95/46/EC Directive is now enforcing informed user consent of cookie tracking. There has yet to be a canonical way of implementing the cookie consent on websites. Many government agencies, including the EU, had yet to find a way to implement it at the time of data collection (Whittaker 2012). At K10, the information about Google Analytics was added to the terms of conditions to make users aware of it. Google Analytics was removed after the experiment (and so were the remarks in the terms and conditions).

5.7.2.3.3. YMLP – Email Vendor

In one of the experiments, emails are sent out via the email provider YMLP (yourmalinglistprovider.com). This is the only case where emails are exported from the K10 MySQL database. The email addresses were deleted from YMLP immediately after the experiment was conducted, and all email addresses were removed from the dataset in order to ensure anonymity, in case one or more of the email addresses could link a user alias to a real person.

After the data are exported to my work computer (see below), they were deleted from the YMLP interface in order to protect privacy and minimize the risk of data theft and third party abuse.

5.7.2.3.4. Survey Monkey

One of the experimental interventions is conducted through a survey. The survey software used to conduct the survey is Survey Monkey. The respondents are asked to fill in their K10 username. After the survey has been conducted, the data is exported to SPSS and merged with MySQL data from K10. At this point all

usernames are deleted from the dataset in order to preserve privacy. As soon as the data is exported to my work computer (see below), they were purged from the Survey Monkey database for privacy and safety reasons.

5.7.2.3.5. Google AdWords

One of the experiments uses Google AdWords as a means to cause the experimental intervention. Google AdWords collects data about how many users see an advertisement and how many users click on an advertisement. However, no identifiable information is made available to the advertiser (me). It is unknown what kind of data Google collects for themselves.

5.7.2.3.6. Work Computer

In order to do data analysis in SPSS, all data is downloaded and merged on my work computer. In the data set from two of the experiments, there are no potential personal identifiers to begin with. In the data from the two experiments relying on emails, the only potential personal identifier, email addresses, are deleted when the data sets are compiled. Thus, as a work computer is not the most secure data repository, no sensitive information is stored on the computer.

In sum, these data repositories are widely used by industry and government agencies around the world, and are considered to be reasonably secure. Despite a general good level of security, personal identifiers are sought not to be collected. When they are, they are deleted as early as possible in the process. And when using third party services for data collection, the data are deleted as soon as they are downloaded in order to eliminate the risk of data theft and abuse.

The next sections describe and discuss the applied methodology; experimental design and analysis. The subsequent sections are dedicated to the following topics: experimental design, experiments in online community research, shortcomings of experimental design, types of experimental design, validity, and the experimental process.

5.8. Experimental Design and Analysis

The applied methodology in this dissertation is experimentation, defined as "a trial-and-error process in which each trial generates new insights into a problem" (Lee et al. 2004, 310). Scientific experimentation is not a new invention. It originates from pyramid building in Egypt around 2613 - 2589 BC (Sundukovskiy 2010).

Experimental methodology is often used in psychology, social psychology and HCI literature (Kim, Barua, and Whinston 2002; Gilovich, Keltner, and Nisbett 2010). As the literature review in Chapter 4 shows, experimentation is the second most used methodology[180] in social psychology inspired online community research. Experimental design is also gaining momentum as an accepted way to gain verified knowledge outside academia. Some of the most successful internet companies such as Google and Amazon are driven by data and experiments.[181] Montgomery (2004) sees experimentation as a nearly universal remedy: "There is not a single area of science and engineering that has not successfully employed statistical designed experiments" (Montgomery 2004: 21). Despite these praising words, experimental methodology is not a superior research methodology. Like all methodologies its strengths and limitations are context dependent (McGrath 1995).

5.8.1. The Experimental Process: An Iterative Stage Model

As experiments are being used for a multitude of research purposes and domains, there is no canonical model for conducting experiments (Sani and Todman 2006). The specific steps vary, although the basic foundations of the experiment are identical. For the purpose of the experiments in this dissertation, Thomke's (1998) experiments within the innovation realm serve as a good model for experimentation.

---

[180] Survey is the most used methodology.

[181] For example, Amazon has a "culture of experimentation" (Kohavi et al. 2009: 141) and "At Google, experimentation is practically a mantra" (Tang et al. 2010, 17).

(Changes in Exogenous information are Due to Unstable User Needs, Rapid Technological Advances, etc.)

DESIGN REQUIREMENTS

Changes in exogenous information

- **(1) DESIGN** — • Use learning from previous cycle(s) to conceive and design an improved solution.
- **(2) BUILD** — • Develop models and/or build prototypes to be used in running experiments.
- **(3) RUN** — • Test model/prototype in real or simulated use enviroment
- **(4) ANALYZE** — • Analyze findings from previous step and learn

DESIGN ACTIVITY
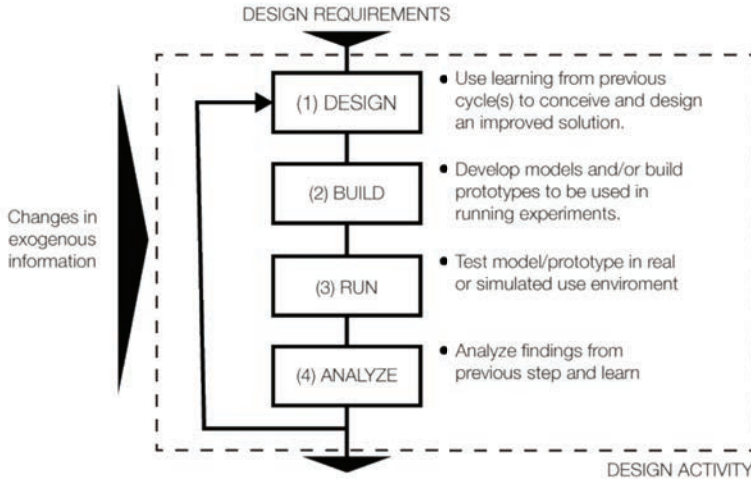
Figure 13: Experimental model adapted from (Thomke 1998: 745).

In this model, an experiment comprises four phases: (1) designing solution, (2) building prototype/product, (3) running experiment, and (4) analyzing the data. These four phases are sequential and illustrated in Figure 13. In most cases, the model is iterative, as follow up experiments often lead to further insights. If the outcome of the first test provides sufficient learning, further iterations are unnecessary (Thomke 1998). The next sections describe the key elements of this experimental process.

5.8.2. True Experiments

The purpose of a scientific experiment is to test the causal claims of a hypothesis (Lee et al. 2004). To ensure validity experiments have to be conducted in accordance with a set of rules and procedures (Sani and Todman 2006). For an experiment to be a true experiment, three conditions need to be fulfilled: (a) control over variables, (b) use of control groups, and (c) use of random assignment of subjects (Dawes 2010):

First, a true experiment is able to control variables. The experimenter needs to be able to manipulate the independent variable, while holding other variables constant (Sani and Todman 2006; Dawes 2010). For example, when conducting an email experiment, the emails sent to the various groups must be identical with the exception of the changes made to examine the independent variable. All other factors such as email subject, sender, delivery time, font size, format etc. must be constant. If they are not, it is impossible to know if observed changes are caused by the independent variable or other variables.

Secondly, a true experiment uses control groups. Causing a change and observing the difference is not a valid experimental design, as it does not account for changes in other factors (Kohavi et al. 2009). This is why a true experiment needs a control group, whose subjects are not exposed to the same changes as the experimental group. In a true experiment, it is insufficient to implement a change and do before/after comparisons.

Thirdly, a true experiment I characterized by random assignment (Campbell and Stanley 1963). Random assignment means that all subjects must have an identical chance of being assigned to the different treatments (McGrath 1995).[182] Experiments not fulfilling this requirement is a quasi experiment.[183]

---

[182] This does not mean that the different treatments must have the same size. For example, 90 % of the subjects can be assigned to the control group, and 10 % to the experimental group. In this example, randomization entails that each subject has 10 % chance of being assigned to the experimental group, and 90 % chance of being assigned to the control group.

[183] A quasi experiment is similar to a true experiment, but without randomized subject assignment. This design is usually applied when the researcher cannot control the subject assignment, e.g. when users are already assigned to groups such as gender and age. This research design is often used when a change in the independent variable has already occurred, meaning it can be examined but not manipulated by the researcher (Broota 1989). Quasi experiments are therefore seen as a "Plan B" for true experiments (Oktay, Taylor, and Jensen 2010). The lack of random assignment means that not all variables are held constant, thereby making causal claims more complicated (Sani and Todman 2006; Oktay, Taylor, and Jensen 2010).

There are two types of quasi experimental designs: (a) comparing the same unit before/after an intervention, and (b) comparing different groups that are naturally subject to different conditions[183]

There are three types of true experimental designs: between-subjects design, within-subjects design, and mixed-design (Broota 1989). None of the designs are superior to the others as they come with different strengths and weaknesses.

5.8.2.1. Within-Subjects Design

In within-subject experiments,[184] each subject is used in both experimental group(s) and control group. The main advantage of within-subject design is that it is possible to hold personal traits constant (Bordens and Abbott 2010). Individual differences do not have any influence on the outcome, as the same individual is exposed to both experimental and control treatment.

Within-subject design is a good research design when subjects are expected to have significant individual differences, or when personal traits are unknown and, thus, cannot be controlled (Sani and Todman 2006). A related advantage of the design is that it requires few subjects due to the reuse. The main disadvantage of within-subject design is carryover effects, occurring "when a previous treatment alters the behavior observed in a subsequent treatment" (Bordens and Abbott 2010, 306). For example, if exposure to the experimental treatment alters the subject, she cannot meaningfully be placed in the control group subsequently. The carryover effect is a potential source of error because it is difficult to estimate the effect when designing the experiment.

5.8.2.2. Between-Subjects Design

In between-subject experiments, each subject is only exposed to one condition (Sani and Todman 2006; Bordens and Abbott 2010). Since each subject is only

---

(Oktay, Taylor, and Jensen 2010). The advantage of quasi-experimental design is that it is quick and cheap. In part because extra data collection often can be avoided, and because there is no need for an experimental platform (Oktay, Taylor, and Jensen 2010). In online settings, any change to a website can be treated as an natural field experiment (Oktay, Taylor, and Jensen 2010). This allows for efficient experiments, as changes already made to a website are treated as an experiment.

[184] Within-subject design is also known as repeated measures design

used once, carryover effects cannot compromise the experimental validity. A compelling aspect of between-subject experiments is that it is relatively simple to design and analyze (Bordens and Abbott 2010). The main disadvantage of the between-subject is that the research design is inefficient in terms of subject use. When each subject is only used once, more subjects are needed to detect significance.

5.8.2.3. Mixed Design / Split-Plot Design

In a mixed experimental design, sometimes called split-plot design (Broota 1989; Bordens and Abbott 2010), some factors are between-subject and others are within-subject. The between-subject part means that the subjects are randomly assigned to different groups, and the within-subject part means that the subjects are measured twice (Bordens and Abbott 2010).

The advantage of this design is that it controls for individual differences by measuring the dependent variable before and after intervention. At the same time, potential carryover effects do not influence the experiment as each subject only is assigned to one condition (Bordens and Abbott 2010).

The disadvantages of this design is that it is an expensive and sometimes infeasible research design due to the need for pre-intervention data. Furthermore, many subjects are needed because each subject only is assigned to one condition.

5.9. Experimental Design in Online Settings

As described in Chapter 1, IS and HCI have traditionally relied on reference disciplines such as psychology, social psychology, sociology, and statistics for knowledge about theory, research design, methodology, and data analysis (Lazar, Feng, and Hochheiser 2010). These reference disciplines are aimed at offline, not online, contexts. Some research practices have yet to be translated fully into online contexts.

	Offline	**Online**
Laboratory	Laboratory Experiment	Online Lab Experiment
Field	Field Experiment	Online Field Experiment

Table 20: Experimentation In offline and online environments

Table 20 illustrates two dimensions of experimentation: laboratory/field and online/offline. Both laboratory and field experimentation has been studied intensively in methodological literature, but not in online contexts (Crook et al. 2009).[185] The vast majority of experimental research is still concerned with offline laboratory and field experiments.[186] Despite the adoption of experimentation in both academia and practice, methodological research is lacking. In the words of Crook et al. (2009), there are significant opportunities in examining experimentation in online settings as "The web provides an unprecedented opportunity to evaluate ideas quickly using controlled experiments" (Crook et al. 2009: 1). In the following sections, the focus is on online field experiments, as this method is used in this dissertation. But it is compared and contrasted with the other types of experiments in order to understand its strengths and weaknesses.[187]

---

[185] Crook et al. (2009) ague that "the theoretical aspects of offline controlled experiments have been well studied and documented, the practical aspects of running them in online settings, such as web sites and services, are still being developed." (Crook et al. 2009: 1)

[186] This research gap is illustrated by the fact that Lazar et al. (2010), *Research Methods In Human-Computer Interaction* does not address experiments in online settings.

[187] In the *laboratory experiment*, the experiment is conducted in an artificial setting (Kim, Barua, and Whinston 2002). Experimental subjects are recruited to a physical laboratory, where the experiment is conducted. Laboratory experimentation is often used in both social psychology and HCI research (Wolf et al. 1989; Levitt and List 2007).

A *field experiment* is conducted in a natural occurring setting (Kim, Barua, and Whinston 2002). A field experiment "uses data or an information set collected from the real world or businesses where some variable is directly manipulated in an otherwise naturally occurring process." (Kim, Barua, and Whinston 2002: 218). A problem with field experiment is limited control over nuisance variables.

An *online laboratory experiment* describes an experiment in an artificial internet environment. In this experimental design, an online environment is created, in which users are exposed to different levels of the independent variable and measured on the dependent variable. This is a good alternative to laboratory experiments when examining behavior in online settings. Dabbish et al.

When studying online communities and open government, the internet is not merely a laboratory; it constitutes a field to be studied. Controlled experiments can be conducted online in what Kim, Barua, and Whinston (2002) call virtual field experiments.[188] The main difference between online laboratory experiments and online field experiment is that the latter takes place in a real environment with real users. If desired, the experiments can often be conducted without the users knowing it. Through rather simple scripts or professional experimental software[189] it is possible to create large-scale controlled field experiments at minimal cost.[190] As described by Nicklisch and Salz (2008), "A field experiment in virtual worlds combines some of the advantages of field experiments and of laboratory experiments." (Nicklisch and Salz 2008: 7). Through online field experiments, it is possible to conduct controlled experiments in the systems people are using:

> *"While offline evaluation can be useful for narrowing down the space of options to test, ultimately these options must be tested on live traffic in order to evaluate how well a particular parameterization of an algorithm works in practice."*
>
> (Tang et al. 2010: 17)

(2012) use an online laboratory experiment to examine how member turnover rate in online groups affects social presence and participation. The researchers create an artificial online setting by creating a Tetris game as a Facebook app. Subjects are recruited to play this game and randomly assigned to different treatments.

[188] 'Online' is here preferred to 'virtual', as virtual is linked to virtual worlds such as Second Life.

[189] Examples of such software include Google Website Optimizer (http://google.com/websiteoptimizer) and Visual Website Optimizer (http://visualwebsiteoptimizer.com).

[190] Tang et al. (2010) give an example of an online field experiment: "An experiment in web search diverts some subset of the incoming queries to an alternate processing path and potentially changes what is served to the user. A control experiment diverts some subset of incoming queries, but does not change what is served to the user" (Tang et al. 2010, 19).

5.9.1. Comparing Experimental Designs

The four experiments conducted in this dissertation are all online field experiments. To understand the strengths and limitations of this methodology (Aronson and Carlsmith 1968; Kim et al. 2002; Reis and Gosling 2010), it is compared and contrasted with the other types of experiments, cf. Table 21. This comparison is fruitful as the literature on offline experiments is much more developed than that of online experiments (op cit.).

Parameter	Laboratory	Online Lab	Field	Online Field
Experimental Settings	Artificial	Artificial	Natural	Natural
Control	High	High	Low	Medium
Costs	Cheap	Cheap	Expensive	Cheap
Scale and sample sizes	Low	High	Low	High
Technical Errors	Low chance	High chance	Low chance	High chance
Experimental Awareness	High	High	Low	Low-High
Internal validity	High	High	Low	Medium
External validity	Low	Medium	High	High
Construct validity	High	High	Medium	Medium

Table 21: Comparative advantages of laboratory and field experiments in online and offline settings

5.9.1.1. Experimental Settings

The main difference between field and laboratory environments is the experimental environment and realism. Realism is important because artificial settings might compromise the validity (Lee, Barua, and Whinston 1997). To describe different aspects of experimental realism, Aronson and Carlsmith (1968) use the terms mundane and experimental realism. *Mundane realism* is concerned with the connection between the experiment and the subject's real life (Aronson and Carlsmith 1968). Field experiments have higher mundane realism than laboratory experiments, as the experiments are conducted in real-life settings where the subjects are carrying out their tasks in real life (Kim, Barua, and Whinston 2002).

Experimental realism refers to whether or not the test subjects take the experiment seriously (Aronson and Carlsmith 1968). In field experiments, the subjects are likely to take the experimental seriously; especially in cases where they do not know they are part of an experiment (Shani and Gunawardana 2011).

5.9.1.2. Control over Experimental Setting

When conducting experiments, all variables except the examined one, must be held constant. Whether it is offline or online, the level of control is higher in a laboratory than in a field (Kim, Barua, and Whinston 2002). Therefore, laboratory experiments are easier to replicate than field experiments, as more factors are controlled. The experimental environment in the field is more likely to change over time, making it time to hold enough variables constant for reliable replications.

One potential challenge in online field experiments is keeping different groups separate (Anderson and Simester 2011). There are two risks here: *Firstly*, if that the same user accesses the website from multiple computers, she might be placed in multiple groups (Reips 2002a). This depends on the assignment technology. *Secondly*, subjects from different groups might communicate, potentially undermining the independence of the different treatments (Oktay, Taylor, and Jensen 2010).

5.9.1.3. Costs

In offline settings, the laboratory is often preferred over the field due to lower costs (Kim, Barua, and Whinston 2002). But in online settings, this gap is reduced or even reversed. Creating an online laboratory and recruiting users is potentially more expensive than using an existing website with existing users. Furthermore, experimentation in general is becoming cheaper due to technological innovations (Thomke 1998; Thomke 2001).

5.9.1.4. Scale and Sample Size

Related to the economic costs is scale. One of the challenges with offline experiments is lack of scale. Many studies are limited to 20-30 subjects because it is expensive to recruit and conduct the experiment. Online experimentation scales better in both laboratory and field. Once the experimental conditions are defined, it is inexpensive to conduct experiments on large n-groups.

5.9.1.5. Technical Errors

Whenever technology is used, errors can occur. Technical errors are likely to occur in online than offline experiments (Reips 2002b). Seemingly small technical errors can create significant biases that create or counterbalances effects of an experiment. Some of the events that can occur during an online field experiments are server breakdown, technical errors, spammers, trolls etc. If accidents are not observed or reported, they might create an error source. Reips (2002a) lists a number of potential technical pitfalls one has to be aware of: unprotected directories, revealing experimental details through URL structures, and the technical variance inherent in the internet. Such effects are, however, likely to be cancelled out by randomization insofar the experiments are true experiments.

5.9.1.6. Experimental Awareness

Whereas subjects in laboratory experiments are highly aware that they are part of an experiment,[191] field experiment subjects might not know it (Kim, Barua, and Whinston 2002). Moreover, the great distance between researcher and subjects in online field experiments decreases the likelihood of biases inflicted by the presence of the experimenter, e.g. experimental demands bias (McGrath 1995) and desirability bias (Fisher 1993).[192] On the other hand, the proximity between researcher and subject increases the risk of undetected errors, accidents, and spurious measures.

5.9.1.7. Internal Validity

An experiment is only valuable insofar it is valid.[193] Internal validity is concerned with the translation of hypotheses into research design, and, thus, with the

---

[191] Although not necessarily of the researcher's actual agenda, as deception is often used.

[192] As noted by Reips (2002a; 2002b), one can unintendedly create experimental awareness in online experiment by giving away technical cues such as logical URL structures (e.g. domain.com/experiment-A).

[193] Three types of validity are discussed in the following sections: internal, external and construct.

strength of the causal claims (Campbell and Stanley 1963; McGrath 1995). Due to the increased control over experimental setting and closer proximity between researcher and subject, the internal validity is usually lower in field experiments than in laboratory experiments (Kim, Barua, and Whinston 2002).

The greatest threat to internal validity comes from changes in the dependent variable caused by nuisance variables (Sani and Todman 2006). Two types of such errors can occur: random errors and systematic errors. *Random errors* create noise in the data, but do not compromise the logic of the experiment. *Systematic errors* are worse than random errors, because they create biases in the study by favoring certain treatments on one or more dimensions, thereby giving misleading results. Randomization helps convert nuisance variables from systematic to random, and thus avoid biases that elude the causal claims (McGrath 1995; Bordens and Abbott 2010). Randomization is therefore used in all four experiments conducted and reported in this dissertation. Chance always allows a non-random distribution of nuisance variables. To find out whether or not changes in the dependent variable are caused by changes in the independent variable or random noise, statistical tests are applied to the data set (McGrath 1995). The selection of statistical tests is further discussed below.

It is tempting to include many constructs in experiments. It makes the experiment more realistic and increases external validity. But complicated experiments are both harder to design and analyze and more prone to errors (Anderson and Simester 2011). The four experiments conducted in this dissertation are kept simple to reduce errors and mistakes. Each experiment only includes one independent variable. The downside to simple designs is that they increase the likelihood of unaccounted for spurious measures. Spurious measures can, however, often be ruled out through a careful research design. Conducting experiments in fields rather than laboratory increases the chance of discovering both spurious measures and the synergistic effects one may encounter in real settings.

5.9.1.8. External validity

External validity is concerned with how the results of an experiment "can be generalized to different contexts and individuals" (Sani and Todman 2006, 34). External validity is important, but it is only valuable insofar the experiment is internally valid. According to Sani et al. (2006), three types of external validity exists: ecological validity, population validity, and temporal validity.

5.9.1.8.1. Ecological Validity

Ecological validity is concerned with how the experimental setting translates into natural settings. Field experiments are usually considered to be fairly ecologically valid, as they are conducted in real settings with people acting as they would otherwise (Bainbridge 2004; Kim, Barua, and Whinston 2002).[194]

Conducting experiments in existing and normal online settings increases external validity: "the experiment that provides the strongest evidence as to the true value of the system is an online evaluation, where the system is used by real users that perform real tasks" (Shani and Gunawardana 2011, 267)

5.9.1.8.2. Population Validity

For an experiment to hold population validity, the subjects need to be sufficiently similar to people in other settings. Field experiments often have a more diverse sample population than laboratory experiments, as the subject sample in the latter often comprises college students, thereby being biased toward traits such as young, rich and white (Wilson, Aronson, and Carlsmith 2010). The field experiment subject sample might not reflect the overall population of the entire country but is likely to more accurately reflect the people in the setting. As discussed further below, one caveat with online field experiments is that there might not be sufficient data about the subjects to evaluate the representativeness of the

---

[194] According to Bainbridge (2004), "Field studies thus have the benefit of greater generalizability because they take place in people's normal settings, but they have the disadvantage of lower control over possible intervening factors" (Bainbridge 2004: 642).

sample. So although the sample composition is more diverse than in other experiments, it is also more complicated to assess and evaluate.

As described in Chapter 3, the aim of this dissertation is not to make any empirical generalizations to the entire Danish population or to all recipients on flex job and early retirement pension. Following Yin (2002), the aim is to make analytical generalizations to the open government community.

5.9.1.8.3. Temporal Validity

One of the great unknowns of online experiments is temporal validity. The internet is changing quickly, which shortens the temporal span in which an experiment is valid (Fogg 2003). Consequently, experiments ought to be replicated after some time to see if the findings still are valid. At this point, it is unclear how long the time gap between replications ought to be.

5.9.1.9. Construct Validity

Construct validity is concerned with the correspondence between theory and empirical domain, and how the theory is translated in the research design and measurements (McGrath 1995, Sani et al. 2006). In the experimental research designs applied in this dissertation, four different social psychology theories are used.[195] To reduce the risk of compromised construct validity, the designs are 'one theory, one experiment' designs. However, one theory can be conceptualized in many different ways, and one risk of the experiments is that the theoretical construct only grasps some parts of the theory. Another risk with such simplicity is that the design might be too simple and thus fail to grasp the essence of a problem (Lee, Barua, and Whinston 1997).

5.10. The Experimental Process

As noted above in Figure 13, Thomke (1998) proposes an experimental design model with four steps. These four steps are concerned with experimentation on

---

[195] Social comparison theory, goal-setting theory, self-efficacy theory, and social identity theory.

a high-level; from idea to evaluation. In terms of academic experimentation, Abdi et al. (2009) adds two steps to experimentation focusing on subject selection and assignment. The six steps in an experiment are:

(1) Based on the overarching research question, research hypotheses are formed; (2) subjects are selected and recruited to the experiment; (3) the subjects are randomly assigned to one of the conditions; (4) the experimental intervention is conducted thy manipulating the independent variable. The subjects assigned on an experimental group are exposed to one variation, while subjects assigned to a control group are exposed to another; (5) the subjects' behavior is tracked and logged; and (6) the results are analyzed and evaluated (Abdi et al. 2009; Kohavi et al. 2009).

The four experiments conducted in chapter 6-9 follow these six steps. Each step is explained below. As an experiment is rarely designed right initially, pilot tests are conducted in all four experiments before the six steps are completed for the reported experiments.

5.10.1. Research Hypotheses

The foundation of any experiment is a set of hypotheses about the cause and effect of the intervention. Following Agresti and Finlay (1997), the statistical hypotheses[196] are formulated in the following format:

H_0: No significant effect of the intervention
$H_a...H_N$: A certain effect of the intervention is expected

---

[196] The null hypothesis assumes there is no difference in the means from the different groups and, thus, the subjects come from the same population with the same mean. If the null hypothesis is true, the alternative hypotheses cannot be true (Agresti and Finlay 1997). However, one alternative hypothesis can be true, while other alternative hypotheses are false.

5.10.2. Subject Recruitment

There are at least four ways to recruit subjects to online experiments: website, email, survey, and advertisement. The different means of recruitment are closely connected to the intervention and are therefore described in section 5.10.4.

5.10.3. Subject Assignment

As noted earlier, randomized subject assignment is the hallmark of a true experiment. The subject assignment procedure is closely associated with the subject recruitment methodology described below. All experiments in this dissertation rely on random assignment. As the technology behind the randomization varies between experiments, the exact randomization methodology for each experiment is carefully described in chapter 6-9.

5.10.4. Intervention

Control over the independent variable is what sets a true experiment apart from a quasi experiment. This control can be exercised in many ways, and several tools and media can be used to control independent variables. In online settings relevant to this dissertation, four different media fulfill this criterion: website, email, survey, advertisement. When using websites, the subjects are assigned to different versions of a webpage or website. When using emails, the subjects in the different groups receive different emails. When using surveys, subjects assigned to different groups receive different questions/wording. When using advertisement, different subjects are exposed to different ads.

5.10.4.1. Website

Users visiting a website are assigned to different conditions through the use of either server-side or client-side software solutions. The existing methodology literature on online experimentation usually studies interventions conducted directly on websites (Kohavi et al. 2009; Kohavi et al. 2009a; Crook et al. 2009). Of the different ways to manipulate independent variables listed here, website

interfaces is the most well-documented.[197] Companies such as Amazon, Google, and Microsoft have built their own experimental platforms to conduct the experiments. But building of such platforms is too big an investment for small and medium sized organizations, such as most government agencies. Conducting experimental invention via websites is also used in research, e.g. Grimmelikhuijsen (2012).[198] However, assignment to different website variations require coding, and is therefore not used in this dissertation.

5.10.4.2. Email

By sending out different emails with different conditions, it is possible to use emails to cause experimental interventions. Some email software vendors facilitate this process (e.g. MailChimp and EmailVision). If such software is unavailable in an experiment, free online tools such as Research Randomizer (randomizer.org), can do the randomization. After randomly assigning subjects to the different treatments, different mailing lists are created to match these treatments. To track the users' behavior on the website, a combination of web analytics and database logging can be used. These will link the subjects' response to the email (open/no open – click/no click) and then to the subjects' subsequent behavior on the website. No actual coding is necessary, as it is limited to inserting different links in the emails.[199]

---

[197] "A large body of work exists on the proper design and analysis of controlled A/B tests … where two similar populations of users are given different user interfaces, and their responses can be rigorously measured and compared. Despite this progress, it can still be challenging to use these tools effectively." (Rodden, Hutchinson, and Fu 2010: 2396).

[198] Grimmelikhuijsen (2012) examines the assumed link between transparency and trust in government through a laboratory website experiment. Four different websites are set up, and subjects from a citizen panel are randomly assigned to the four variations. Using four different websites is a slightly different approach than most other studies, where different webpages on the same website are used.

[199] The methodology literature on using emails as experimental interventions is scarce. Yet, it is being used quite often in research practice. Several experiments cited in Chapter 4 rely on emails to manipulate the independent variable. For example, Beenen et al. (2005) and Chen et al. (2010) use emails to expose users to social comparison information. They send out different email variations (experimental and control conditions), and subsequently measure users' behavior (dependent variable) on the MovieLens website. Other experiments relying on emails to cause an experimental

5.10.4.3. Survey

Another way to cause changes in the independent variable is through surveys. Using surveys to manipulate the independent variable is the most common way to conduct experiments in political science, as it is an efficient and inexpensive methodology to identify causal relations (Gaines, Kuklinski, and Quirk 2007).[200] As with emails and websites, different versions of surveys can be set up and used as a way to influence experimental subjects. The interesting thing is not only what the subjects answer in the survey, but how they act in the community subsequently.

---

intervention include Brajer & Gill (2010) who use email experiments as a proxy to examine if women use more words than men. The experimental subjects are sent an open-ended email, and the responses are coded qualitatively. Examining volunteer recruitment in Rock Your Vote, Keane (2010) uses an email experiment to test which types of incentives and events volunteer prospects are most receptive to. This research uses email proxy to examine the effectiveness of different organizational core activities. Aiming to find out if US state legislators respond differently to citizens based on their race, Butler and Broockman (2011) set up an email field experiment. The researchers find that legislators respond differently to emails, depending on the name of the sender (traditional "white names" vis-à-vis traditional "African American names). On a related note, Bosch, Carnero, and Farre (2010) and Baldini and Federici (2011) use a similar design to test ethnic discrimination on respectively Spanish and Italian housing rental market. Violi et al. (2011) use email experiments to examine different motivations for joining an online neighborhood watch community. Four different emails highlighting different motivational aspects (egoism, altruism, collectivism, and principlism) are sent out to users. No significant results are found. An interesting aspect of this study is that emails are not just used as a proxy. The system that is being tested usually relies on email invitations.

[200] In political science, survey experiments are used to measure the consequence of priming, wording of questions, and question order (Gaines, Kuklinski, and Quirk 2007). Examples of using survey experiments in political science include Sniderman and Piazza (1993) and Katz (2011). Sniderman and Piazza (1993) use survey to show that reference to race increases stereotype salience, and thus affects further attitude. Katz (2011) uses a survey experiment to test a new design feature on eBay. The subjects are asked the same questions, but see different images at a shopping page. The subjects' answers to the question vary according to the amount of context shown next to each image. In the study by Katz (2011), both the independent and dependent variable are measured in the survey environment.

5.10.4.4. Advertisement

Advertisement and experimentation are no strange bedfellows. For many decades, direct mail advertisement has relied on experimentation. Split testing has frequently been used to test different ad copy, layout, placement, and even paper quality (Lewis and Reiley 2010).

Surprisingly little literature exists on paid advertisement in online settings. In the perspective of this dissertation, advertisement is an effective way to conduct the experimental intervention and randomly assign new users to different treatments. Through online advertisement platforms such as Google AdWords, Facebook Adverts, and LinkedIn Ads, it is possible to serve different variations of an advertisement to different users.[201]

Using online advertisement platforms as intervention comes with three challenges: (1) creating groups of equal size, (2) few demographic data, and (3) lower user satisfaction. *First*, One downside to advertisements is that advertisement platforms such as MySpace and Google AdWords[202] seek to optimize earnings by showing ad copies with higher CTR to users more often than ad copies with low CTR (Dow et al. 2010). This entails that the different variations are not necessarily served consistently over time, meaning that some treatments are more likely

---

[201] Examples of research relying on online advertisement as part of experimentation include Guerini, Strapparava, and Stock (2010), Chan (2011), and Dow et al. (2010). Guerini, Strapparava, and Stock (2010) use Google AdWords to measure the persuasiveness of a message. The main benefit of using Google AdWords as the dependent variable is the low costs and high accessibility, as it "allows conducting experiments with thousands of subjects for a few dollars in a few hours, by tweaking and using existing commercial tools for advertising on the web, such as Google AdWords" (Guerini, Strapparava, and Stock 2010: 3459). In an attempt to encourage university students to connect to the university library's Facebook page, Chan (2011) find Facebook ads to be an effective marketing channel. Dow et al. (2010) use paid advertisements on MySpace (http://myspace.com) as the dependent variable in an experiment on different advertisement prototyping processes. The effectiveness of the ads is measured through CTR, and two post-click variables: time on site and pageviews.

[202] Although Google AdWords offers a feature that rotates ads more evenly (described further in chapter 9), ads with higher CTR are still favoured by the algorithm. However, the bias is smaller.

to get subjects than others. This does not comprise the validity of the experiment, as the subject assignment is still random. All subjects have the same likelihood of being assigned to the treatments, but the different treatments do not have the same chance of receiving subjects.

Secondly, platforms such as Google AdWords offer few data, limited to language and location. Some advertisement programs only provide scarce demographic data, making it hard to assess the representativeness of the subject sample, and thus the external validity of the study. In comparison, Facebook Adverts offer many demographical data such as age, gender, and location.

A *third* challenge with using online advertisements is that citizens might find them obtrusive, which can lead to a decline in satisfaction with government (Teerling and Pieterson 2011). To examine potential effects like this, follow up surveys can be helpful.

5.10.5. Tracking Behavior

The four different intervention strategies have different opportunities and limitations in terms of data tracking. In three of the experiments, the primary data source is a MySQL database, in which most user behavior is logged and stored. This data needs to be supplemented with a cross platform tracking tool when the intervention is conducted through a proxy. In the four experiments, Google Analytics[203] is used to track across platforms. By using custom URL parameters (Google 2011), subjects are tracked from the proxy to K10. This enables tracking of different user behavior across groups. This tracking methodology is good, but not perfect. Likely sources of error include URL sharing[204] and URL cleaning.[205] It should be noted, that both sources of errors are unlikely to be severe enough to comprise the data validity.

---

[203] Google analytics is free web analytics software provided by Google, http://google.com/analytics.

[204] If subjects share a URL containing custom parameters with others, noise will be inflicted in the data.

[205] If users remove parts of the URL string, the page is still available but the tracking is disabled.

5.10.6. Data Analysis

A potentially critical error source in experiments is variance caused by other factors than the independent variable. The best way to detect such variance is through statistical tests.

5.10.6.1. Parametric Assumptions

It is desirable to use parametric tests when the dataset allows it, as these tests are more sensitive to detecting significant difference than non-parametric tests (Purchase 2012). Parametric tests have three assumptions (1) random sampling from the population; (2) normally distributed sample; and (3) homogenous within-subject variance (Bordens and Abbott 2010; Gravetter and Wallnau 2012). Normal distribution is satisfactory when each group has n > 30 subjects due to the central limit theorem (Feld 2009). If a data set lives up to all three assumptions, a parametric test should be used. If not, non-parametric tests are to be applied, as they are assumption-free tests (Field 2009). In borderline cases both parametric and non-parametric tests will be used.

Meeting parametric assumptions is often challenging in the online community literature as participation usually follows an inverse power law distribution causing zero-inflated data (See for example Dabbish et al. (2012) and Farzan et al. (2012). But as long as n > 30 subjects in each group, the central limit theorem is fulfilled and the data is assumed to be normal.

5.10.6.2. Selecting Statistical Test

The dependent variables examined in all experiments are continuous variables. In comparison, the independent variables are categorical. Combined with the experimental design, these variable choices determine which statistical tests are used (Feld 2009).

In this dissertation, all three types of true experiments (op cit.) are applied: between-subject design, within-subject design, and mixed design.[206] The three different designs have to be analyzed using different statistical tests.

5.10.6.2.1. Between-Subject Design

When using between-subject design, the preferred statistical test is ANOVA (analysis of variance). This test examines the variation among subjects in each group and compares this variation to the between groups variations (Feld 2009). ANOVA is a parametric test and can only be used insofar the data meets parametric conditions and/or the central limit theorem (Howell 2010). When the data fails to do meet the parametric assumptions, the Kruskal-Wallis Test is used as an alternative (Feld 2009).

5.10.6.2.2. Within-Subject Design

In the within-subject design, each subject is used in all conditions. To analyze this data, the paired-sample t-test[207] is used when the parametric assumptions are met. When these assumptions are not met, the non-parametric version of this test, the Wilcoxon signed-rank test,[208] is used. The aim of these tests is to see if the means differ across the different treatments. The test logic of both tests is identical, with the notable differences that data values are converted into ranks in the Wilcoxon signed-ranks test (Feld 2009). In the Wilcoxon signed rank test, all data is treated as ordinal data. The test creates ranks for each value based on the differences between the before and after conditions, and uses these ranks to calculate significance.

5.10.6.2.3. Mixed Design / Split-Plot Design

In the mixed design, subjects are assigned to different conditions and measured repeatedly. If the parametric assumptions are met, the repeated measures general linear ANOVA is preferred (Feld 2009). If case assumptions about linearity

---

[206] The reasons for choosing the different designs are provided in chapters 6-9.

[207] This test is sometimes referred to as the dependent t-test (e.g. Feld 2009).

[208] This test is sometimes referred to as the Wilcoxon Matched-Pairs test (e.g. Feld 2009).

are not met, the repeated measures general linear ANOVA cannot be used, even when the central limit theorem is met (Feld 2009). When linearity assumptions are not met, the most appropriate method to analyze paired data points is the paired-sample t-test and the Wilcoxon signed ranks test (Field 2009; Bordens and Abbott 2010). Compared to the repeated measures general linear ANOVA, the paired-sample t-test and the Wilcoxon signed ranks test can only detect significance within groups, but not across groups.

5.10.6.3. Alpha Level

In both psychology and HCI research, 0.05 is the conventional α-level (Abdi et al. 2009; Feld 2009). Therefore, statistical results with a p value ≤ 0.05 will be considered statistically significant, while results with a p value > 0.05 will be considered non-significant.[209]

5.10.6.4. Effect Size

As pointed out by Ellis (2010), effect sizes ought to be reported separately from significance tests. For small samples, a significance test might show that the experimental intervention yields non-significant effects, but nevertheless show a large effect size. This indicates that further research with larger samples might be useful, as this might reveal a significant effect.

Cohen's *d* is used to calculate effect sizes for *t*-tests (e.g. paired sample *t*-tests) and F-tests (e.g. ANOVA) (Howell 2010). Cohen's *d* is calculated by taking the difference between two means divided by a standard deviation (d = M1 - M2 / σ)

---

[209] An important part of statistical tests is the choice of significance level. Although statistics is considered to be a "hard science", the alpha level (α) is a somewhat arbitrary and subjective measure (Sani and Todman 2006; Abdi et al. 2009). The two standard α-levels used in experimental research is either α = 0.05 or α = 0.01 (Agresti and Finlay 1997). The choice of α-level is an inherent trade-off between Type I errors and Type II errors. When selecting a α-level at 0.05, the likelihood of a Type II error is reduced, while the likelihood of a Type I error is increased. On the other hand, a stricter α-level at 0.01 will reduce the likelihood of a Type I error while increasing the chance of a Type II error.

(Cohen 1988; Rosnow and Rosenthal, 1996). Cohen (1988) defines effect sizes as: 0.2 = small; 0.5 = medium; and 0.8 = large.

The effect size r is used to calculate the Wilcoxon signed ranks test. Effect size r is calculated using this formula: $r = Z/\sqrt{n}$.[210] Effect sizes range from 0-1, and Cohen (1988) describes effect ranges as: 0.10 = small; 0.30 = medium; 0.50 = large.

5.11. Chapter Summary

In this chapter the ontological, epistemological, methodological and axiological choices are described. This dissertation is firmly rooted in ontological realism, while the epistemological stance is characterized by positivism, objectivism, and being an outsider. This entails that the purpose of research is to identify causal relationships between observable events. These fundamental choices constitute the framework for the dissertation and guide the methodological choices. In terms of methodology, this dissertation is quantitative, exploratory, deductive, field based, and nomothetic. On the axiological level, relevance is prioritized over rigor.

The empirical research comprises four online field experiments. All four experiments are true experiments. Three different experimental research designs are used in the data collection: within-subject design, between-subject design, and split-plot design. In order to conduct the experiments with only few resources available, four different proxies are used to cause the experimental intervention:

---

[210] There is some confusion in statistics textbooks on how to calculate the effect size r for a Wilcoxon signed ranks test. There is consensus about calculating $r = Z/\sqrt{n}$. The disagreement is concerning n. In a Wilcoxon signed ranks test there are two observations per participant (one before, one after). Field (2009: 588) argues that n = total observations, while Corder and Foreman (2009: 59) argue that n = total participants. A plausible explanation is that Field (2009) is wrong, as he writes "The effect size [for Wilcoxon signed ranked test] can be calculated in the same way as for the Mann-Whitney test." (Field 2009: 558). However, the effect size in the Mann-Whitney test is based on observations, not participants. Thus, I use the interpretation suggested by Corder and Foreman (2009).

email, internal message system, online survey, and Google AdWords advertisement. The applied statistical tests used in the data analysis are described. Depending on the data, ANOVA, Kruskal-Wallis, dependent sample t-test, and Wilcoxon Signed Ranks test are used. α-level at 0.05 is used as criterion to evaluate the test results. The tests are followed up with effect size tests relying on either r or Cohen's *d*, depending on the data.

Finally, the four field experiments are subject to both privacy laws and ethical considerations. In terms of *privacy*, the data is collected and stored in a way that ensures it is fully compliant with the EU *Data Protection Directive* and the Danish *Act on Processing of Personal Data*. The *ethical* guidelines and practices are less clear in online than offline research. In order to maximize the realism and validity of the experiments, it is chosen to conduct the experiments without asking for users consent. Although this decision is contestable, it follows the practices of much online experimentation.

Part 3

The Four Experiments

Chapter 6: How Social Comparison Information Affects Participation - An Email Experiment

6.1. Introduction

It is often difficult to know how often to perform certain tasks, e.g. posting, in an online community; for there is no objective "right amount" of activity, such as a daily, weekly, or monthly post. Due to the lack of objective and stringent yard-sticks people will turn to their peers for comparison of behavioral patterns (Festinger 1954). One way to determine the socially accepted and appropriate contribution level is to compare one's own behavior to that of others. As described in section 4.4.2, social comparison theory contends that people compare themselves to others in order to help identify appropriate behavior and evaluate capabilities, opinions, and actions.

The purpose of this experiment is to detect if exposing users to social comparison information will affect subsequent participation. If social comparison theory can be successfully applied to K10, it might be a way to alter and improve the participation patterns in open government communities. For details on how social comparison theory is operationalized into a research question and hypotheses, refer to section 4.4.2.2.

This chapter offers three contributions. The *first* contribution is knowledge about how an open government community – K10 – reacts to a certain construct of social comparison theory. Such knowledge makes it possible to design more effective information systems. The *second* contribution is a methodological contribution to the IS community, resulting from the application of email as an experimental instrument. *Finally*, this study is partly a replication of Chen et al. (2010). This makes it possible to compare and contrast the results of social comparison theory across two different communities: MovieLens and K10.

6.1.1. Research Question and Hypotheses

This experiment seeks to answer the question of how exposure to social comparison information affects users' participation in K10?

As the research is relying on experimental design and statistical analysis, the research question is translated into seven statistical hypotheses:

H_0 Receiving social comparison information has *no* effect on subsequent participation.

H_a Receiving social comparison information has an effect on subsequent participation

H_{a1} Receiving social comparison information reduces the contribution level among above-means contributors

H_{a2} Receiving social comparison information increases the contribution level among below-means contributors

H_b Receiving non-comparative information has an effect on subsequent participation, but smaller than comparative information.

H_{b1} Receiving non-comparative information increases the contribution level among above-means contributors

H_{b2} Receiving non-comparative information increases the contribution level among below-means contributors

In this study the *independent* variable is different levels of participative information, and the *dependent* variable is number of contributions to K10. The effect of the manipulation is measured over a two-week period before and after the experimental intervention.

6.2. Research Design

The research design is outlined in the following sections. It addresses the experimental methodology, subject selection and assignment, and statistical data analysis.

6.2.1. Split-plot Design

This experiment is a split-plot design, entailing a between-subject element as well as a within-subject element. Split-plot design, as shown in Table 22, is a good fit to both theory and research question. It manages to track before and after contribution levels while assigning subjects to different conditions.

	Pre-intervention contribution level	Post-intervention contribution level
Group 1	T_0 observation	T_1 observation
Group 2	T_0 observation	T_1 observation
Group 3	T_0 observation	T_1 observation

Table 22: Split-plot research design used in experiment 1. Adapted from Bordens and Abbott (2010)

To answer the research question of this experiment, the subjects are divided into three groups. Each group is treated with a different intervention, which comprises the between subject variation in the experiment.

Group 1   Receives an email with comparative participation information

Group 2   Receives an email with non-comparative participation information

Group 3   Receives no email

There are two different experimental groups who each receive an email. The second experimental group is included in part to examine if comparative information is more effective than non-comparative information, and in part to investigate if receiving non-comparative information is more effective than receiving no information.

		Pre-intervention contribution level	Post-intervention contribution level
Group 1	Above mean	T_0 observation	T_1 observation
	Below mean	T_0 observation	T_1 observation
Group 2	Above mean	T_0 observation	T_1 observation
	Below mean	T_0 observation	T_1 observation
Group 3	Above mean	T_0 observation	T_1 observation
	Below mean	T_0 observation	T_1 observation

Table 23: Split-plot research design for experiment 1, segmented between above and below means subjects

As described in section 4.4.2, one strand of social comparison theory predicts regression toward the mean. This entails that the *above-means contributors* are expected to contribute less post-intervention while *below-means contributors* will contribute more. To examine this effect, the subjects are segmented into *above-mean* and *below-mean* contributions. This gives six groups in the second part of the analysis, as illustrated in Table 23.

6.2.2. Sampling and Subject Assignment

At the time of the experiment, K10 had 8,810 registered users. Since K10 was launched in 2005, a total number of 3,959 users have posted one or more post in the community.

As mentioned in section 5.9, online field experiments scale well compared to their offline counterparts. The costs of conducting an experiment with 1,000 subjects are almost identical to conducting an experiment with 10 subjects. In the case of K10, there is no economic rationale for drawing a sample instead of using all community members as experimental subjects.[211]

---

[211] Some online communities have huge populations, making a study of all members impracticable (e.g. Marlow et al. 2009; Chen et al. 2010). To make the data collection and analysis more manageable, a sample is drawn. However, sampling carries the risk of biases, making it preferable to conduct a study on the entire population whenever it is possible (Bryman and Cramer 2005).

Including all community members in the study, and not just the active users, increases the likelihood of Type I errors (accepting the H_0 hypothesis when it should have been rejected) because it entails increased within-group variability.[212] Including all users also creates the risk of ending up with a zero-inflated dataset, unfit for parametric tests. However, as long as each group has n > 30 subjects, it meets the assumptions in the central limit theorem and thereby parametric assumptions as well (Feld 2009).

The benefit of removing users who have made zero contributions during the pre- and post-intervention period is increased likelihood of detecting significance. The disadvantage is that it limits the scope to those users who contributed, and thus ignores the non-contributors. When examining participation in online communities, contributors and non-contributors alike constitute important data points. As both approaches comprise benefits and disadvantages, the data analysis is conducted on both datasets. In the first part of the data analysis, in section 6.4.4, all subjects are included. This is followed by an analysis of the dataset containing only subjects with at least one contribution in the four-week experimental period.

To reduce the risk of emails being reported as spam and limit the disturbance of the experiment, emails are only sent to K10 users who have written one or more post in their K10 lifetime. The 3,959 users are randomly assigned to one of three conditions. Each group has approximately 1,320 users. By using email click through rate (CTR) benchmarks it is possible to estimate the number of emails clicked and thereby the number of experimental subjects. Epsilon (2011) reports the average CTR across all industries to be 5.5 %. MailChimp (2010) reports 2.85 % CTR in emails sent out in the social networks and online communities' niche (Mailchimp 2010). With approximately 1,320 users in each experimental group and assuming a CTR between 2.85 % and 5.5 % each experimental group is expected to have between 38 and 72 subjects.

---

[212] The within-group variability is increased because many observations from the inactive users are likely to be 0.

As the section on split-plot design above reveals, each group will be segmented into two subgroups, thereby making the group sizes approximately half the size. This means that some of the groups are likely to have fewer than 30 subjects and not meet the assumptions of parametric tests. To increase the likelihood of finding a statistically significant effect, and thus reducing type I errors (missing an effect), no sampling will be made (Abdi et al. 2009). This means the sample in this experiment comprises all users who made at least one contribution in the past.[213]

For an experiment to hold internal validity, random assignment of subjects is crucial (op cit). To ensure random assignment of subjects in this experiment, three MySQL queries are executed in the database (The queries can be found in Appendix 5). The first query assigns all users to *group 1*. The second query assigns 50 % of the users to *group 2*. The third query assigns 50 % of the members from *group 1* and *group 2* to *group 3*. This approach has created three groups of approximately the same size.

Each query randomizes the subject assignment as all users have the same likelihood of being assigned to each group. The user's user ID in the MySQL database is used as identifier, ensuring that each user only appears in the dataset once.

6.3. Conducting the Experiment
6.3.1. Intervention: Email Design
The subjects in the two experimental groups receive an email, while the subjects in the control group do not receive an email. The subject line of the email is identical in both conditions. The content of the emails is kept short and simple. This helps reduce confusion, and equally important, it makes the experimental ma-

---

[213] One weakness of large samples is that they increase the likelihood of Type II errors, because it is easier to detect significant differences in large samples. To show the impact of the intervention, the effect size for each group is calculated in the data analysis.

nipulation more salient and thereby increases the internal validity of the experiment.[214]

Besides static text, the emails contain three dynamic variables:

- **User name**. This variable personalizes the email.
- **Number of posts**. This variable counts the number of posts per user in the last 30 days. This data is extracted from the MySQL database.
- **Above-mean / below-mean**. This variable is calculated by comparing the number of posts in the last 30 days with the mean number of posts. The mean is calculated to 6.3 posts. This mean is based on the number of posts from active users, and is much higher than the overall population mean and median. Thus, the mean does not reflect the entire experimental population, but merely the active users. A high mean is used to increase potential effects of the intervention.

Screenshot 1 and 2 show the two emails sent out to the subjects in *group 1* and *group 2* respectively.
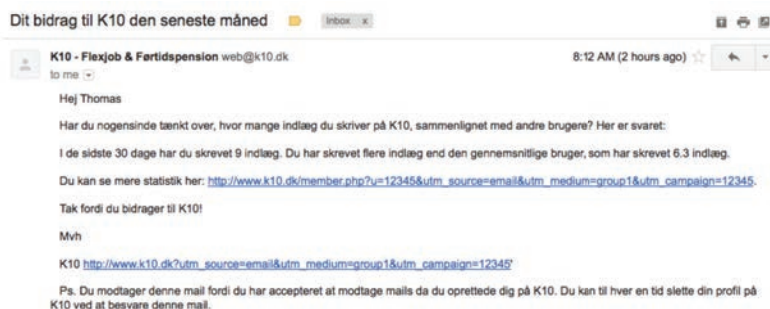
Screenshot 7: The email with comparative information sent to G*roup 1* subjects

---

[214] The wording of the email was pilot tested on three colleagues and three K10 users, and it led to small adjustments of the content.

Dit bidrag til K10 den seneste måned   Inbox x

K10 - Flexjob & Førtidspension web@k10.dk   8:12 AM (2 hours ago)
to me

Hej Thomas

Har du nogensinde tænkt over, hvor mange indlæg du skriver på K10? Her er svaret:

I de sidste 30 dage har du skrevet 9 indlæg.

Du kan se mere statistik her: http://www.k10.dk/member.php?u=12345&utm_source=email&utm_medium=group2&utm_campaign=12345.

Tak fordi du bidrager til K10!

Mvh

K10 http://www.k10.dk?utm_source=email&utm_medium=group2&utm_campaign=12345

Ps. Du modtager denne mail fordi du har accepteret at modtage mails da du oprettede dig på K10. Du kan til hver en tid slette din profil på K10 ved at besvare denne mail.

Screenshot 8: The email with non-comparative information sent to G*roup 2* subjects

The emails were sent on January 11th 2012 at 3 pm CET. A total of 2,684 emails were sent. 1,339 emails for *group 1 subjects*, and 1,345 emails for to *group 2 subjects*.

All users with 1+
post: N = 3,959

2/3 users get emails:
2,684 emails sent out

219 users clicked
in the email
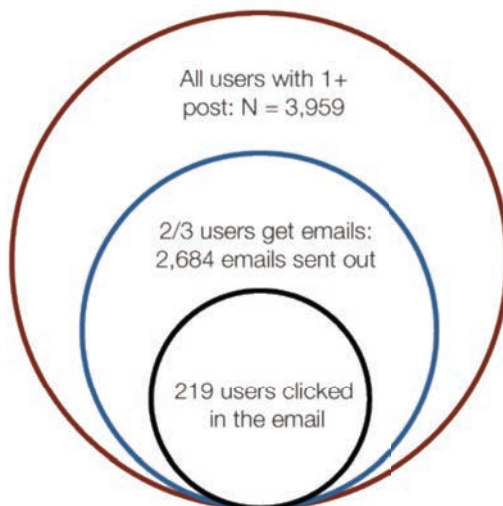
Figure 14: Visualizing of the funnel from K10 user to experimental subject

As Figure 14 shows, of the 2,684 emails sent, 219 users have clicked on a link in the email. This gives a CTR on 8.16 %. Although this CTR might sound low, it is high compared to the industry benchmarks that ranges from 2.85 % - 5.5 % (op cit.).

In the email, the user is informed about the number of posts during the prior 30 days. But the data analysis is limited to 2 weeks before the experimental manipulation, and 2 weeks after the experiment. Although the potential effect over time is unknown, it is very likely that if there is an effect, it will likely be more salient right after the experiment and slowly fade away as the subjects forget about the email. Since this experiment primarily is concerned with identifying potential causality, it makes sense to limit the period to four weeks total.

6.3.2. Tracking and Logging

As illustrated in Figure 15, the data logging begins four weeks prior to the experiment, and ended two weeks after the experiment. The data from the four weeks prior to the intervention is used to calculate the mean number of contributions as well as the individuals' participation within the past 30 days. However, only the data from two weeks before the intervention and two weeks after the intervention is used in the actual data analysis. Two different methodologies are used to track the user metrics: MySQL database logging and javascript logging.
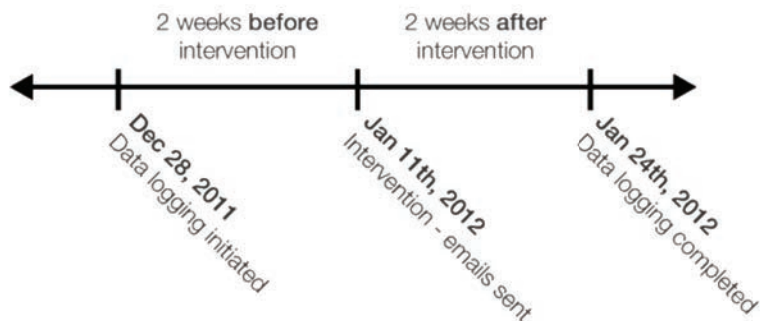
Figure 15: Timeline over data collection in experiment 1

As mentioned above, the 2,684 emails each contained three different dynamic variables. Neither the vBulletin, the forum software that powers K10, nor any known low-priced email provider can conduct these calculations automatically. Therefore, the emails are sent out via a PHP-script directly from the K10 server.

240

The main caveat of this approach is that the script – contrary to state of the art email providers – is incapable of measuring email open rate. Through javascript tracking, it is however possible to track when a user clicks a link in the email. Therefore, click through is used as a measure of the intervention.

If the user clicks a link in the email she is included in the sample. Consequently, some users might be exposed to one of the experimental groups, but is excluded from the dataset. This is not a problem in terms of validity, as the subject is excluded from the dataset rather than transferred to the control group. The downside is that it reduces the group size and makes it harder to detect significance.

6.3.2.1. Database Logging
The MySQL database contains all user data and content. The information in the database relevant to this study is:

- Number of posts
- Email address
- User ID

6.3.2.2. Javascript Logging
As vBulletin does not facilitate email tracking, a Google Analytics javascript is placed on all pages on K10. Logging is activated when the subjects click the link in the email that contains a unique identifier, and javascript logging can therefore only be started at the time of the intervention. Each subject receives an email with a link containing the user's user id. The analytics tracking looks like this *&utm_source=email&utm_medium=group&utm_campaign=$userid*.[215]

---

[215] The link contains three different parameters: (1) &utm_source=email shows the visitor comes from an experimental email; (2) &utm_medium=group shows the user's experimental group (Group 1 or Group 2); and (3) &utm_campaign=$userid shows which user id clicked on the link. The parameter $userid is connected to vBulletin.

Two weeks after the intervention, the data is exported from the database and Google Analytics and merged into a simple dataset that can be analyzed in SPSS.

6.4. Data Analysis

6.4.1. Cleaning the Dataset

A subject started a discussion thread in which the experiment is discussed. [216] 10 subjects from the control group participated in the discussion. As these subjects are exposed to the experimental intervention even though they should not have been, they are removed from the *control group*.[217] Furthermore, the admin of the site was removed from *control group*, as he was well aware of the experiment. Besides these minor adjustments to the dataset, no changes are made.

6.4.2. Sampling Distribution

To determine if the data meets the parametric assumptions, Kolmogorov-Smirnov tests are used on the entire data set and on each of the three groups to check for normal distribution. The distribution across all three groups shows that participation is heavily skewed to the left.

---

[216] The discussion of the experiment can be found here: www.k10.dk/showthread.php?t=21058.

[217] User IDs 573, 2994, 5142, 5196, 5565, 7581, 7686, 8577, 8592, and 8625

Mean = .81
Std. Dev. = 5.755
N = 1,484

Number of posts two weeks prior to intervention
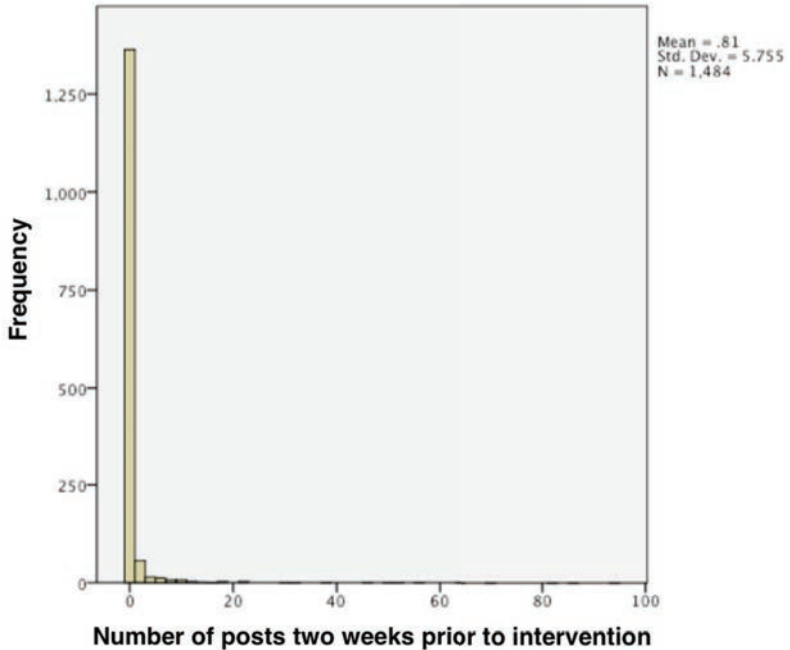
Figure 16: Distribution of posts two week before the experiment among all subjects

The majority of the users, who have contributed at least one post in their K10 lifetime, have written zero posts during the two weeks prior to the experiment. The same pattern can be detected in all three groups. The distribution is skewed so much to the left that the median number of posts in all three groups is 0. The Kolmogorov-Smirnov tests give the following results: for all subjects, K-S(df: 1484): 0.475, p < 0.000. For Group 1 K-S(df: 107): 0.474, p < 0.000; Group 2 K-S(df: 111): 0.406, p < 0.000; and Group 3 K-S(df: 1265): 0.488, p < 0.000.

All three tests show p < 0.000, meaning that the likelihood of normal distribution in the data is smaller than 0.000 %. The most appropriate model for this analysis, repeated measures ANOVA, is a general linear model. As the histograms above show, this data does not fit a general linear model and this model is therefore a poor fit for the data.

However, as all three groups have n > 30, the data does fulfill the central limit theorem, making it suitable for parametric tests. The most appropriate parametric test for this data, *not* assuming linearity, is the paired sample t-test.

When segmenting the experimental groups into above-means and below-means contributors, some of the groups have n < 30 (c.f. Table 26 later in this chapter). In these cases, a non-parametric test is used to analyze the data. The Wilcoxon signed ranks test is the non-parametric test best suited to do the analysis (Wilcoxon 1945). Both tests are conducted by comparing pre-intervention performance with post-intervention performance *in* groups. This entails that the between group differences are not tested. Significant variation in groups is a prerequisite for meaningful variation between groups, which makes this the more important consideration.

Two different results are reported for this experiment. The first part of the analysis comprises all subjects, n = 1,483. The second part of the analysis removes all the subjects who have not written any posts in the four-week experimental period. By removing all the subjects with zero-values, the variance in the data is also removed. This makes it easier to detect significant effects. This data set has n = 165.

6.4.3. Results

Fewer posts were written in the two weeks after the experiment than before. In the two weeks pre-intervention, a total number of 1,195 posts were written. The mean number of posts per user is 0.81. In the post-intervention period, the total number of posts is 1,111, giving a mean post of 0.75 per user over a two-week period. The median number of posts in both periods is 0.

A paired samples t-test across all users shows that there is no significant effect of the intervention across the entire data set, as t(1,482) =0.663, p < 0.507. Effect size Cohen's d = -0.03 effect size. This very low effect size shows that the intervention had basically no effect across all subjects. The effect size is skewed in

part by the different group sizes. Because the control group is so big compared to the two experimental groups, any potential effect in the small groups gets crowded out by the big group.

		Posts	Median	N	Mean	Std. Dev.	Minimum	Maximum
Group 1	Before	492	0	107	4.60	17.320	0	94
	After	406	0	107	3.79	14.326	0	111
Group 2	Before	171	0	111	1.54	5.812	0	50
	After	182	0	111	1.64	4.707	0	30
Group 3	Before	531	0	1265	0.42	3.058	0	56
	After	518	0	1265	0.41	2.971	0	52
Total	Before	1194	0	1483	0.81	5.757	0	94
	After	1106	0	1483	0.75	4.967	0	111

Table 24: Posts per user two weeks before and two weeks after the experiment

One interpretation of this stability in the data before and after intervention is that the data set seems valid, and no spurious variables are affecting the experiment. The next step in the analysis is a pairwise analysis on the group levels: here is can be detected whether or not there are any significant differences in the contribution level in each of the three groups after the experimental manipulation. As all three groups have n > 30, the paired samples t-test is used.

	Mean	Std. Dev.	S.E. Mean	T	df	Sig. (2-tailed)	Effect size
All Subjects	-0.057	3.289	0.085	-.663	1482	0.507	-0.03
Group 1	-0.804	9.329	0.902	-0.891	106	0.375	-0.13
Group 2	0.099	3.258	0.309	0.32	110	0.749	0.05
Group 3	-0.007	2.1	0.059	-0.12	1264	0.904	-0.01

Table 25: Paired samples t-tests of posts per user two weeks before and after the experiment

Table 25 describes the number of contribution across groups. The table shows that the contributions among *group 1* subjects have decreased post-intervention. More specifically, the mean number of posts per user has decreased from 4.60 in the two weeks before the experiment to 3.79 in the two weeks after the experiment. The median is 0 posts per week in pre-intervention as well as post-intervention conditions.

Running a paired samples t-test test on *group 1* shows that there is no significant effect of the manipulation in this group, and only a minor effect size t(106) = -0.891, p < 0.375. Cohen's d = -0.13 indicates only a small change after the intervention. This test suggests that the social comparison data has no significant effect on participation.

Similarly, the paired samples t-test shows non-significant results for *group 2* and *group 3*. In *group 2*, the median is 0 in both the before and after condition. The mean number of posts decreased from 1.64 posts in the two weeks before the intervention to 1.54 posts in a comparable period after the intervention. The test results show no significant result and small effect size t(110) = 0.32, p < 0.749, Cohen's *d* = 0.05. A Cohen's d of 0.05 indicates a small effect size. This test also indicates that the non-comparison participatory information has no significant effect on participation.

For *group 3*, the median is 0 both before and after the intervention. The mean is 0.34 posts per subject in the pre-intervention period and 0.38 posts in the post-intervention period. No significance is detected in the paired samples t-test: $t(1264) = -0.12$, $p < 0.904$, Cohen's *d* = -0.01. The effect size is also here very small. No significance test results in the *control group* are positive, and it indicates that the experiment is valid and largely unaffected by nuisance variables.

When comparing pre-intervention and post-intervention participation in the three groups, no effect of the intervention can be detected. This data analysis seems to suggest that neither social comparison information nor non-comparative information have any effect on subsequent participation.

According to one strand of social comparison theory, people react differently to social comparison information depending on their score relative to the group median (op cit.). To see if this is the case at K10, the data can be separated into two groups: *above-mean*, and *below-mean*.

6.4.4. Above-Mean and Below-Mean Contributors

For *group 1*, it is straightforward to separate the experimental subjects, as some users got an email stating their contribution level were above the mean, while other users got an email saying they were below the mean. The mean used in the email experiment was calculated to be 6.3. As 6.3 posts in the prior month was used in the experimental manipulation, 6 posts serves as the cutoff point.

For *group 2* and *group 3*, there is not an explicitly given cutoff point, as the subjects have not been informed about the mean level of contribution. However, in order to maintain consistency in the analysis, 6 posts in the prior month will also be the cutoff point for *group 2* and *group 3*.

		Posts	Median	N	Mean	Std. Dev.	Min	Max
Group 1	Above mean	Before	492	9	15	32.80	35.779	0	94
		After	387	21	15	25.80	30.713	0	111
	Below mean	Before	0	0	92	.00	.000	0	0
		After	19	0	92	.21	1.043	0	9
Group 2	Above mean	Before	151	6	11	13.73	13.661	0	50
		After	114	9	11	10.36	10.443	0	30
	Below mean	Before	20	0	100	.20	.603	0	3
		After	68	0	100	.68	2.054	0	14
Group 3	Above mean	Before	456	4	48	9.50	12.649	0	56
		After	407	6	48	8.48	12.546	0	52
	Below mean	Before	73	0	1217	.06	.388	0	5
		After	122	0	1217	.10	.648	0	11

Table 26: Posts per user two weeks before and two weeks after the experiment

Table 26 shows that the group sizes of above-mean contributors in *group 2* and *group 3* are n < 30, thereby not fulfilling the central limit theorem. As both these groups are above-means groups, the data is not zero-inflated.

		Mean	Std. Dev.	S.E. Mean	t	df	Sig. (2-tailed)	Effect size d
Group 1	Above							
	Below	.207	1.043	0.109	1.898	91	0.061	-0.91
Group 2	Above							
	Below	0.48	2.082	0.208	2.306	99	0.023	0.38
Group 3	Above	0.48	2.082	0.208	2.306	99	0.023	-0.14
	Below	-1.021	10.317	1.489	-0.686	47	0.496	0.09

Table 27: Paired samples t-tests for the groups meeting parametric assumptions: Posts per user two weeks before and two weeks after the experiment

For *group 1* above-means contributors, the Kolmogorov-Smirnov test of normality shows $D(10) = 0.267$ p < 0.008 for the pre-intervention period and $D(10) = 0.268$ p < 0.007 for the post-intervention period. For *group 2* above-means contributors the Kolmogorov-Smirnov test of normality shows $D(10) = 0.290$ p < 0.017 for the pre-intervention period and $D(10) = 0.198$ p < 0.200 for the post-intervention period. In other words, the Kolmogorov-Smirnov tests of normality show that the number of contributions in neither of the two groups is normally distributed.

Group		Z	Assymp. Sig. (2 tailed)	Effect size R (r = Z/√ n)
Group 1	**Above mean**	-1.006	.315	-0.260
Group 2	**Above mean**	-1.277	.202	-0.385
Group 3				

Table 28: Wilcoxon signed ranks test for the groups *not* meeting parametric assumptions: Posts per user two weeks before and two weeks after the experiment

As the *above-means* segment in *group 1* and the *above-means* segment in *group 2* both have n < 30, they do not meet the central limit theorem. The two n < 30 groups are examined with the Wilcoxon signed ranks test, c.f. Table 28, while the paired samples t-test is used for the four n > 30 groups, c.f. Table 27.

6.4.4.1. Group 1: Social Comparison Information

Above mean subjects: The mean contributions dropped from 32.80 posts per user in the two weeks prior to the experiment to 25.80 posts after. In spite of the lower mean, the median increased from 9 posts before to 21 posts after. This indicates that more subjects in the lower end of the above-means group displayed increased activity after the intervention. At the same time, the overall decline in mean indicates that the top contributors contributed less.

The Wilcoxon signed ranks test on the above mean subjects in *group 1* shows that there is no significant effect of the manipulation, T = 36.50, Z = -1.006, p < 0.315, r = 0.260 (For the sake of consistency, it is worth noting that the paired samples t-test does not detect significance either: t(15) = -1.103, p < 0.289, Cohen's d = -0.41). Although the effect is not significant, the effect size is close to medium. This means that a change of some magnitude happened to the group, although it is debatable what caused it. In other words, the number of contributions decreased a lot, but the change is not significant.

Below mean subjects: For the *below-means subjects* in *group 1*, the mean contributions increased from 0.00 posts per user in the two weeks before the experiment

to 0.21 posts after. The median is constant at 0 posts per user. For the below means subjects in *group 1*, the paired samples t-test shows that there is no significant change at the 0.05 α-level, although it is close. The test statistics show: $t(92) = 1.898$, $p < 0.061$, Cohen's $d = -0.90$. This effect size is considered high, even though it is non-significant. As the p-value is close to significant and effect size, it is likely that the results would be significant with a larger sample size.

6.4.4.2. Group 2: Non-Comparative Information

Above mean subjects: For the subjects in this group, the mean contributions fell from 13.73 posts per user in the pre-intervention period to 10.36 posts after. In spite of the lower mean, the median increased from 6 posts before to 9 posts post-intervention. The Wilcoxon signed ranks test on the above-mean subjects in Group 2 shows that there is no significant effect of the manipulation, $T = 15$, $Z = -1.277$, $p < 0.202$, $r = 0.39$ (For the sake of consistence, it is worth noting that the paired samples t-test does not detect significance either, despite a very large effect size: $t(11) = -1.449$, $p < 0.178$, Cohen's $d = -1.41$).

Although the effect is not significant, the effect size is between medium and large. As the group only comprises 11 subjects, this 0.39 effect size indicates that a larger dataset would show a significant effect.

Below mean subjects: For the *below-mean subjects* in *group 2*, the mean contributions increased from 0.20 posts per user in the two weeks leading up to the intervention to 0.68 posts after. The median is constant at 0 posts per user. For the *below means subjects* in *group 2*, the Wilcoxon signed rank test shows that there is a significant effect change at the 0.05 α-level as the paired samples t-test shows: $t(99) = 2.306$, $p < 0.023$, Cohen's $d = 0.38$. In sum, the statistical test shows that the *below-mean subjects* in *group 2* contribute significantly more to K10 after being sent non-comparative participatory information.

The effect size is Cohen's $d = 0.38$, which indicates that it falls well above a small effect size, but below medium. Thus, the *below-mean subjects* who received in-

formation about their own number of posts, but nothing about the community average, contributed significantly more after the intervention.

6.4.4.3. Group 3: Control Group

Above mean subjects: For the subjects in the group, the mean contributions fell from 9.50 posts per user pre-intervention to 8.48 posts post-intervention. In spite of the lower mean, the median increased from 4 posts pre-intervention to 6 post-intervention posts. The paired samples t-test on the above mean subjects in the *control group* shows that there is no significant difference between the contribution level before and after the experiment, as $t(47) = -0.686$, $p < 0.496$ and Cohen's $d = -0.14$. The effect size of -0.14 is considered small. The high p-value on 0.496 indicates that it is a valid experiment and that the fluctuations in the K10 participation caused by the independent variable.

Below mean subjects: For the subjects in the group, the mean contributions increased from .06 posts per user in the two weeks before the experiment to 0.10 posts after. The median is constant at 0. For the *below means subjects* in the *control group*, the paired samples t-test shows that there is no significant difference between the contribution level before and after the experiment: $t(1,216) = 1.752$, $p < 0.080$, Cohen's $d = 0.009$. This group has a surprisingly low p-value for a control group. This could be a cause of concern for the experimental validity. The group is a very large group, $n = 1,217$, making it much more likely to detect significance. Moreover, the effect size is very small, Cohen's $d = 0.009$, which supports the claim that the low p-value at least partly is a result of the large n.
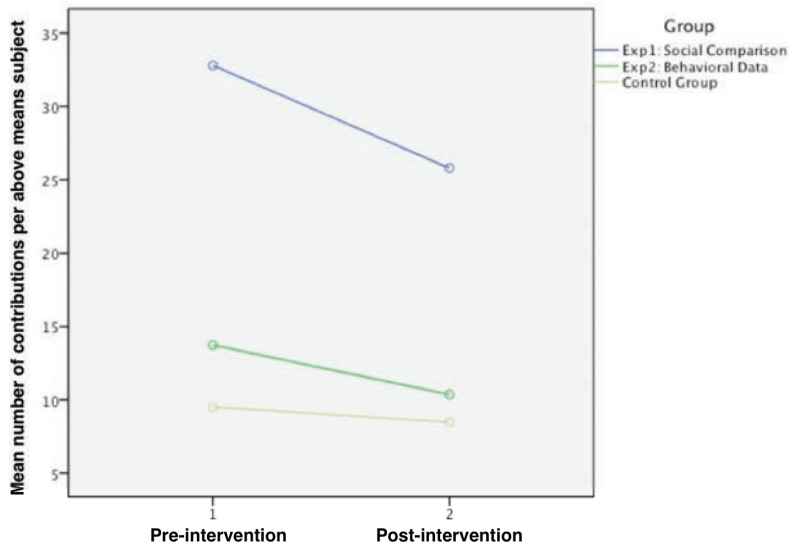
Figure 18 shows the participation level for the above-means subjects in *group 1*, *group 2*, and *group 3*. The figure shows that *group 1* is declining more rapidly than *group 2*, while *group 3* is relatively constant. However, the changes are not significant.
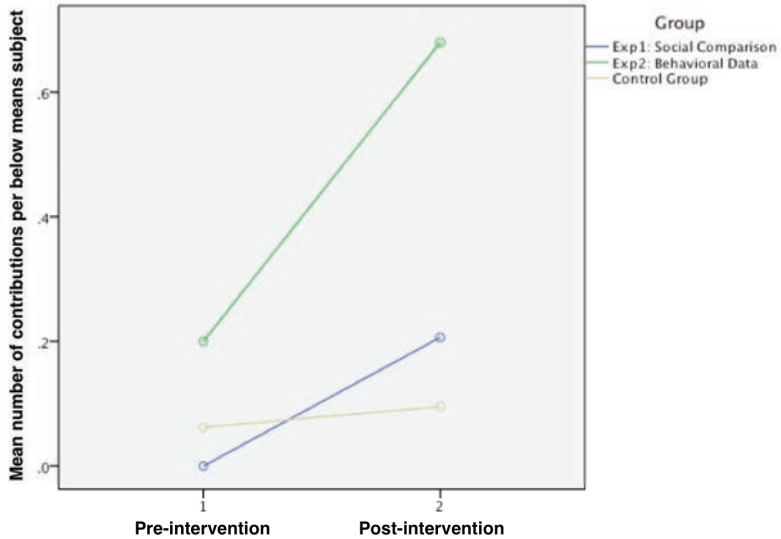
Figure 19 shows that the largest – and only significant effect – is among the *below-means group 2* subjects. Although the participation among the subjects in *group 1* also increases quite a bit, it is not significant.

6.4.5. Second Iteration: Removal of Zeros

As noted above, the data is heavily zero-inflated. This makes it harder to detect significance. To reduce the likelihood of Type II errors – false negatives – this part of the data analysis is similar to the first, but it only contains subjects who made at least one contribution in the four-week experimental timeframe.

		Posts	Median	N	Mean	Std. Dev	Min	Max
Group 1	**Before**	**492**	**4**	**20**	**24.60**	**33.994**	**0**	**94**
	After	406	7	20	20.30	28.146	0	111
Group 2	**Before**	**171**	**1**	**33**	**5.18**	**9.831**	**0**	**50**
	After	182	3	33	5.52	7.357	0	30
Group 3	**Before**	**532**	**2**	**112**	**4.75**	**9.258**	**0**	**56**
	After	523	2	112	4.67	8.696	0	52
Total	**Before**	**1106**	**2**	**165**	**7.24**	**15.893**	**0**	**111**
	After	1194	2	165	6.73	13.507	0	94

Table 29: Posts per user two weeks before and two weeks after the experiment

One interesting finding presented in Table 29 is that the standard deviation for *group 1* and *group 2* is smaller post-intervention than pre-intervention. This indicates that the intervention has homogenizing effects on the subjects' participation. This is what at least one strand of the theory predicts as the outcome of social comparison information. The standard deviation in *group 3* also declined, but less than for the two other groups.

In the following analysis, the paired samples t-test is used for groups with n > 30, and the non-parametric equivalent, the Wilcoxon Signed Ranks test, is used for groups with n < 30. The overarching data shows that *group 1* only has n = 20 subjects, and therefore does not fulfill the central limit theorem.

For all subjects who contributed at least one comment within the experimental timeframe, the mean number of posts per subject decreased from 7.24 before the intervention to 6.73 posts after the intervention. The median was 2 posts per subject both before and after the experimental intervention. The paired samples t-test shows that $t(164) = -0.662$, $p < 0.509$, Cohen's $d = .08$. This means that there is no significant change among all subjects, and the effect size is small.

			Posts	Median	N	Mean	Std. Dev.	Min	Max
Group 1	Above mean	Before	492	19	14	35.14	35.916	0	94
		After	**387**	**21.50**	**14**	**27.64**	**31.000**	**0**	**111**
	Below mean	Before	0	0	6	00.00	0.000	0	0
		After	**19**	**2**	**6**	**3.17**	**2.92**	**0**	**9**
Group 2	Above mean	Before	151	9.50	10	15.10	13.577	0	50
		After	**114**	**8.50**	**10**	**11.40**	**10.394**	**0**	**30**
	Below mean	Before	20	1	23	0.87	1.014	0	3
		After	**68**	**2**	**23**	**2.96**	**3.457**	**0**	**14**
Group 3	Above mean	Before	456	6	45	10.13	12.820	0	56
		After	**407**	**4**	**45**	**9.04**	**12.765**	**0**	**52**
	Below mean	Before	73	1	67	1.13	1.242	0	5
		After	**122**	**1**	**67**	**1.73**	**2.206**	**0**	**11**

Table 30: Posts per user two weeks before and two weeks after the intervention

On the overarching group level the tests show no significance. For *group 1*, the mean number of posts declined from 24.60 posts over a two-week period before the social comparison information is sent out to 20.30 after the emails. Despite the decrease in mean, the median increases from 4 posts before the intervention to 7 posts after. The Wilcoxon Signed Ranks test shows T = 103.50, z = -0.056, p < 0.955, r = -0.13, indicating a small effect size. This test shows that there is no significant change in *group 1 subjects'* participatory behavior after the intervention.

For *group 2*, the mean number of contributions increases slightly from 5.18 posts before the intervention to 5.52 posts after the intervention. The median increased from 1 post before the email intervention to 3 posts after. The paired samples t-test indicates non-significant results of the intervention, as the test result is t(32) =0.317, p < 0.753. This test shows that there is no significant change in the behavior of the subjects in *group 2* after the experimental intervention.

Similarly, for *group 3* the paired samples t-test also indicates non-significant results of the intervention. The mean is nearly constant: 4.75 posts before the intervention compared to 4.67 posts after the intervention. The median remains constant at 2 posts across the experiment. According to the paired samples t-test, this change is not significant: t(112) =-1.20, p < 0.905, Cohen's d = -0.02. In other

words, the *control group* seems reliable as the p value is very high and effect size very low.

		Mean	Std. Dev.	S.E. Mean	t	df	Sig. (2-tailed)	Effect size d
Group 1	Above							
	Below							
Group 2	Above							
	Below							
Group 3	Above	0.597	2.747	0.336	1.779	66	0.08	-0.14
	Below	-1.089	10.66	1.589	-0.685	44	0.497	0.31

Table 31: Paired samples t-tests for the groups meeting parametric assumptions: Posts per user two weeks before and two weeks after the experiment

The three statistical tests above show that the experimental interventions have little or no effect on the subjects in all three groups. Table 31 and Table 32 tell a different story when separating the groups into *above-means* and *below-means* subjects.

Group		Z	Assymp. Sig. (2 tailed)	Effect size R $(r = Z/\sqrt{n})$
Group 1	Above mean	-1.006	0.315	0.269
	Below mean	-2.226	.0.026	0.909
Group 2	Above mean	-1.277	0.202	0.404
	Below mean	-2.272	0.023	0.474
Group 3				

Table 32: Wilcoxon signed-rank tests for the groups *not* meeting parametric assumptions: Posts per user two weeks before and two weeks after the experiment.

6.4.5.1. Group 1: Social Comparison Information

In *group 1*, the *above-mean contributors* contribute less post-intervention than pre-intervention. As Figure 20 shows, the decline is considerable. The mean number of posts is 35.14 posts before the intervention and 27.64 posts after the experiment. Even though the mean declines, the median increases from 19 to 21.50 posts in the period. In the post-intervention period, the mean and median are much closer to each other compared to the pre-intervention period. This in-dicates a homogenizing effect of the intervention. Despite these rather large

shifts in value, the Wilcoxon signed ranks test shows that the drop is non-significant: T = 36.50, z = -1.006, p < 0.315, r = 0.269. Although the intervention did not yield a significant effect, the effect size of 0.269 suggests that it had a medium sized impact. Again, the statistical test shows that there is no significant change in *group 1 above-means subjects'* behavior after the experimental intervention.
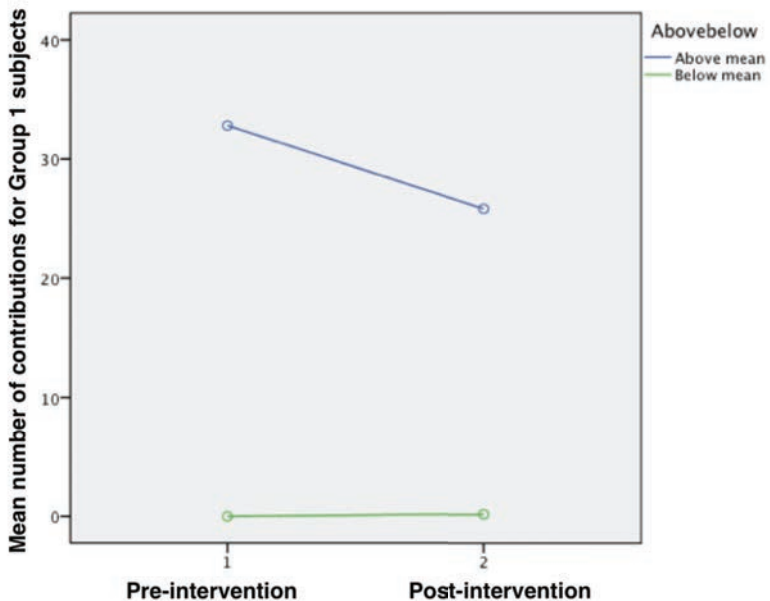
Figure 20: Contributions by above-mean and below-mean subjects in *group 1*

Figure 20 only shows a small increase among the *below-means contributors*. The figure does not do the data justice due to the scales. The mean number of contributions increases from 0 contributions before the experimental intervention to 3.17 after. The median increases from 0 posts before the intervention to 2 posts after the intervention. The effect is significant as T = 0, z = -2.226, p < 0.026, r = 0.909. This is a very large effect size, strongly suggesting that the intervention has a high impact on the subjects. Figure 21 below compares the pre-intervention and post-intervention scores for all below-means groups, and

shows the effect more clearly. In sum, the Wilcoxon signed ranks test shows that the *group 1 below-means subjects* contribute significantly more after receiving social comparison information in an email.

6.4.5.2. Group 2: Non-Comparative Information

For *group 2*, the *above-mean contributors* contribute less after receiving the intervention email: The mean is 15.10 posts pre-intervention and 11.40 posts in the post-intervention period. The mean decreases from 9.5 posts before the intervention to 8.5 posts after the intervention. The Wilcoxon signed ranks test shows T = 15, z = -1.277, p < 0.202, r = 0.404. The test shows that the observed change is non-significant.
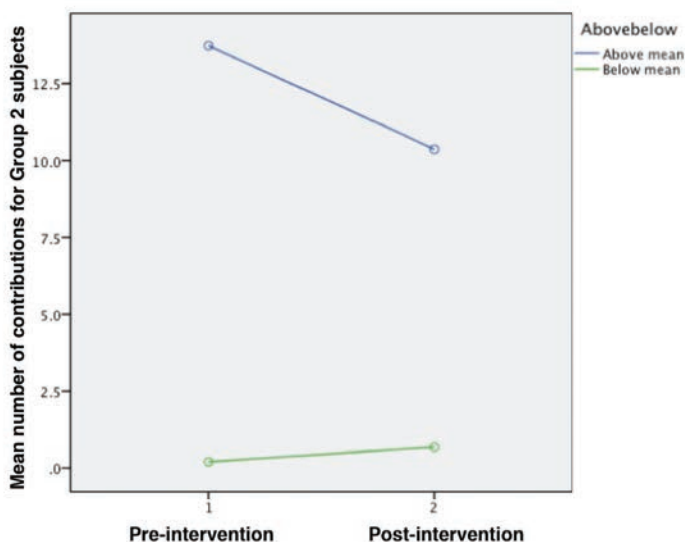
Figure 21: Contributions by above-mean and below-mean subjects in *group 2*

For the *below-means contributors*, the intervention caused a change from a mean of 0.87 posts in the pre-intervention period to 2.96 post-intervention posts. The median increased from 1 to 2 posts after the intervention. The Wilcoxon signed ranks test shows that the intervention effect is significant, as T = 50.50, z = -

2.272, p < 0.023, r = 0.474. The effect size r shows that the effect size of the intervention is close to large. Thus, the Wilcoxon signed ranks test shows that the *group 2 below-means subjects* contribute significantly more after receiving an email with non-comparative participatory information.

6.4.5.3. Group 3: Control Group

Although there is some variation in *group 3*, the variation is non-significant for both above-means and below-means contributors. For the *above-means contributors*, the mean drops from 10.13 posts pre-intervention to 9.04 post-intervention. The median also declines from 6 posts before the intervention to 4 posts after. The paired samples t-test shows that $t(44) = -0.685$, $p < 0.497$. The effect size is small as Cohen's d = -0.14
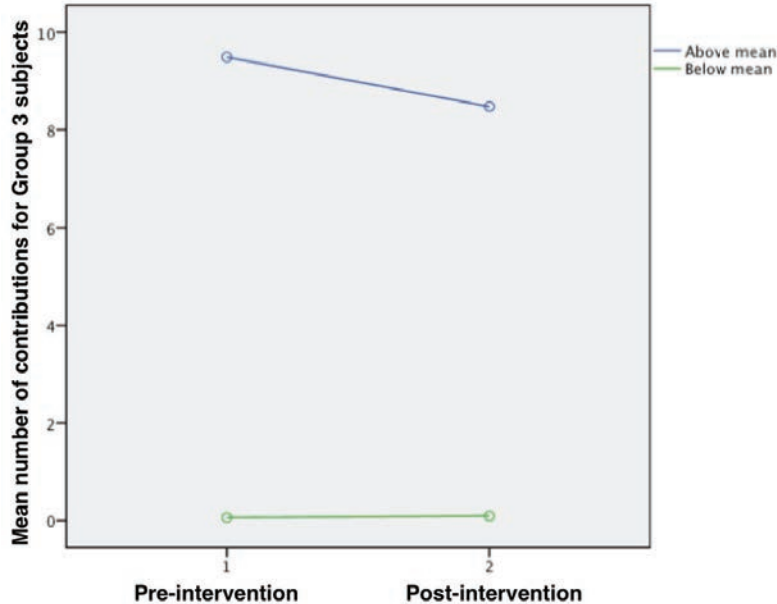
Figure 22: Contributions by above-mean and below-mean subjects in *group 3*

For the *below-mean contributors*, the number of posts increased from 1.13 before the intervention to 1.73 after the intervention. The median remained constant at 1 post per user across the two time periods. Although this change is quite large, it is not significant: $t(66) = 1.779$, $p < 0.08$, Cohen's $d = 0.38$. Surprisingly, this indicates a medium effect size, and therefor that some spurious variables might be affecting the experiment. The relatively low p-value in this group vis-à-vis the p-values in other groups might be caused by the relatively large group size. The group size does increase the likelihood of identifying significant effects. Despite the medium effect size in this group, both the above-mean and below-mean contributors from the control group have non-significant changes. Although there is some noise from spurious variables, it is not sufficient to compromise the internal validity.

Figure 23 and Figure 24 below show the mean contributions pre- and post-intervention for respectively above-means contributors and below-means contributors.
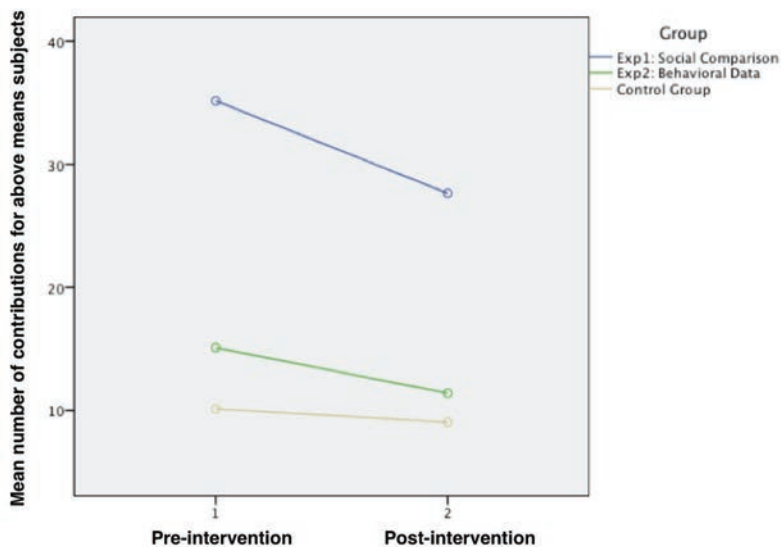
Figure 23: Contributions by above-mean subjects in all three groups

For the above-means contributors, the tendency is identical to the first data analysis iteration, comprising the full dataset. *Group 1* is dropping more than *group 2*, while *group 3* is fairly constant. None of the changes are significant.
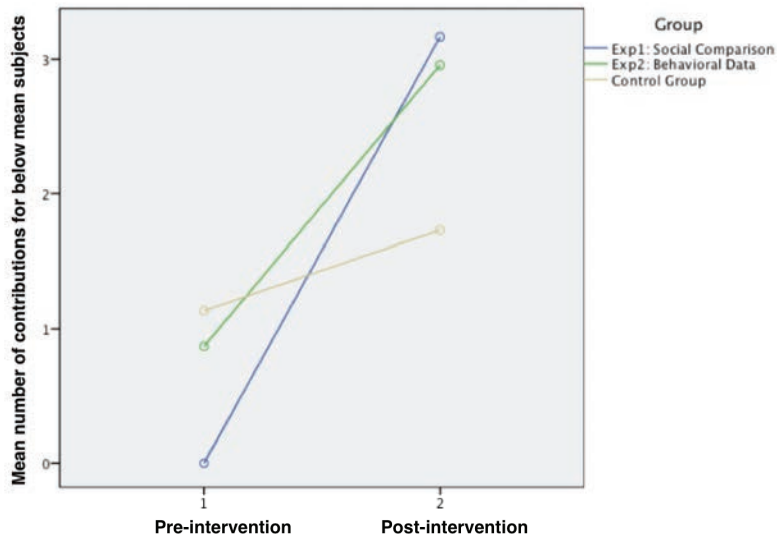
The tendency among the *below-means subjects* has changed. The change in *group 1* subjects' behavior is now larger than *group 2* subjects', which is what would be expected according to social comparison theory. *Group 3* is still fairly constant.

6.4.6. Summary of Results

As shown in Table 33, the first part of the data analysis shows only a significant effect in the *below-means contributors* of *group 1*. The second part of the data analysis, only comprising active participants in the experimental period, shows a significant effect in two groups: the *below mean subjects* of *group 1* and *group 2*.

	Hypothesis	Verdict (all subjects)	Verdict (all participatory subjects)
H_0	Receiving social comparison information has *no* effect on subsequent participation.	Rejected	Rejected
H_a	Receiving social comparison information has an effect on subsequent participation	Rejected	Rejected
H_{a1}	Receiving social comparison information reduces the contribution level among above-means contributors	Rejected	Rejected
H_{a2}	Receiving social comparison information increases the contribution level among below-means contributors	Rejected	Verified
H_b	Receiving non-comparative information has an effect on subsequent participation, but smaller than comparative information.	Rejected	Rejected
H_{b1}	Receiving non-comparative information increases the contribution level among above-means contributors	Rejected	Rejected
H_{b2}	Receiving non-comparative information increases the contribution level among below-means contributors	Verified	Verified

Table 33: Evaluation of the seven research hypotheses in experiment 1

The results suggest that social comparison information can lead to improved participation among the users who usually do not participate a lot. An interesting aspect is that the comparative information is not significantly different from non-comparative information. In fact, the statistical significance and effect sizes are almost identical for the below mean users in *group 1* and *group 2*. This might indicate that non-comparative participation information can be as effective as comparative information. One possibly explanation for this lack of difference between comparative and non-comparative information is that the subjects have a fairly good sense of how their contribution level over the past month compares to the average. Although the *group* 2 subjects can only see how much they participated in the time frame, it is reasonable to assume that they know if this number is above or below the group mean. Especially for those subjects who are either highly active or inactive.

Although the experiment is non-significant on the above-means subjects, they did contribute less after the experiment than they did before. This results in a

decline of posts in the aggregate level from 1,195 to 1,111 posts. The overall drop is a result of a skewed participation level where a few users are doing a lion share of the contributions.

6.4.7. Qualitative Insights

As previously stated, Chen et al. (2010) recruited community members to their experiment. This approach does not show how regular users respond to social comparison information when they have not explicitly consented to receive such an email. In order to increase experimental realism and thus expose potential unintended consequences of using social comparison design patterns, no explicit recruitment was made in this study. All active community members are a part of this experiment.

Emails were sent out to 2,684 users, who at some point had written one or more post at K10. The emails caused more reactions than expected. Subjects responded in the following ways: by replying to the email, by contact form on the website, by sending an email to the admin, and through private messages in the community. The reactions were mostly harsh.

80 people (2.98 %) wrote a negative respond, whereof 65 people (2.42 %) asked to get their profiles deleted. This is a high unsubscribe rate compared to the average unsubscribe rate for social networks and online communities, which averages at 0.74 % (Mailchimp 2010). It is worth noting that K10 very rarely sends out emails, making it more likely that some passive users will unsubscribe to this email. This indicates that social comparison information might make users more upset than other types of email content such as newsletters and notifications. Moreover, 22 subjects (0.82 %) wrote a response because they were confused about the email. This response is likely due to the fact that some of the recipients have not been active for several years. Finally, 4 people (0.15 %) wrote a positive response to the email. To get a sense of the tone in the email, a few of the replies are presented here.

Some users became quite upset. One user wrote: "It would suit you not to write to me as I already know that I did not write any posts recently. I do not need you to tell me that. And if it can't be in any other ways, you can delete me. Ridiculous email." One user felt unwelcome after receiving the email: "You can just delete me if I am not welcome". Another tried to explain the lack of participation: "I did not contribute more than I did, because I participate in many different communities". Curiously, some users thought they got the emails because they wrote too many posts: "I thought you wrote because you were tired of me. I read the email as a warning:"

A few were more positive. For example, one user replied "Thank you for the information. I would like to contribute to stop doctor consultants or make them responsible for their actions."

The responses were similar to those of subjects in *group 1* and *group 2*. It does not seem to be the comparative information *per se* that created the negative responses. Simply informing people about their level of contribution is enough to upset some users. This might also explain why the comparative and non-comparative emails cause the same behavioral effects in the subjects. This can also help explain why the above-means users contributed less after the intervention, although the decline is non-significant.

6.5. Limitations

This experiment has a number of limitations in terms of internal validity, external validity, and construct validity.

6.5.1. Internal Validity

Unaware of its existence, the subjects cannot know they are part of an experiment and behave accordingly. This increases the realism and internal validity of

the experiment, as experimenter biases are eliminated. [218] The increased participation among the below mean users in *group 1* and *group 2* might not have anything to do with the information in the email, but rather the fact that they got an email reminding them about the existence of K10. In other words, the content of the emails might be irrelevant. Merely sending out emails to users will make them participate more. Although it is a possibility, it is not very likely. The results found in this study generally support social comparison theory and verifies the findings in Chen et al. (2010).

The lack of difference between subjects who receive comparative (*group 1*) and non-comparative (*group 2*) information means it is relevant to question the applied theory operationalization. Again, since the purpose is to improve practice rather than verify theory, only a simple construct of the theory is applied here. But it is surprising that no differences between the groups are detected. This makes it relevant to do follow up studies with more rigorous theory operationalization to find out if this unexpected result is caused by the theory or the theory operationalization. As noted above, one likely explanation for the similar behavior among the subjects in *group 1* and *group 2* is that the subjects in *group 2 already* have a sense of how much they contribute compared to group the mean., even though they don't have a specific number of the group mean. Especially users with 0 or 1 post in the period might have a strong feeling that they are below

---

[218] The experimental intervention in this experiment is conducted via email. In many ways, this research design resembles Chen et al. (2010). Both studies seek to improve participation through an online field experiments conducted via email, relying on social comparison theory. Chen et al. (2010) invite users to take part in the experiment, rather than making the intervention directly in the field. Of the users who clicked in the email, only 63 % consented to participate in the study. Based on a Kolmogorov-Smirnov test of the equality of distributions, the authors find a significant difference in participation between the users who opted in and those who did not. There is thus a risk of sampling bias and volunteer effect by excluding critical users (Rosnow et al. 1969; Chen et al. 2010). Furthermore, the users were motivated to participate through prize drawings. In this dissertation users were not asked for their consent to participate, nor were they awarded any prizes. In other words, all current users of K10 were placed in one of the three groups. This increases the realism and external validity of the experiment.

the mean in terms of participation. If this is true, social comparison theory can account for the significant effects in both *group 1* and *group 2*.

6.5.2. Construct Validity

In the first experiment, social comparison theory is translated into a quantitative variable: number of posts the last 30 days. This choice is made for several reasons. It is a fairly objective unit of measurement that is easy to compute, and easy for the experimental subjects to understand. However, social comparison theory can be translated in many ways, and the translation used in this study might not match the users' most salient comparison metrics. They could, for example, be concerned with other quantitative metrics such as how many people read their replies, and/or how many users respond to their posts. It is also possible that they are more preoccupied with qualitative metrics such as the quality of the posts.

6.5.3. External Validity

Omitting explicit user recruitment not only increases the realism of the experiment, but also the external validity. Based on the research findings in this experiment, other online communities can try implementing the design pattern for some user groups with the purpose of increasing participation. Assuming the users have already opted in to receive some emails from the website, it is unlikely that the users will be asked to opt in to social comparison emails. Thus, *not* asking users to participate makes the experimental setting more similar to real online community settings. The external validity of the experiments is further discussed in Section 10.3.1.

6.6. Implications for Open Government Community Design

6.6.1.1. Altered Participation Patterns

The aim of the experiment is to examine degree to which social comparison information can affect participation in K10. The overall participation quantity did

change, but not for the better as it declined. The participation regressed toward the mean. A lesson learned here is that it is important to analyze the overall participation patterns in a community before enacting social comparison. Figure 16 above shows a heavily skewed participation pattern, where a small number of users contribute most of the posts. More specifically, if one considers using social comparison design elements in an open government community, it is critical to examine the existing contribution pattern, as it is likely to be altered.

In the two weeks prior to the experiment, the top 0.5 % of the users wrote 50 % of the posts. In the two weeks following the experiment, the top 0.5 % of the users contributed 42 % of the content, entailing reduced participation by the top users and increased participation by other users. This participation pattern is even more skewed than most of the numbers cited in chapter 2. The skewed data entails that the regression toward the mean yielded an overall negative result.

In an open government perspective, it is worth considering if a total decline in participation is acceptable, as long as it creates a more balanced participation pattern. More voices are heard, and the loud voices are less vocal. From a participative democracy perspective (op cit.), the aim of open government is to enable and encourage as many citizens as possible to participate with the goal of furthering democracy. The regression toward the mean entails that the users have a more similar voice.

Furthermore, the results of this study indicate that more targeted exposure to participatory information can result in better outcomes. *First*, one can make segmented comparisons so people are compared to peers. If top contributors are compared to other top contributors, the regression will be toward a top contributor mean, and not the mean in the entire community. A *second* strategic use of participatory information is to send it exclusively to the below means contributors, and thus not send anything to above-means contributors. This would provide some of the positive effects while reducing the negative.

When sending out emails to users, it is almost inevitable that there will be some collateral effects. Many people might not like receiving emails from online communities and perceive them as spam. Other people might be confused by seemingly random emails. By asking users to opt in on the experiment, these collateral effects are removed. Ignoring such effect might not have any consequences for the research experiment *per se*, but it has consequences for other online communities implementing similar measures. Not reporting these effects may lead other people to underestimate the negative consequences of delivering social comparison information via email.

6.6.1.2. Upset Users

The responses to the experimental emails show that some users got upset when receiving these emails. The negative reaction among some users is important to note when considering sending emails containing participatory information. But why did some users get upset? At least three things might account for it. *First*, some people do not like receiving emails. Their dissatisfaction might be caused by the fact that the intervention was conducted via email instead of on the website. One user wrote: "I was upset about the emails. I thought I had unsubscribed from all social networks, and then came this email from K10". According to this explanation it is not the social comparison information but rather email, the experimental proxy used here, that is the culprit.

Second, as described in Chapter 3, one important trait of K10 users is their poor health conditions. The users who are on early retirement pension are considered too sick to maintain a job while the users involved with the flex job program can maintain limited work functions. Thus, K10 users are already under tremendous stress, and the frustration might be more characteristic of a such population than a young, tech savvy one. For example, one user wrote that "We feel less worthy to begin with, and cause a state of submissiveness". This explanation is supported by Taylor, Buunk, and Aspinwall (1990), who find that the cognitions underlying social comparison processes are different for people in stressful situations. In

other words, the fact that users became upset might be case-specific and limited to K10 and similar open government communities.

Third, sending information about participation might make some users feel they are being pressured to participate more. Surprisingly, this can happen even with emails not containing comparative information.

Social comparison theory can help explain why some users get upset when receiving social comparison information. As described in Section 4.4.2, social comparison is only possible when the users have knowledge about other users' participation. Not being able to 'keep up with the joneses' can be uncomfortable and cast doubt about one's own performance (Strahan et al. 2006). This can induce negative feelings in the users and thus make them upset. According to Swallow and Kuiper (1988) people who are depressed and/or likely to experience stress will often react negatively to comparative self-evaluations. But social comparison theory cannot explain why subjects receiving non-comparative information find the email equally upsetting. One explanation might be that the subjects that have received an email with below-mean information know they have contributed less than the average, even though it is not written explicitly in the email.

It is also entirely possible that the negativity has little to do with the content of the email, but rather the two reasons given above; email-as-a-medium and case specific traits.

Altered participation patterns and unintended consequences are further discussed in section 10.2, which is a discussion of the broader implications for open government.

6.7. Chapter Summary
Based on social comparison theory, it is hypothesized that the K10 users will contribute more to the community when compared to other users level of participation. This chapter reports an email experiment using social comparison theo-

ry to change the participation on K10. Providing both social comparison and non-comparative information affected the participation among below-means subjects significantly. Although below-means subjects contribute more post-intervention, the overall effect of the intervention is negative. The above-mean users contribute less, although the effect is non-significant. A surprising finding in this experiment is that comparative and non-comparative information seem to have almost similar effects on the subjects.

The significant changes in behavior among some groups show that social comparison theory – and social psychology in general – hold a potential to alter the participation patterns to open government communities. This strongly indicates that this experiment, as well as those reported in the following three chapters, is able to contribute to improving the current state of open government communities.

From a methodological perspective, emails are successfully used to cause the experimental intervention through a between-subject design. However, the emails containing participatory information also had an unexpected side effect. A surprisingly large amount of users were upset about receiving the email. This suggests that the combination of email and social comparison information has some costs in terms of declined user satisfaction. These negative reactions might, however, be due to some of the special characteristics of the K10 community users.

Chapter 7: How Goal-Setting Affects Participation in K10 - A Survey Experiment

7.1. Introduction

Goal-setting theory finds that goals often increase the motivation and performance for those who are subject to the goal. In open government communities this might entail increased participation and engagement. In this experiment, goal-setting theory is operationalized to an open government community context. The purpose of this operationalization is to examine if goals can help increase participation, and thus help solve the participation challenge.

In this experiment, K10 users are encouraged to answer a survey in which they are asked how many posts they will write in the next 14 days. The main aim of the experiment is to examine if the subjects live up to their goal commitment. If goal-setting is to have an effect on participatory behavior, it is a necessary condition that the subjects assign themselves an ambitious goal. This experiment is therefore designed to get the subjects to assign themselves as high a goal as possible. But this is only a means in the experiment in order to examine how goal-setting affects participation in the K10 community. For a full description of operationalization of goal-setting theory, please refer to Section 4.4.3.

If the subjects assign themselves high goals and subsequently live up to them, it might help open government communities overcome the participation challenge by encouraging users to assign themselves participation goals.

This chapter offers two contributions. The first contribution is a substantial one, offering knowledge about how an open government community – K10 – reacts to a certain construct of goal-setting theory. The second contribution is a methodological one, resulting from the application of online survey as a tool to cause the experimental intervention.

7.1.1. Research Question and Hypotheses

Based on goal-setting theory, this experiment seeks to address the following research question: *how does a self-assigned goal to a number of contributions to K10 affect subsequent participation?*

To examine this research question, one null hypothesis and one alternative hypothesis are tested:

H_0 Setting a participation goal has no effect on subsequent participation

H_a Setting a participation goal makes subjects strive to achieve the goal

The *independent variable* in this experiment is the goal commitment. This independent variable has three levels:

- *Group 1*: Subjects are asked to assign themselves a participation goal in a survey
- *Group 2*: Subjects are asked to answer a survey, but are not asked about participation goals.
- *Group 3*: Subjects are not asked to answer a survey.

The *dependent variable* is number of contributions to K10 following the survey submission. The dependent variable is measured two weeks before and two weeks after the intervention.

7.2. Research Design

As described in chapter 5, four different mediums are considered to cause the experimental intervention in this dissertation: website, email, survey, and advertisements. Setting a goal requires a medium that can facilitate interactivity. Advertisements do not allow this, as the users can click an advertisement, but not type in any input. Emails could be used in this experiment, but it does not scale very well. If emails containing questions are sent out to users, each email has to be coded manually. This leaves two viable options for the intervention: survey and website.

The two options are quite similar, but with one important difference: setting up a survey is much cheaper than coding the feature and running the experiment on the website. As the resources at K10 are scarce, the low-cost alternative is preferred.

7.2.1. Split-Plot Design

The research design for this experiment is a split-plot design, comprising both a between-group comparison and a within-group comparison. In the between group comparison, the subjects from the three conditions are compared. In the within-group comparison, subjects in the experimental group are analyzed to examine whether the different goals cause different behavior among the subjects.

	Pre-intervention contribution level	Post-intervention contribution level
Group 1	T_0 observation	T_1 observation
Group 2	T_0 observation	T_1 observation
Group 3	T_0 observation	T_1 observation

Table 34: Split-plot research design used in experiment 2. Adapted from Bordens and Abbott (2010)

The split-plot design, illustrated in Table 34, is very compatible with goal-setting theory. It is possible to compare pre- and post-intervention contribution levels for the subjects assigned to different groups. This helps avoid the risk of carryover effects inherent in a regular within-subject design.

7.2.2. Sampling and Subject Assignment

A link to the survey was placed on the homepage of K10 on February 10th 2012, at 8 am CET.[219] The survey was open for one week.

---

[219] After the negative responses to the emails in the first experiment (c.f. section 6.4.7) the K10 owner prefers to conduct the second experiment without sending emails.

Screenshot 9: K10 homepage showing survey link in the right sidebar.

This experiment is subject to a self-selection problem since not all K10 users are taking part in the survey. No passive users will be dragged in, which might create a bias toward the most active users. This means there might be systemic differences between users who answer the survey, and those who do not. To control for this potential response-bias, the non-survey answers are used as a control group. But this group is not directly comparable to the two survey groups due to potential unknown biases. As it cannot be used in a direct comparison, it is used as a baseline to help identify potential biases between the survey-takers and the non-survey-takers.

As K10 users do not see the survey link at the same time, it is unlikely that they will answer the survey simultaneously. To account for this time disparity the two-week pre-intervention and post-intervention timeframe is calculated for each subject in the data analysis.

In this experiment, the assigning of subjects is conducted through Survey Monkey. The assignment is done through the following steps: (1) the survey is created and the two conditions defined: an experimental group and a control group. As the aim is to create two groups of similar size, (2) the desired subject assignment is set to 50 %. This does not entail that each group will have a similar

amount of subjects, but that each subject has 50 % chance of being assigned to the experimental group and 50 % chance of being assigned to the control group. Based on this assignment, (3) each survey respondent is randomly assigned to one of the conditions once clicking the link from K10 to Survey Monkey.

7.3. Conducting the Experiment

7.3.1. Intervention: Survey Design

Surveys are often used as a tool to conduct experiments. In this research design, different treatments are constructed. This experiment contains two survey conditions (*group 1* and *group 2*) and a non-survey condition (*group 3*). In the *experimental treatment*, the subjects are asked how many posts they will write in the following two weeks. The subjects in the *survey control treatment* are not asked this question. The subjects in the *no-survey control treatment* are not asked any questions at all.

Survey Monkey[220] is used to conduct this experiment. The software has an experimental design feature, making it feasible to set up a robust experiment in which the subjects are randomly assigned. One limitation of the Survey Monkey experiment feature is that it is designed for question wording and priming. It does not offer the opportunity to include an extra question in one of the conditions. Therefore, the surveys in both conditions have an equal number of questions in the two groups with one question being different. [221]

To avoid respondent fatigue and high dropout rates, the survey is limited to seven questions. [222] The questions are:

1. In which year were you born?

---

[220] http://surveymonkey.com

[221] Another solution would have been to set up two different surveys and assign users randomly to them. Adding an extra question to the control group is preferred to running two different surveys as it minimizes common online experimental pitfalls such as differences between the two variants, e.g. file name structures (Reips 2002), and non-random subject assignment.

[222] The full survey and original wording in Danish is enclosed in Appendix 6.

2. What gender are you?
3. What part of K10 do you like the best?
4. What would you like to be different at K10?
5. **Intervention**
   a. *[Experimental Group]:* How many posts do you expect to write at K10 during the next two weeks?
   b. *[Control Group]:* How many times a month do you visit K10?
6. Is there a K10 member who does something particularly good for other users?
7. We have a prize draw for three gift cards. It is completely optional if you want to be in the draw. In order to win the gift card, enter your user name.

According to (Tversky & Kahneman 1974), people are influenced by anchors when making decisions. Constructing categories with high values is a way to create high reference points and anchors. Reference points with high values make it more likely that subjects will make higher commitments than they otherwise would. Therefore, the response categories in the intervention question have higher values than the mean and median contributions at K10. The categories to the question *how many posts do you expect to write at K10 the next two weeks?* are:

- 0 posts
- 1-5 posts
- 6-10 posts
- 11-20 posts
- 20+ posts

To motivate users to fill out the survey, three gift cards of DKK 250 each are drawn among the participants. This practice of prize drawing is common in surveys (Porter & Whitcomb 2004), although one can speculate that some respondents will give more positive and desirable answers than they otherwise would (Balakrishnan et al. 1992).

7.3.2. Responses

In the seven days the link to the survey was on the homepage, 2,251 unique users visited the homepage (it is worth noting that K10 had a total of 11,574 unique users visited K10 in this period. This indicates that it was a mistake only to post the survey link on the homepage). Of these 2,251 people who visited the K10 homepage in the 7-day period, 135 users responded. This gives a response rate on merely 6 %. This number is likely to be higher as an unknown amount of users are likely to have overlooked the survey link. Of these 135 subjects, 84 subjects wrote their username in the survey and can thus be used in the experiment. This means an even lower effective response rate on 3.7%. Again, this number is likely to be somewhat higher, but it is unknown how many people noticed the survey link.

7.4. Data Analysis

The data analysis of this experiment begins with a description of the dataset and sample distribution. Afterward, the effect of the experimentation across the three groups is analyzed for both within-group variation and between-group variation. Subsequently, the subjects in the experimental condition are segmented into three groups based on their survey goal compared to their pre-intervention participation level.

7.4.1. Cleaning the Dataset

As noted above, 84 of 135 respondents typed in their username. This means 51 subjects were removed from the dataset as it is impossible to couple these survey answers to K10 users. No other subjects were removed from the dataset.

After the data collection, the data is exported from the K10 MySQL database into excel. Here, the numbers of pre-intervention and post-intervention contributions are calculated for each user. The sums are placed in two columns: pre-intervention post count and post-intervention post count.

7.4.2. Sample Distribution

As in the first experiment, the data is zero-inflated and therefore not normally distributed. The Kolmogorov-Smirnov test of normality shows $D(4,082) = 0.477$ $p < 0.000$ for the pre-intervention period and $D(4,082) = 0.476$ $p < 0.000$ for the post-intervention period. Thus, the normal distribution assumption is not met for the entire population. As the significance level of the Kolmogorov-Smirnov test of normality is < 0.05 for all three groups,[223] general linear models such as the repeated measures ANOVA are a poor fit.

The group size in all three groups is more than 30: The experimental group (*group 1*) has n = 51, the *survey control group* (*group 2*) has n = 33, and the *no-survey control group* (*group 3*) has n = 3,998.[224] This means that all groups fulfill the central limit theorem, making the data suitable for non-linear parametric tests. To test the between-group variation, an ANOVA test is applied. To examine the within-group variation in the three groups, the paired samples t-test is used.

7.4.3. Results

When looking at the contribution level across all three groups, the number of posts written before the intervention is 2,088 posts and 2,357 posts after the intervention. The mean number of contributions over the 14-day period prior to the intervention is 0.51 post while the number is 0.58 after the intervention. Due to the zero-inflated data, the median remains at 0.

---

[223] For *group 1* subjects, the Kolmogorov-Smirnov test of normality shows $D(51) = 0.309$ $p < 0.000$ for the pre-intervention period and $D(51) = 0.264$ $p < 0.000$ for the post-intervention period. For *group 2* subjects, the Kolmogorov-Smirnov test of normality shows $D(33) = 0.322$ $p < 0.000$ for the pre-intervention period and $D(51) = 0.326$ $p < 0.000$ for the post-intervention period. For *group 3* subjects, the Kolmogorov-Smirnov test of normality shows $D(3,988) = 0.485$ $p < 0.000$ for the pre-intervention period and $D(3,988) = 0.481$ $p < 0.000$ for the post-intervention period.

[224] *Group 3* comprises all users who have written one or more post in their K10 lifetime and *not* answered the survey.

7.4.3.1. Between Group Variance

A one-way ANOVA test shows that the differences between the groups are highly significant both in the two weeks before the experiment and two weeks after the experiment. For the pre-intervention period, the ANOVA tests shows $F_{(2632, 14,5)} = 181.539$, $p < 0.000$. This means that the difference between the groups before the experiment is highly significant. As the ANOVA test is an omnibus test, Tukey's HSD test is used to make pairwise post-hoc comparisons. With the value $p < 0.00$, this test shows that there are significant differences between *group 1* (m = 8.45 posts) and *group 3* (m = 0.35 posts) subjects. In addition, there are significant differences between *group 2* (m = 8.18 posts) and *group 3* subjects, as $p < 0.001$. However, there are no significant differences between *group 1* and *group 2* as $p < 0.946$.

This suggests that significant changes exist between those who did answer the survey and those who did not. However, the two different survey groups do not differ. This means that a self-selection bias probably is present.

		Sum of Squares	Df	Mean Square	F	Sig.
Number of posts before intervention	Between Groups	5264.61	2	2632.303	181.539	0.000
	Within Groups	59145.35	4079	14.5		
	Total	64409.96	4081			
Number of posts after intervention	Between Groups	9053.431	2	4526.715	233.083	0.000
	Within Groups	79218.61	4079	19.421		
	Total	88272.04	4081			

Table 35: Results of the omnibus ANOVA test for all three groups

The ANOVA test for the post-intervention period shows similar results. For the post-intervention period, $F_{(4526, 19)} = 233,08$, $p < 0.000$, meaning that the difference between the groups after the experiment is highly significant. Like the case is for the participation in the two weeks before the intervention, the difference between *group 1* (m = 10.55 posts) and *group 3* (m = 0.36 posts) subjects

are significant. Equally so are there significant differences between *group 2* (m = 11.3 posts) and *group 3* subjects, as p < 0.001. However, there are no significant differences between *group 1* and *group 2* as p < 0.946.

Dependent Variable	(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.
Number of posts **before** intervention	Experimental	Control	0.269	0.851	0.946
		No-Survey Control	8.104*	0.537	0.000
	Control	Experimental	-0.269	0.851	0.946
		No-Survey Control	7.835*	0.666	0.000
	No-Survey Control	Experimental	-8.104*	0.537	0.000
		Control	-7.835*	0.666	0.000
Number of posts **after** intervention	Experimental	Control	-0.75401	0.98454	0.724
		No-Survey Control	10.18734*	0.62102	0.000
	Control	Experimental	0.75401	0.98454	0.724
		No-Survey Control	10.94135*	0.77031	0.000
	No-Survey Control	Experimental	-10.18734*	0.62102	0.000
		Control	-10.94135*	0.77031	0.000

Table 36: Tukey's HSD post hoc test. (* = The mean difference is significant at the 0.05 level)

Since significant differences existed before the intervention, potential significant differences *after* the intervention must to be treated with caution due to the self-selection bias.

7.4.3.2. Within-Group Variance

When looking at the contribution level across all three groups, the number of posts written before the intervention is 2,088 posts and 2,357 posts were written after the intervention. The paired samples t-test shows that the change across all three groups is non-significant as $t(4081)$ -1.681, $p < 0.093$, Cohen's *d* =0.04. Thus, when measuring all subjects, no significant differences exist before or after the intervention. The low effect size indicates that the overall variation in the data set is small. This is to be expected as the non-survey control group is very large and should contain little variation since no intervention has been made here.

		Posts	Median	Mean	N	Std. Dev.	Std. Error mean
All subjects	Before	2088	0	0.51	4082	3.973	0.062
	After	2357	0	0.58	4082	4.65081	0.07279
All survey re-spondents	Before	701	3	8.35	84	16.942	1.848
	After	911	5	10.8452	84	17.22879	1.87982
Group 1 Experimental	Before	431	3	8.45	51	16.958	2.375
	After	438	4	10.549	51	15.22933	2.13253
Group 2 Control	Before	270	3	8.18	33	17.178	2.99
	After	373	7	11.303	33	20.17809	3.51255
Group 3 No Survey	Before	1387	0	0.35	3998	2.973	0.047
	After	1446	0	0.3617	3998	3.69574	0.05845

Table 37: Participation behavior across the three experimental groups

When examining the 84 survey-respondents (comprising both *group 1* and *group 2* subjects), it is clear that something has changed around the time of intervention. The mean number of posts pre-intervention is 8.35 posts per user in the two-week period before the intervention, while the mean number is 10.85 posts per user post-intervention. The median increases from 3 posts per user to 5 posts per user in the same period. The paired samples t-test shows t(84) -2.830, p < 0.006, Cohen's *d* =0.44. Thus, the difference in means before and after the intervention is statistically significant and the effect size is small to medium.

	Mean	Std. De-viation	Std. Error Mean	t	df	Sig. (2-tailed)	Co-hen's D
All subjects	-0.07	2.50514	0.03921	-1.681	4081	0.093	0.04
All survey respondents	-2.50	8.09507	0.88324	-2.83	83	0.006	0.44
Group 1 Experimental	-2.10	8.87751	1.2431	-1.688	50	0.098	0.34
Group 2 Control	-3.12	6.79502	1.18286	-2.639	32	0.013	0.84
Group 3 No Survey	-0.02	2.21804	0.03508	-0.421	3997	0.674	0.00

Table 38: Paired samples t-test results for all three groups

As Figure 25 shows, the increase is not limited to the *experimental group*. Both *survey groups* (*group 1* and *group 2)* are contributing more after the intervention.

This suggests that it might not be the specific questions in the survey that have an effect, but answering a survey *per se*.
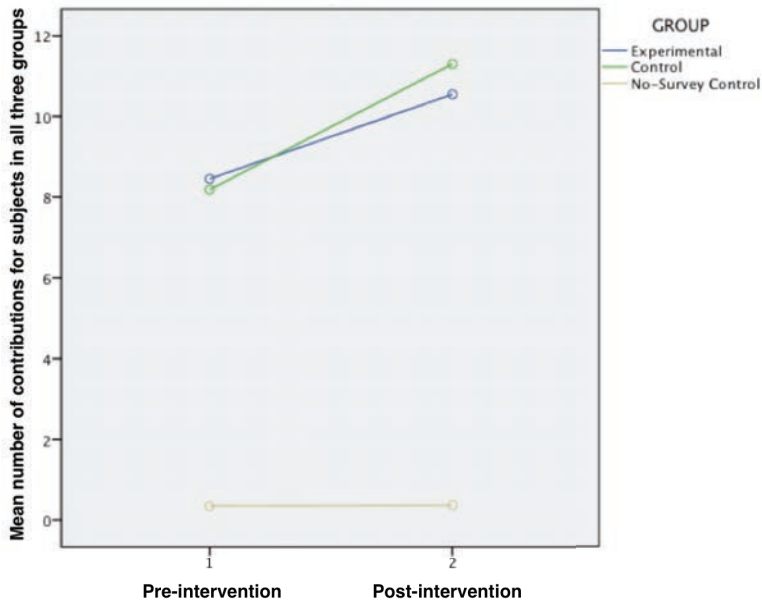
Figure 25: The pre-intervention and post-intervention contributions among *group 1, group 2,* and *group 3* subjects

For the *experimental group* (*group 1*) alone, the mean number of contributions increases from 8.45 posts per user pre-intervention to 10.55 posts post-intervention. The median increases from 3 to 4 in the period. The observed change is not significant as t(50) = -1.688, p < 0.098, Cohen's *d* = 0.34. Thus, the intervention only caused a non-significant change with low effect size. Contrary to the pre-experiment expectations, setting a goal does not lead to increased participation among the subjects.

In the *control group* (*group 2*) the number of posts per subject increases from 8.18 posts before the intervention to 11.30 posts after the intervention. The median increases from 3 posts in the pre-intervention period to 7 posts post-intervention. The paired samples t-test shows that this change is significant as

282

t(33) = -2.639, p < 0.013, Cohen's *d* = 0.84. Contrary to expectations, the change in the control group is significant with a high effect size. In other words, answering a survey that does *not* contain a question about participation level increases the subjects' participation significantly.

For the *non-survey control group (group 3)* – comprising all users who have not answered the survey – the mean number of contributions increases slightly from 0.35 posts per users in the period before the intervention to 0.36 posts after the intervention. The median remains at 0 posts. This change is not statistically significant, as t(3998) = -0.421, p < 0.674, Cohen's *d* = 0.00. An effect size on 0.00 shows that this control group is valid, and no important nuisance variables are at play in the control group.

In sum, only the change in *group 2* is significant. This contradicts the hypothesis and theoretical assumption of this experiment. Goal-setting theory might be able to explain this surprising finding and add further granularity to the study. A possible explanation for the non-significant findings in *group 1* is that the group includes all subjects from *group 1*, regardless of the goal they committed to. Those subjects who commit to a low goal might have fulfilled their goal and, thus, contributed less. In the next section, *group 1* is limited to subject who made a higher goal commitment than their pre-intervention contributions.

7.4.4. Goal-setting among Subjects in the Experimental Treatment

The response categories are created with a high anchor value, thereby increasing the likelihood of subjects committing to more contributions than they made in the two weeks prior to the intervention.

The first interesting aspect of this data is how high or low the subjects in the experimental condition set their goal. When asked how many contributions they expect to make the following 14 days, 41.2 % of the subjects chose a lower number of posts than they had contributed during the previous two weeks. 25.5 % of the subjects selected a number of posts identical to their contributions the pre-

vious two weeks, and 33.3 % set a higher goal. The anchor effect does not seem to have any impact on the subjects. 33 % of the users committing to a higher goal could just as well be a result of random goal assignment.

As noted above in section 7.4.3.2, one explanation for the lack of significant change in *group 1* is that the subjects' move toward their goal commitment: those setting a high goal contribute more, and those setting a low goal contribute less. This entails that the overall effect of the experimental treatment is cancelled out across subjects. To examine this possibility, the subjects are segmented based on their goal commitment compared to posts the previous 14 days. This creates three segments: (1) subjects who set a higher goal than their participation the previous 14 days, (2) subjects who set a goal that maintains participation status quo, and (3) subjects who set a lower goal than their participation the previous 14 days.

Goal		N	Mean	Median	Std. Dev.	Minimum	Maximum
Higher	**Before**	22	8.36	3.5	11.512	0	47
	After	22	13.7727	9.5	12.62625	0	46
Status Quo	**Before**	15	7.27	1	20.002	0	79
	After	15	4.8667	2	10.14091	0	41
Lower	**Before**	14	9.86	1	21.386	0	74
	After	14	11.5714	0	21.62874	0	71

Table 39: Participation among *group 1* subjects segmented on the relation between pre-survey participation and goal level in the survey

As noted, the *group 1 subjects* are divided into three segments based on their current goal compared to previous contributions. As a result, the group sizes are small: n = 22 for *higher goal subjects*, n = 15 for *status quo goal subjects*, and n = 14 for *lower goal subjects*. The three groups are summed up in Table 39.

None of the three sample sizes fulfill the central limit theorem. For the *higher goal group*, the Kolmogorov-Smirnov tests shows $D(22) = 0.238$ $p < 0.002$ for the post-intervention contributions. For the *status quo goal group*, $D(15) = 0.428$ $p < 0.000$ for the post-intervention period. For the *lower goal group*, the Kolmogo-

rov-Smirnov tests shows D(14) = 0.334 p < 0.000 for the post-intervention contributions. As the data is not normally distributed, the non-parametric variation of the paired samples t-test, the Wilcoxon signed ranks test, is used. The results are listed in Table 40.

Group	Z	Assymp. Sig. (2 tailed)	Effect size R (r = Z/√ n)
Higher	-2.934	0.003	-0.63
Status Quo	-0.357	0.721	-0.09
Lower	-.567	0.571	-0.15

Table 40: test Results of the Wilcoxon signed ranks test among Group 1 subjects

7.4.4.1. Subjects Committing to a Higher Goal

For the *higher goal subjects*, the mean increases from 8.36 posts in the pre-intervention period to 13.77 posts in the post-intervention period. The median also increases from 3.5 posts before the intervention to 9.5 posts afterward. This is a substantial difference. The Wilcoxon signed ranks test shows that the change is significant, as T = 26.50, Z = -2.934, p < 0.003, r = 0.623. An effect size on 0.63 is considered high, entailing that the goal commitment has a strong impact on the subjects' subsequent behavior. In other words, the subjects committing to a higher goal in the survey contribute significantly more after the intervention. This result supports the claims made on the basis of goal-setting theory.

The correlation between goal and number of posts written post-intervention is high, but non-significant for those subjects committing to a higher goal. The non-parametric correlation, Spearman's Rho, shows that r = (21) 0.606, p < 0.116. This indicates that the subject's goal is a poor predictor of her subsequent participation behavior. The high r might suggest that the lack of significance is caused by low n.

7.4.4.2. Subjects Committing to a Status Quo Goal

For the subjects setting a *status quo goal*, the mean declines from 7.27 posts in the pre-intervention period to 4.87 posts in the post-intervention period. Despite this decreased mean, the median increased from 1 post before the intervention

to 2 posts afterward. The Wilcoxon signed ranks test shows no effect in this group, as T = 29, Z = -0.357, p < 0.721, r = -0.09. An effect size on 0.09 is considered small. Thus, the subjects who set a goal similar to their earlier contribution number do not change their behavior significantly. This is what goal-setting theory leads us to expect.

The correlation between goal and number of posts written post-intervention is high and significant. The non-parametric correlation, Spearman's Rho, shows that r = (14) 0.614, p < 0.015. Thus, the goal seems to be a reasonable predictor of subsequent participation.

7.4.4.3. Subjects Committing to a Lower Goal

For the subjects setting a *lower goal* than their previous participation level, the mean surprisingly increases from 9.86 posts in the pre-intervention period to 11.57 posts in the post-intervention period. Despite this increasing mean, the median decreases from 1 post before the intervention to 0 posts afterward. The Wilcoxon signed ranks test shows no significant effect in this group, as T = 14, Z = -0.567 p < 0.057, r = -0.15. An effect size of 0.15 is considered small. Although the Wilcoxon signed ranks test is borderline significant, the subjects setting a low goal do not change their behavior significantly afterward. This contradicts goal-setting theory as goals ought to be a predictor of future behavior.

The correlation between goal and number of posts written post-intervention is high and significant for those subjects committing to a lower goal. The non-parametric correlation, Spearman's Rho, shows that r = (13) 0.846, p < 0.009. Thus, the goal seems to be a good predictor of subsequent participation.

7.4.5. Summary of Results

The paired samples t-test for all subjects in *group 1* shows no significant change in contributions after the intervention. Taking the subjects' goal into consideration provides better insights. The *subjects setting a higher goal* are contributing significantly more after setting the goal. This result is what goal-setting theory

predicts. Despite a high effect size, no significant correlation can be detected between goal and behavior.

As expected, the *status quo goal subjects* do not have significantly different participation patterns after the intervention. The Spearman's correlation shows, however, that goal is a reasonable predictor of subsequent participation.

Contrary to the expectations and predictions of goal-setting theory, the *lower goal subjects* increase their participation after the intervention, although the change is non-significant. The Spearman's correlation does show that there is a strong association between the two variables. In other words, goal-setting theory predicts a decreased participation from *lower goal subjects*, but no significant effect is detected.

	Hypothesis	Evaluation
H_0	Setting a participation goal has no effect on subsequent participation	Rejected
H_a	Setting a participation goal makes subjects strive to achieve the goal	Partially verified

Table 39: Hypothesis evaluation for the two hypotheses in experiment 2

As the high goal subjects contribute significantly more post-intervention, the null-hypothesis is rejected. A serious caveat is that the change in the survey control group (*group 2*) is also is significant, thereby casting doubt about the validity of the experiment. The alternative hypothesis is partially verified due to the significance in the high goal subject segment and no significant change in the low goal segment.

7.5. Limitations

7.5.1. Internal Validity

The data analysis shows that there is no significant difference between the two survey groups, but vast differences between the each of the survey group and the non-survey group. Another problem with this experiment is that the change in one of the control groups (*group 2*) is significant. This shows that one or more

spurious variable might have affected the results. There are two plausible reasons for this: the Hawthorne effect[225] and self-selection bias. The *first* explanation is that completing a survey might affect the subsequent behavior, cf. the Hawthorne effect (Adair 1984). If this is true, the survey questions had no effect, but merely taking the survey did. This explanation is supported by the fact that no significant change is detected in the second control group (*group 3*) that consists of *non-survey* subjects.

The *second* explanation is self-selection bias. As the survey respondents – and thus experimental subjects – chose to answer the survey, they might systematically differ from those who did not volunteer. This can explain the different behavior among survey takers vis-à-vis non-survey takers.

Regardless of the reason for this change in behavior among the survey control group subjects, it is still an indication that even the significant changes within *group 1* must be interpreted with caution. There is a risk that the observed changes are wholly or partly caused by spurious variables and not the independent variable. On the positive side, no significant changes are detected in the *no-survey control group* (*group 3*). This indicates that the overall test environment is stable, and that no nuisance variables are interacting outside the two survey groups.

7.5.2. External Validity

As internal validity is a prerequisite for external validity, this experiment holds little external validity. Due to the significant changes detected in the control group it is not recommended that the findings of this study are transferred to other settings. Furthermore, the self-selection sampling makes it very hard to make claims about how the results of this experiment extrapolate.

---

[225] The Hawthorne effect is defined as "the problem in field experiments that Ss' knowledge that they are in an experiment modifies their behavior from what it would have been without the knowledge." (Adair 1984: 334).

The external validity of the experiments is thoroughly discussed in discussed in Section 10.3.1.

7.5.3. Construct Validity

This experiment seeks to increase contributions to K10 by asking the subjects to set a goal for their future contributions to K10. No valid significant effect is detected across the dataset. Thus, using survey as a mean of intervention does not yield the expected results in this experiment. From a technical perspective, Survey Monkey is useful for assigning subjects to different treatments. The lack of valid significant results is more likely caused by the research design.

A second concern about the operationalization of goal-setting theory in this experiment is that no feedback is given after the experimental intervention has been conducted. As explained in Chapter 4, ongoing feedback on goal-fulfillment is an important factor in goal-setting effectiveness. When the subjects cannot easily track their performance, it is hard to know if the goal is met. The lack of significant results among the subject committing to a lower goal might thus be caused by the specific construct of the theory. A more advanced theory construct might show results more in line with the theoretical predictions. Another and more complicated research design is necessary to provide this feedback. This shows that simple proxy experimentation has its limitations in thorough theory operationalization. As described extensively, the experiments conducted and reported here are aimed at improving practice rather than rigorously falsifying or verifying theories. Therefor, these results do not make give any indications about the theory *per se,* but only of the applied construct.

7.6. Implications for Open Government Community Design

The statistical tests in this experiment do not produce any unambiguous insights into goals effects' in open government communities. However, assuming the significant increase in contributions among the *high goal* subjects is valid, this result offers a number of useful implications for design.

The most important implication for open government communities is that simply asking users how much they will contribute might increase participation. That is, as long as the users commit to a goal that is higher than their current participation effort. Interestingly, this experiment did *not* find that subjects who assigned themselves a low goal participated significantly less than they did before. If this finding extrapolates to other open government communities, asking subjects to set goal is a low-risk design strategy as it has a positive effect among some user groups, without having a negative effect on other user groups.

7.6.1.1. Implementation Possibilities

On a practical design level, goal-setting can be implemented into open government communities in several ways. Users can be asked to commit to a goal when signing up. As noted in section 4.3.5, Drenner et al. (2008) examine a design pattern close to this. Using cognitive dissonance theory, they make the barrier of entry for new users higher. It makes fewer people complete the sign-up process, but those subjects who do complete this process contribute significantly more post sign-up. Existing users can also be asked to make a commitment regularly (e.g. once a month or once a year) through surveys or interface design. In smaller communities, it is also an option to do direct outreach to community members, and ask them how much they expect to contribute.

In the theory operationalization in section 4.4.3, it is described how this experiment relies on self-assigned goals. If implemented on an online community, other types of assignments could also be used. The community owner could define the participation goals, or the community itself could set goals for the members and community as a whole.

As described in section 4.4.3, feedback tends to improve the effect of the goal. In this experiment, no feedback is provided to the subjects. But if the concept is turned into design, it is likely to increase the participation level further by

providing feedback on the goal. This can be done by having a post counter or sending out notifications to the users with a status on how they are performing.

One could also consider making both the goal and the status public. Making the goals public is likely to affect both the chosen goal and the efforts to attain them. This would combine goal-setting theory with social-comparison theory as the users will be able to make some comparisons. As described in chapter 6, this might reduce the above-means contributors' participation level due to a regression toward the mean.

The broader implications for open government, e.g. participation patterns, opportunity costs, and designing for niche audiences, are further discussed in section 10.2.

7.7. Chapter Summary

This experiment seeks to make K10 users commit to a high participation goal. Based on goal-setting theory, it is expected that subjects will increase their participation effort in order to meet their self-assigned goal.

With the means of a survey, subjects exposed to the experimental treatment are asked how many contributions they will make to K10 during the following 14 days. Compared to a control group of survey respondents who are *not* asked this question, there is no significant difference in the experimental group's participation level post-intervention. Curiously, the subjects in the control group contribute significantly more. This questions the validity of the experiment.

The second part of the analysis looks only at the subjects in the experimental treatment and separates them into three segments based on their assigned goal: *higher goal*, *status quo goal*, and *lower goal*. For the subjects who set a *higher goal*, the participation is increased significantly after the intervention with a large effect size. There are no significant changes among the *status quo subjects* and *lower goal subjects*. Nonetheless, the correlation between the goal and post-

intervention participation for all three groups is relatively high, ranging between 0.614 and 0.846.

The contribution of this chapter is that goal-setting theory can help improve participation in K10, an open government community. However, due to the significant changes in the *survey control group*, no firm conclusions are made. It does, however, seem likely that K10 users who commit to a higher goal participate more in order to meet the goal. This entails that it might be a good strategy to make users set participation goals for themselves. This is, however, only effective insofar one succeeds in getting the users to commit to higher goals.

The limited valid results are disappointing. But as discussed further in chapter 10, the majority of online experiments do not yield improvements. Not detecting significant results is part of experimentation, as well as one of the main reasons for conducting experiments in the first place. If one is certain about the effect of an intervention, experiments are rendered irrelevant.

Chapter 8: The Effect of K10 Users' Self-Efficacy - An Email and Message Experiment

8.1. Introduction

Different social psychology theories rely on different types of emotions. The first experiment (reported in Chapter 6) leverages social comparison theory. Some of the subjects responded negatively to this kind of transparency, comparison, and competitiveness. An alternative and more positive way to try to increase participation is to increase the subjects' self-efficacy.[226] If one believes she has the chance to succeed in a specific situation, she will approach the task differently than she would if she does not hold such beliefs (op cit.). Interventions that rely on emphasizing positive emotional states might also be more appropriate in non-competitive contexts such as K10 and other open government communities.

This experiment aims to improve the subjects' self-efficacy by thanking them for their invaluable contributions to the K10 community. The purpose of thanking the subjects is to help them build a conviction that they have a positive impact on K10, and that their contributions matter to other people. More specifically, this experiment seeks to give feedback on previous performance accomplishments in order to increase self-efficacy and thereby subsequent participation. According to the theory (op cit.), previous performance accomplishment is the most important information source to people's self-efficacy.

This chapter offers two contributions. The first contribution is a substantial contribution encompassing knowledge about how a certain construct of self-efficacy theory affects participation in an open government community. The second contribution is a methodological one stemming from the application of email and online messages as experimental instruments.

---

[226] Self-efficacy is defined as "the conviction that one can successfully execute the behavior required to produce the outcomes" (Bandura 1977: 193).

8.1.1. Research Question and Hypotheses

This experiment examines the effect of providing self-efficacy promoting information to K10 users. The experiment seeks to answer the following question: *How does knowledge of other users' gratitude for previous contributions affect future contributions?*

This hypothesis helps show how positive emotional states can be used in open government community design. One null hypothesis and one alternative hypothesis are tested in the experiment:

H_0 Knowledge about users gratitude for the subjects' previous contributions does *not* affect the subjects' future contributions

H_a Knowledge about users gratitude for the subjects' previous contributions affects the subjects' future contributions

In this experiment, the *independent variable* is thanking the subjects. This independent variable has two levels:

- Experimental group: receives an email and message at K10 containing the thank you note
- Control group: receives no email

 The *dependent variable* is number of post-intervention contributions. The dependent variable is measured two weeks after each intervention round.

8.2. Research Design

In the survey-experiment reported in the previous chapter, all respondents are asked the following question: "Is there a K10 member who does something particularly good for other users?" From the 84 answers in the survey, 29 different users are mentioned. In this experiment, these 29 subjects are explicitly thanked for their contributions to K10. As K10 users have selected the subjects who are being thanked, this experiment is a field experiment with high realism: The sub-

jects, who are being thanked for their previous contributions actually deserve to be thanked.

Containing only 29 subjects, this experiment faces a high risk of Type II errors due to small sample sizes. The best way to mitigate this risk is to use the research design most suitable for small sample sizes: within-subject design.[227] This experimental research design allows all subjects to be used in both experimental and control conditions.

8.2.1. Within-Subject Design

In this experiment, each subject is used in both the *experimental group* and *control group*. In the first round of the experiment, *group 1* subjects are exposed to the experimental condition and *group 2* subjects to the control condition. This is reversed in the second round, in which *group 1* subjects are exposed to the control condition and *group 2* subjects to the experimental condition. The research design is summed up in Table 41.

	T_0	T_1
Group 1	Experimental Treatment (email/message sent)	Control Treatment (no email/message)
Group 2	Control Treatment (no email/message)	Experimental Treatment (email/message sent)

Table 41: Within-subject design used in experiment 3

The greatest concern with within-subject design is potential carryover effects. As shown in Table 41, the carryover effect is only a concern for *group 1* subjects, as they are exposed to the experimental condition in the first round. In this experiment, carryover effects are likely to occur as long as the email and/or message is present in the subjects mind. As nothing happens in the control condition in the first round, no carryover effect can take place for *group 2 subjects*.

---

[227] Please refer to section 5.8 in for a review of the different research designs' comparative advantages and disadvantages.

Carryover effects are complicated because they are unknown at the time of the intervention. The time between the two experimental rounds ought to be long enough to reduce carryover effects, and short enough to ensure that the experimental context does not change between the rounds. For this experiment, two weeks are considered to strike an appropriate balance between these two concerns.

8.2.2. Sampling and Subject Assignment

As noted above, 29 K10 users are mentioned in the survey as being particularly helpful to other users. This experiment is conducted 40 days after the survey. In the time between the two experiments one member died, and, curiously, another was banned from the community. This leaves a sample of 27 users. These 27 users are used as experimental subjects in this experiment. Using this non-random sample instead of a random sample is appropriate in the field context. As mentioned, it improves the realism of the experiment and strengthens the internal validity. A non-random sampling does not compromise the internal validity or construct validity insofar the subjects are randomly assigned to groups. As this experiment is a within-subject experiment, all users are used in both the experimental treatment and the control treatment. The 27 subjects are divided into two groups, using Research Randomizer.[228] Based on this randomization, *group 1* has 14 subjects, and *group 2* has 13 subjects.

There are two caveats of this sampling procedure. *First*, this sample methodology is not random. Differences might exist between the users who are thanked and those who are not. External validity is sacrificed to improve the realism and internal validity of the experiment. *Secondly*, merely 27 remaining users are mentioned in the survey-experiment. The sample size is therefore small in this experiment.

---

[228] http://randomizer.org

8.3. Conducting the Experiment

8.3.1. Intervention: Email and Message Design

In this experiment, two different channels are used to cause the experimental intervention: email and internal community messages.

8.3.1.1. Emails

As noted in Chapter 6, the click through rate (CTR) in emails is usually modest. In the experiment reported in the previous chapter, the email CTR is 8.16 %. This CTR is acceptable in the first experiment, as it is conducted on a large user database. Since this study only contains 27 subjects and no dynamic variables except username, the emails can be hand crafted. This gives greater flexibility in terms of email delivery services. Flexibility is positive as it enables the use of an email service that tracks open rates. Thus, whereas it was not feasible to track email *open rates* in a reliable manner in Experiment 1,[229] this experiment now has that option.

The ability to track open rate instead of click through rate is beneficial to this experiment, as email open rate is higher than click through rate. It is therefore more likely to acquire a sufficient number of subjects in the sample. According to the MailChimp industry benchmark report (Mailchimp 2010), the average open rate in Social Networks and Online Communities is 22.37 %. The same study indicates that the open rate is often 8 – 10 times higher than the CTR. The first experiment has an 8.16 % CTR. If the findings from the MailChimp benchmark study extend to K10, the expected open rate is 8 – 10 times the 8.16 % CTR. This equals a 64 % – 82 % open rate. Assuming that the open rate reaches this interval, the experiment ought to have a sufficiently large sample in order to keep the risk of Type II errors at a reasonable level.

---

[229] Due to technical challenges with a large number of emails all containing dynamic variables (op cit,).

The email program used, YourMailingListProvider,[230] places an invisible 1x1 pixel in the email. When the email is opened, this pixel loads and sends the information back to the email program. The email program then knows which email recipients have and have not opened the email.[231]

8.3.1.2. Internal Community Messages

Solely relying on emails as the mean of intervention is risky in this experiment. Even though the average open rate assumed to be in the 64 % – 82 % interval, the actual number might be different in such a small sample. Attempting to reduce this risk, the message is delivered via an additional medium: internal community messages on K10. The internal community message system works like messages in other online communities and social networks such as Facebook, LinkedIn, and Twitter. Each user has a message inbox and outbox used for private communication with other members. Each subject is potentially receiving the message on two platforms.

Screenshot 10: Delivery confirmation prompt for internal messages at K10

---

[230] http://ymlp.com

[231] One problem with this tracking methodology is that it is based on HTML. The pixel is only activated if the email is opened in an email client supporting HTML and allows images to be opened. Although all major email services such as Gmail, Hotmail, Outlook, and Mail allow this, there is still a risk that some emails will be opened without being tracked.

The internal message system at K10 is running on vBulletin. vBulletin allows senders to ask for confirmation that the message has been read. However, this tracking relies on the recipients accept. As seen in Screenshot 10: Delivery *confirmation* prompt for internal messages at K10, the recipient is shown a prompt asking her to confirm that she has read the message. However, the recipient can read the message without confirming it. Only subjects confirming that they have read the message are included in the experiment.

In sum, both delivery systems have a risk of false negatives, as they potentially underestimate how many people read the emails and messages. This leads to smaller group sizes, but does not compromise the validity of the experiment, since the subjects are removed from the dataset, rather than transferred to control groups.

As this experiment is a within-subject design, the experimental intervention is conducted twice. The first round is conducted on March 29th 2012, where 14 subjects receive the message and email. The remaining 13 subjects receive them during the second round 14 days later, on April 12th, 2012. The message and email contains the following content:

Hi [username],

Last month, we conducted a survey on K10. Among the questions was this: *Is there a K10 member who does something particularly good for other users?*

Several users highlighted you for your contributions in the community.

So we just want to let you know that other users are aware of you and appreciate your contributions to K10.

If you have any questions, you can just reply to this message.

Best regards,
K10

Ps. Many users also wrote that K10 pull together – and that all users contribute in their own way. We agree. We do not wish to highlight some members over others, but rather express our gratitude for your contributions to K10. It means a lot to other people.

The two delivery systems, email and messages, have very different open rates. As shown in Figure 26, delivery via K10's internal message system is more effective than email.

27 emails were sent and 23 delivered. Four could not be delivered because the subjects' inboxes were full. Out of the 23 delivered emails, 9 were opened. One user replied to the email. All 27 messages sent via the K10 message system were delivered. 25 subjects confirmed they had received and read the message, and five subjects wrote a reply.

	Email	Message
Sent	27	27
Delivered	23	27
Opened	9	25
Not Opened	14	1
Mean Open	92.9 %	33.3 %
Std. Deviation Open	0.480	0.267
Std. Error Mean Open	9.092	0.051
Replies	1	5

Table 42: Delivery and open rate for emails and messages

As Table 42 and Figure 26 show, the email open rate is merely 33 %, or half of the estimated number of opens. Nine subjects are insufficient to make any meaningful claims. However, the double delivery strategy pays off, as the message open rate is 92.9 %.

	N	Mean Rank	Sum of Ranks
Negative Ranks	16	8.5	136
Positive Ranks	0	0	0
Ties	11		
Total	27		

Table 43: Wilcoxon signed ranks test for email and message open rates

Although this experiment does not seek to study and compare different delivery forms, it is relevant to examine the date briefly. The data does not meet the assumption about normal distribution, as a Kolmogorov-Smirnov tests shows $D(27) = 0.535$ $p < 0.000$ for the messages, and $D(27) = 0.423$ $p < 0.000$ for the emails.

Open Message – Open Email	
Z	-4.000
Asymp. Sig. (2-tailed)	0

Table 44: Results of Wilcoxon signed ranks test for email and message open rates

As the two groups have $n < 30$ and fail to meet parametric assumptions, it is best analyzed through the Wilcoxon signed ranks test. This test shows that messages sent via the internal K10 message system are significantly more likely to be opened than a message sent through emails: $T = 0$, $p < 0.000$, $r = 0.77$. This effect size is considered very high. Although the external validity is unknown, internal community messages are a significantly more effective communication medium than email at K10.

8.4. Data Analysis

Before addressing the results of this experiment, the dataset and sample distribution are described.

8.4.1. Cleaning the Dataset

As it is the case in the two former experiments, this dataset does not meet the assumption about normal distribution. To improve this fit, four users with zero contributions in both the pre-intervention and post-intervention period are removed from the dataset. This leaves a dataset with 23 users: 12 in *group 1*, and 11 in *group 2*.

After the data collection, the data is exported from the K10 MySQL database into excel. In here, the numbers of pre-intervention and post-intervention contributions are calculated for each user. The sums are placed in two columns: pre-intervention post count and post-intervention post count.[232]

8.4.2. Sample Distribution

The Kolmogorov-Smirnov test of normality shows that the data is borderline significant.

---

[232] To make meaningful statistical comparisons, the conditions of the *group 2* subjects are reversed so there is one column containing subjects' post count under the experimental condition, and one column containing subjects' post count under the control condition.

For the *no thank you* intervention, the test of normality shows that $D(23) = 0.178$ p < 0.056. The test for the *thank you* intervention variable is $D(23) = 0.185$ p < 0.040. Although the difference is small between the two samples, the former meets the parametric assumptions, but the latter does not.

Mean = 24.22
Std. Dev. = 21.538
N = 23
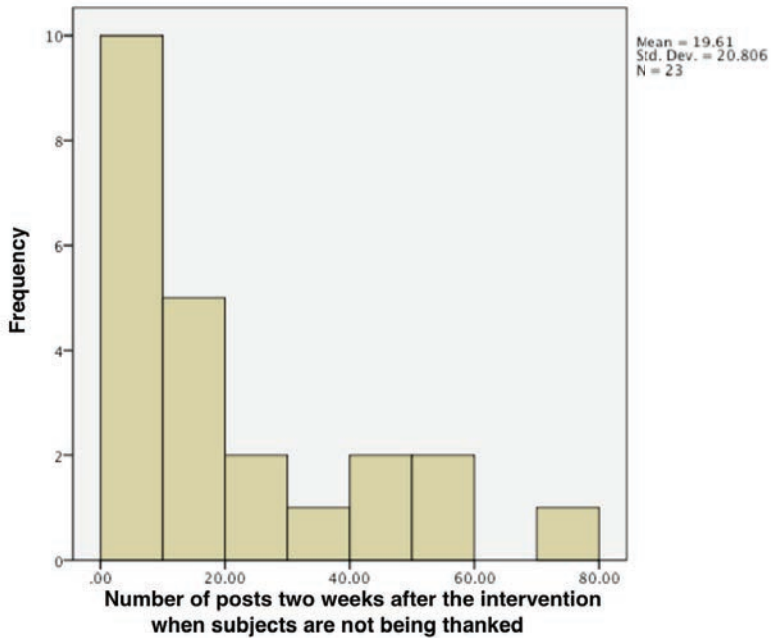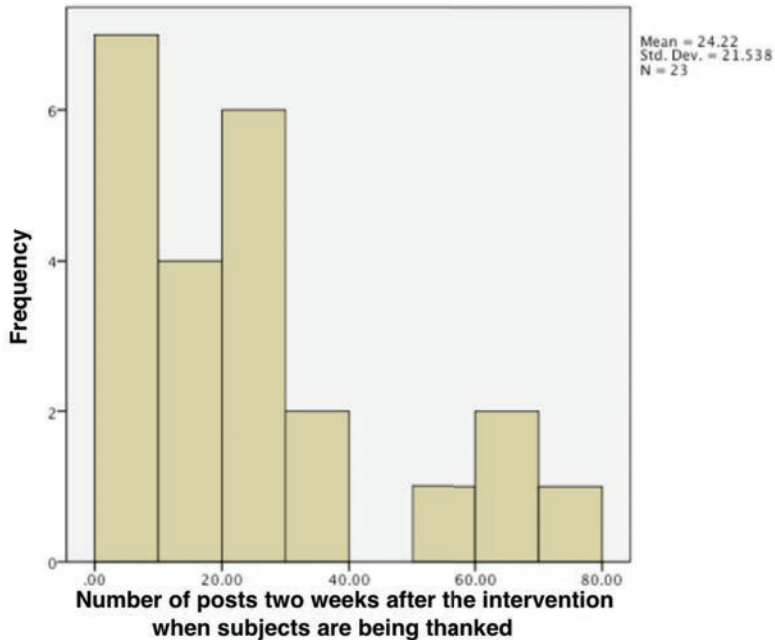
Figure 28: Participation distribution when subjects are being thanked

The tests for this data are fairly simple: the parametric test is a paired samples t-test, and the not-parametric equivalent is a Wilcoxon signed ranks test. A decent solution for such data is to run both a parametric and a non-parametric test. Therefore, the dataset is analyzed with both statistical tests.

8.4.3. Results

As shown in Table 45, the 23 subjects wrote a total of 451 posts when they were subjected to the *no thank you* treatment and a total of 557 posts when assigned to the *thank you* treatment. Table 45 shows that the mean contribution for the subjects is 19.6 posts over a 14-day period, when the subjects *do not receive thank you emails* and messages. The median is 12 posts per subject. When the subjects did *receive a thank you email*, the mean number of posts per user in a 14-day period is 24.2 with a median of 23. The minimum and maximum observations are consistent across the two treatments: the minimum observation in both

treatments is zero and the maximum in the pre-intervention and post-intervention is respectively 71 and 75.

	No Thank You Treatment	Thank You Treatment
Number of posts	451	557
Mean	19.6087	24.2174
Median	12	23
Minimum	0	0
Maximum	71	75
Std. Deviation	20.8059	21.53846
Std. Error of Mean	4.33833	4.49108
N	23	23

Table 45: Pre-Intervention and post-intervention participation among all subjects

Both the mean and the median increase post-intervention. This indicates that explicitly acknowledging users' impact and value affect their subsequent behavior. As seen in the two histograms above in Figure 27 and Figure 28, the intervention changes the participation pattern. Several subjects are moving from low contributors to medium contributors, which explains why the median value is almost doubled, compared to pre-intervention level.

	N	Mean Rank	Sum of Ranks
Negative Ranks	16	12.84	205.5
Positive Ranks	7	10.07	70.5
Ties	0		
Total	23		

Table 46: Paired samples t-test for all subjects

The paired samples t-test shows that the difference between the means in the *experimental group* and *control group* is significant $t(22) = 2.186$, $p < 0.040$, Cohen's $d = 0.454$. This means that thanking the subjects makes them participate significantly more in the following two weeks. An effect size on Cohen's $d = 0.454$ is considered to be slightly below medium.

There is a high correlation between the pre-intervention and post-intervention treatment, as $r(23) = 0.885$, $p < 0.000$. This shows that many subjects' participate incrementally more after the intervention.

Test Statistics	
Z	-2.055
Asymp. Sig. (2-tailed)	0.040

Table 47: Paired samples t-test for all subjects

The non-parametric equivalent of the paired samples t-test, the Wilcoxon signed ranks test, also detects a significant difference between the experimental and control treatments: T = 70.5, p < 0.041, r = 0.429. An effect size of 0.429, is considered to be a medium-large effect. In sum, both statistical tests show that the subjects contribute significantly more to K10 when they receive an email aiming to increase their self-efficacy. This indicates that self-efficacy theory can be operationalized effectively into the open government community domain.

8.4.4. Responses

The text in the email and message encouraged the users to reply if they had any questions. No users had any questions about the content. This indicates that the content of the email and message was easy to understand and conveyed the message appropriately. Although no one had any questions, five subjects replied to the message and one subject replied to the email.[233] Extracts from the replies are translated and analyzed in the following sections.

8.4.4.1. Feeling Good

K10 is primarily a community for people suffering from long-term illnesses, as this is a prerequisite for receiving early retirement pension and flex job benefits. User C replied that "Such a pad on the shoulder really makes me feel better... Especially in a time where I struggle to get acknowledged that I cannot do as much anymore – and acknowledge it myself.... It is messages like this that helps me keep going". Two other users expressed similar thoughts. This shows that expressing gratitude can make people feel good.

---

[233] All responses can be found in Appendix 7.

One user wrote that she did not think she deserved the praise more than anyone else. Nonetheless, she was very grateful for receiving the message.

8.4.4.2. Self-Efficacy

As described in Chapter 4, self-efficacy theory argues that having an impact is motivating. The importance of self-efficacy is highlighted by User A, who wrote that "I do not have any question but I want to say thank you for letting me know that people can use my posts". This reply shows that she appreciates knowing that other users find her posts useful.

8.4.4.3. Reciprocity

Another user interprets the message in terms of generalized reciprocity. User E replies "Thank you very much.... I am happy to be able to give something back after all the help I have received here at K10".

8.4.5. Summary of Results

Both the parametric and non-parametric tests find the difference in means to be statistically significant, and the effect size to be medium-large. This means that expressing gratitude and impact to K10 users for their efforts has an impact on the users' subsequent behavior. This indicates that self-efficacy theory can be applied to this domain. Making users feel they are heard, appreciated, and have an impact makes them contribute significantly more to the community.

	Hypothesis	Evaluation
H₀	Knowledge about users gratitude for the subjects' previous contributions does not affect the subjects' future contributions	Rejected
Hₐ	Knowledge about users gratitude for the subjects' previous contributions affects the subjects' future contributions	Verified

Table 48: Hypothesis evaluation for experiment 3

The results are not highly significant as the probability level is $p < 0.04$, entailing a 4 % risk that the variance is caused by nuisance variables and not the independent variable. According to the two effect size calculations, the effect size is either just below medium or between medium and large. This might indicate that the relatively high p-value is a result of the low sample size. Carryover effect is another potential reason for the quite large p-value. This possibility is discussed below.

The qualitative responses indicate that expressed gratitude can be understood not only in terms of self-efficacy, but also generalized reciprocity. On a higher level, this message underscores that participation motivators are complex and multidimensional, and cannot be limited to a single variable. But this concern is more relevant to research than to practice.

8.5. Limitations

This experiment has three limitations in terms of internal validity, construct validity, and external validity.

8.5.1. Internal Validity

One caveat of within-subject design is potential carryover effects.[234] This effect entails that subjects exposed to the treatment in the first round of intervention are influenced by the treatment when exposed to the treatment in the second intervention round. Carryover effect is a complicated matter since the extend is often unknown.

---

[234] Carryover effects are explained section 5.8.2.1.

	No Thank You Treatment	Thank You Treatment
Group 1	158 posts	165 posts
Group 2	293 posts	392 posts

Table 49: Intervention across the two groups shows that carryover effect can occur in *group* 1 but not *group 2*

In this experiment, the carryover effect was estimated to be two weeks. However, the data shows that the subjects in *group 1* (who were introduced to the treatment condition in the first intervention round) have a similar amount of participation in week 3-4, as they do in week 1-2. In week 1-2 post-intervention, the subjects contributed 158 posts. In week 3-4 after the intervention (when these subjects functions as control group), they contributed 165 posts.

The pattern for *group 2* is different. The subjects in this group are used as a control group in week 1-2, and as treatment group in week 3-4. This group wrote 293 posts in week 1-2 and 392 posts in week 3-4.

This suggests that the carryover effect is longer than 14 days. This does not undermine the experimental logic. It does, however, underestimate the effect of the experimental intervention in terms of both significance and effect size. If the experiment is replicated with more time between the interventions - for example 6 or 8 weeks – the statistical significance and effect size are expected to be improved.

In other words, the potential carryover effects lead to increased likelihood of Type II errors: accepting the null hypothesis when it ought to be rejected. But as argued previously in this dissertation, Type II errors are less critical than Type I errors in the scientific communities relevant to this research.

8.5.2. Construct Validity

8.5.2.1. Who is Thanking the Subjects?

As discussed further below, this experiment indicates that it will be beneficial to K10 to integrate expression of appreciation and impact into the design. In this regard, the experiment has a severe limitation: it might not be clear to the user *who* is thanking her - the other users or the website owner. This uncertainty is caused by the experimental design.

As noted above, the survey gave members of K10 the chance to mention other members who they thought were particularly useful and valuable to the community. Based on these answers, the website owner sends a thank you message and email to those who were mentioned in the survey. Thus, the expression of gratitude is done between users, but communicated via and by the K10 owner.

The wording in the message and email clearly states that it is other users who have decided to thank the subject. But the effect of the gratitude might be altered – for better or worse – because it is sent by the site owner. This means that the causality is not clear, and therefore compromises the construct validity somewhat.

8.5.2.2. Theory Operationalization

The email and message expressing appreciation and impact causes a change in the subjects' behavior. Although there are several validated scales to measure self-efficacy (op cit.) they are not used in this experiment. It is therefore unknown if the email and message has indeed increased the subjects' self-efficacy. It is a possibility that other social psychological aspects are causing the change in behavior, e.g. generalized reciprocity. However, one of the subjects replied to the email and explicitly remarked he was happy that he was having an impact. This suggests that self-efficacy is indeed at play. However, as argued, validating theories is outside the scope and purpose of this dissertation.

As described in section 4.4.4, self-efficacy stems from four information sources: performance accomplishments, vicarious experience, verbal persuasion, and physiological states. Although this experiment leverages performance accomplishments, vicarious experience, and verbal persuasion, the most salient one is information about previous accomplishments. Other results might be attained when making one of the other three information sources the most salient.

8.5.3. External Validity

The sample of this experiment is not randomly drawn from the K10 population. Rather, it comprises users who were explicitly thanked by other users in the experiment reported in Chapter 7. This increases the experimental realism, as the subjects truly deserve to be thanked.

It is, however, unclear how these thank-worthy K10 users differ from the users who are not found thank-worthy. Based on this experiment alone, no conclusions can be made about how other users will react. This being said, the important thing is that users who deserve to be thanked appreciate the gratitude and increase their participation subsequently. The external validity of the experiments is further discussed in Section 10.3.1.

As described above, K10 comprises users that are challenged due to illnesses. This probably made them respond more negative to the participatory information conveyed in Experiment 1 (Chapter 6), than healthy people would have. It is equally that the K10 subjects suffer from low self efficacy and thus respond more positively to acknowledgement than healthy people would. As the answer from User C above demonstrates, these subjects are probably not used to a lot of praise. Highlighting performance accomplishment of previous performances might thus be rare for the subjects here. This means the effect sizes detected here might be smaller if replicated on communities with more healthy users.

8.6. Implications for Open Government Community Design

This experiment yields a significant effect with a medium effect size. This means K10 is likely to benefit from implementing a design feature allowing users to be thanked. Again, it is primarily one of the four information sources, previous performance accomplishments, that is made salient in this experiment and thus used in this section. More design ideas can be generated by including other information sources.

Expressing gratitude to users for their contributions, and thus reassuring them of their performance accomplishments, is a popular design practice that is applied in various ways. Recognition of certain behaviors make it possible for other community members to see what kind of actions are appreciated (Kriplean et al. 2008).

8.6.1.1. Automated Recognition

Some online communities offer automated means of recognition. In this way of designing for appreciation, certain behavior leads to certain forms of recognition. An example is provided in Screenshot 11, which shows that certain check-in behavior on Foursquare[235] unlocks badges.

---

[235] http://foursquare.com

Thomas' foursquare Badges

These are the core badges dreamt up by the members of the foursquare team, for things like regular workouts at your gym or being a local at your neighborhood coffee shop.

Newbie
August 25, 2011

Adventurer
August 28, 2011

Explorer
September 03, 2011

Superstar
September 17, 2011

Bender
September 08, 2011

Crunked
September 24, 2011

Screenshot 11: Foursquare badges are an example of automated recognition

Similarly, accumulating a sufficient amount of points awards users recognition on Quora.[236] In some online communities, the automated recognition is paired with privileges. For example, Klout[237] offers benefits such as movie theater benefits to highly influential users.

8.6.1.2. Peer to Peer Recognition

Another way to recognize users is to allow members to thank each other manually. Several big online communities use this design pattern: Wikipedia[238] allows users to give Barnstars for work and behavior they value.[239] See Screenshot 12 for an example. Similarly, Google allows users to thank other users for sharing articles on Google+.

---

[236] http://quora.com

[237] http://klout.com

[238] http://wikipedia.org

[239] http://en.wikipedia.org/wiki/User_talk:ForteTuba. Retrieved on April 2nd , 2012.

I, EWS23, hereby
award **ForteTuba** the
da Vinci Barnstar for
his programming and
implementation of the
ever-useful
SuggestBot. Keep up
the good work! 20:38,
2 July 2006 (UTC)
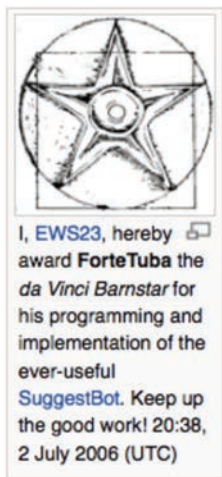
Screenshot 12: Wikipedia barnstars are an example of peer to peer recognition

8.6.1.3. Expert Recognition

A third way to integrate gratitude into design is through expert recognition. As shown in Screenshot 13, Moz[240] has implemented such a feature in their Q&A. Here staff members can endorse an answer written by a user. This entails that both the answer and author receive an editorial approval. One advantage of expert recognition vis-à-vis automated recognition and peer to peer recognition is that the former not only provides performance accomplishment information, but also one of the other self-efficacy information sources: verbal persuasion. A message from an expert ought to be more persuasive than a similar message from a peer, due to the implicit credentials of experts.

---

[240] http://moz.com. Retrieved on April 2nd , 2012.
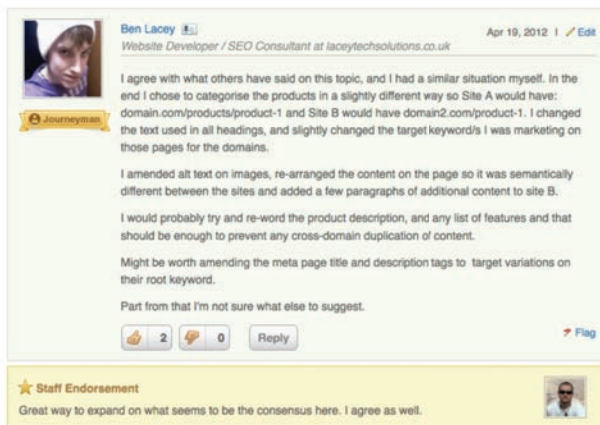
Screenshot 13: Staff endorsement in Moz Q&A is an example of expert recognition

Expressing recognition is not the only implication this experiment has for open government communities. The broader implications for open government, e.g. participation patterns, opportunity costs, and designing for niche audiences, are further discussed in section 10.2.

8.7. Chapter Summary

The first two experiments rely on more competitive and negative feelings, whereas this one relies on a positive feeling: self-efficacy. In a within-subject experiment, 27 users identified in the survey reported in the previous chapter are thanked for their contributions to K10. The paired samples t-test and Wilcoxon signed ranks test both show that the experimental subjects participate significantly more after being thanked. Depending on the method, the effect size is either just below medium or between medium and high. The finding is strengthened by six subjects who replied to the emails and messages to express their happiness after receiving the email. Thus, the main contribution of this chapter is to document that thanking users increases participation in K10.

More by accident than design, this experiment also compares internal community messages with emails. The open rate is found to be significantly higher for the

315

internal message system. No claims are made about the generalizability of this finding.

The experiment has two limitations: a potential carryover effect between the treatments, and uncertainty about *who* is thanking the subject.

Chapter 9: How Does Group Construct Affect Behavior at K10? An Advertisement Experiment

9.1. Introduction

Two types of identities help guide human behavior in social: personal identity and group identity (op cit.). Social identity theory therefore argues that people seek to enhance the groups they belong to in order to enhance their self-image (op cit.). Moreover, people often prefer to contribute to groups comprising similar others.

People are concerned with the groups they belong to. In an open government community context, it might therefore be beneficial to make the group welfare more salient in the users' mind. Based on social identity it can be expected that the beneficiaries of a contribution affect users' participation level. The aim of the fourth experiment is to examine the link between knowledge about beneficiaries of contributions and participation level. Taking the point of departure in social identity theory, the link is examined by highlighting different types of beneficial scenarios: a contribution can be *beneficial to self*, be *beneficial to a small group*, be *beneficial to a large group*, or *no beneficiary* can be highlighted. In other words, the aim of this experiment is to make it clear whom a contribution benefits in order to examine if such knowledge increases participatory and pro-social behavior. For details on operationalization of social identity theory, please refer to section 4.4.5.2.

This chapter offers two contributions. The *first* contribution is a substantial contribution about how a certain construct of social identity theory influences participation at K10, an open government community. The *second* contribution is a methodological contribution resulting from the application of online advertisements as an experimental instrument.

9.1.1. Research Question and Hypotheses

In this experiment the effect of making contribution beneficiary salient is examined at K10. The research question in this experiment is: *How does benefit of contribution affect subsequent behavior on K10?*

This research question can help understand why people contribute to open government communities by examining if knowledge about who benefits from the contribution matters to the contributors. To answer this research question, a null hypothesis and two alternative hypotheses are tested. Based on social identity theory, it is hypothesized that the *beneficial to other* groups will outperform *beneficial to self* group in terms of engagement and contributions. Furthermore, social identity theory predicts that people, under certain conditions, favor contributing to small groups compared to larger groups (Brewer and Kramer 1986). The three statistical hypotheses tested in this experiment are therefore:

H₀ The group constructs have no effect in terms of engagement and contributive behavior.

Hₐ Subjects assigned to the *beneficial to other groups* will engage and contribute more than subjects assigned to *beneficial to self* and *control group*.

H_b Subjects assigned to the *beneficial to small* group will engage and contribute more than subjects assigned to *beneficial to large* group.

In this experiment, the *independent variable* is the benefit of contribution. Building on Beenen et al. (2004) and Rashid et al. (2006), this study uses three levels of contribution benefit, supplemented with a control group where no benefit is highlighted. Thus, this independent variable has four levels:

- *Group 1*: Beneficial to self
- *Group 2*: Beneficial to others (small group)
- *Group 3*: Beneficial to others (large group)
- *Group 4*: No benefit (control group)

No real groups are formed in this experiment. Rather, the group constructs are used as a way of priming the experimental subjects. The priming is done through advertisements in Google,[241] where different ad copies are served to different subjects. When priming the benefit of a contribution to the experimental subjects, the *beneficial to other* is here defined as other people in the flex job benefit program.

In order to examine the effects of the intervention, seven *dependent variables* are measured: click through rate (CTR), new accounts, new posts, new threads, time on site, pageviews, and bounces. The first three experiments in this dissertation rely on only one dependent variable: number of posts. The reason behind the growth in variables in this experiment is that it targets new users rather than existing. This means that the number of posts written is expected to be much lower than in the other experiments, making it harder to detect statistical significance on this variable. Including more variables will give a more nuanced view of new behavior from the four groups, and it increases the likelihood of identifying significant differences across the four groups. The seven dependent variables are described in section 4.4.5.2.

Contrary to the first three experiments reported in this dissertation, this experiment not only includes existing members but also seeks to engage new users in K10. This means that the experimental subjects are not necessarily members of the K10 community when the experiment is conducted. They are, however, likely to be part of the flex job and early retirement pension social groups, and thus have a common identity that overlaps with K10.

9.2. Research Design

Using Google AdWords to manipulate the independent variable limits possible research designs in three ways. *First*, although it is possible to see how many users saw the advertisement, the population demographic is unknown. *Second*,

---

[241] For a full description of Google AdWords, please refer to section 5.10.4.4.

the same subject cannot deliberately be exposed to multiple treatments. This effectively eliminates the possibility of within-subject design. *Third*, there is no data about pre-intervention user behavior. This effectively eliminates the possibility of mixed design.

The only attainable experimental design is between-subject design. This design complements this experiment, as advertisement gives access to a large pool of potential subjects. Running a between-subject design is more expensive than within-subject design because more subjects are needed. But as described further below, the price for recruiting subjects via AdWords is low.

9.2.1. Between Subject Design

As noted, the research design in this experiment is a between subject design. Each user is randomly assigned to one of the four groups, and thus exposed to one of the four treatments. The main challenge of this research design is that the subjects in each group are different.

	Group 1	Group 2	Group 3	Group 4
T_0	Beneficial to self	Benefit to others (small group)	Benefit to others (large group)	No benefit (control group)

Table 50: Between-subject design used in experiment 4

As described below, Google does not offer any behavioral data in their AdWords program. It is therefore impossible to know if any differences between the groups exist. The statistical tests performed in the data analysis helps identify the risk of such biases.

9.2.2. Sampling and Subject Assignment

Subjects are recruited through Google AdWords. As this is an online field experiment, the recruitment is taking place in a real environment. The subjects in this study includes people that (1) conduct a search in Google.dk using a keyword matching one of those advertisements was bought for, (2) are exposed to this

advertisement, and (3) click on this advertisement. Thus, a Google.dk user is treated as an experimental subject if she clicks the advertisement to K10.

One problem with advertisement platforms such as Google AdWords is that it is biased toward high performing ads. To increase revenue, Google prefers to show ads with the highest CTR. This does not compromise the internal validity of the experiment, as different ad variations are rotated. In this between-subject design, each subject has the same likelihood of seeing any advertisement.

The research design does, however, raise questions about the external validity, as it is unclear which subject group constitutes the population. As described in Chapter 5, an advertisement in Google.dk is triggered when there is correspondence between the keyword used by advertiser and user. This entails a challenge with a self-selecting sample, as people chose to perform a search.

9.3. Conducting the Experiment

9.3.1. Intervention: AdWords Design

Compared to advertisement platforms such as Facebook, Google offers little demographic and behavioral targeting in their AdWords. The following advertisement targeting variables are used:

AdWords Settings	
Location	Denmark
Language	Danish
Network	Google Search (Search partners and display networks are deselected in order to reduce data variance).
Devices	Desktop, mobile, and tablet
Ad rotation	Rotate: Show ads more evenly (This option gives a more even distribution of ad impressions despite higher costs).
Exact match keywords	Flexjob Fleksjob
Phrase match keywords	Flex job Fleks job

Table 51: Advertisement set-up in Google AdWords

Advertisements are bought for "broad match" keywords[242] containing the words Fleksjob and Flexjob, as well as "phrase match" keywords[243] including "fleks job" and "flex job". Based on these keywords and the targeting choices made above, Google estimates 51-63 clicks per day, costing between DKK 1.92 – 2.90 per click.

	AdWords Settings
Average cost per click	DKK 1.92 – 2.90
Average clicks per day	51-63
Average cost per day	DKK 120.92 – DKK 147.79

Table 52: AdWords settings for the campaign in experiment 4

The four ad copies used to measure different levels of the independent variable are listed in Table 52. The AdWords were originally worded in Danish. Please refer to Appendix 8 for the Danish wording of the ad copies, which were shown to Google.dk users.

Group	Ad Copy
Group 1 Highlight individual benefit	**Receive Advices About Flex Job** Get advices and help about Flex Job from others at K10
Group 2 Highlight small group benefit	**Share Your Knowledge About Flex Job** Share your experiences with Flex Job with people like you on K10
Group 3 Highlight large group benefit	**Share Your Knowledge About Flex Job** Share your experiences with Flex Job with other people in Denmark on K10
Group 4 Control group / generic adver-tisement	**Forum About Flex Job** Forum about flex job. Sign up at K10

Table 53: Advertisement copies of the four ads in the experiment

---

[242] Google explains broad match keywords as "When you use broad match, your ads automatically run on relevant variations of your keywords (including misspellings) … Keyword variations can include synonyms, singular and plural forms, possible misspellings, and phrases containing your keywords." (Google 2012a).

[243] Google explains phrase match keywords as "With phrase match, you can show your ad to customers who are searching for your exact keyword and close variants of your exact keyword, with additional words before or after." (Google 2012b).

The experiment is conducted over a two-week period, starting April 8th 2012 and ending April 21st 2012.[244] An example of one of the advertisements as shown in Google search results can be seen in Screenshot 14.
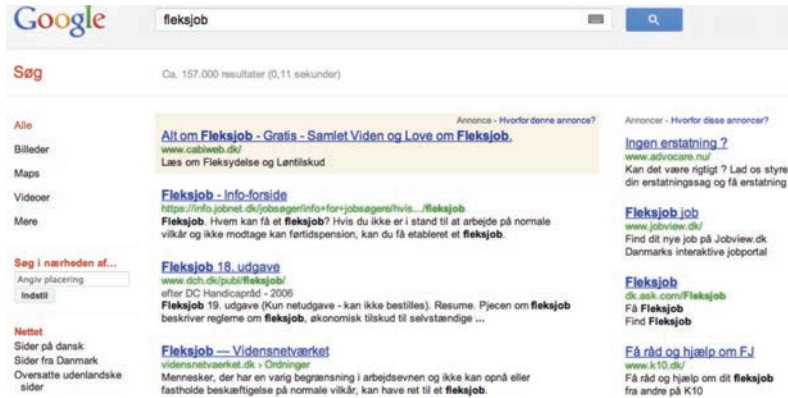
Screenshot 14: K10 advertisement as seen on a Google search result page

The four different advertisements send Google.dk users to the K10 homepage (k10.dk). Each of the four advertisements has unique URL parameters, making it possible to track users from each group to K10. The users are first tracked when

---

[244] To test the price and CTR of the different ad copies, a pilot study was conducted between April 3rd and April 7th ((2012). In this pilot phase, different costs per click, keywords, and wording were tested. Two problems arose in the pilot testing phase: *First*, many companies are bidding for keywords involving "pension". This entails that the cost per click is much higher for early retirement pension keywords than for flex job keywords. The average cost per click for early retirement pension keyword is between DKK 5.08 – 7.68, while the price for a flex job keyword is DKK 1.92 – 2.90. *Second*, the CTR on early retirement pension ads was much lower than the CTR on flex job ads. One explanation is that the advertisement was crowded out by many other advertisements, as keywords containing "pension" draw many advertisers. When ads have low CTR, Google will show them less or not show them at all because it entails opportunity costs for Google. After a couple of days with low CTR, Google basically stopped showing the ads. The original plan for the experiment was to run parallel advertisements on early retirement pension and flex job, but the pilot study shows that it is more efficient to spend the money on flex job advertisements, as more subjects can potentially be recruited within the budget of the experiment.

they click the advertisement, and there is therefore no pre-intervention data of the users in this experiment; hence the between-subject design.

9.4. Data Analysis

Before reporting the results of the experiment, this section describes how the data set is cleaned, the sample characteristics, and the sample distribution.

9.4.1. Cleaning the Dataset

Google Analytics does not provide individual level data for users reached via AdWords. This is a problem when relying on statistical tests, as these rely on the variance on the individual subject level. Averages cannot be used for statistical tests. However, Google Analytics does allow advanced segmentation on metrics such as date, keyword, city, browser, screen resolution, number of pageviews etc. By applying a sufficient amount of filters, it is possible to create conditions that only one subject fulfills. By such advanced segmentation, it is possible to get individual level data for all the subjects, and thus use statistical tests to examine the data. After doing the data segmentation in Google Analytics, the data is exported to excel and prepared for analysis in SPSS.

One outlier was removed from the dataset, as the subject had been on the website for over 3 hours while only seeing two pages. 3 hours is 400 % higher than the second highest observation, which was 45 minutes. A plausible explanation is that the user got to K10 and then left the computer for several hours before coming back.

9.4.2. Sample Characteristics

During the 14-day experimental period, K10 had a total of 39,411 visits from 21,179 unique visitors. In the same time frame, 461 subjects clicked on one of the AdWords in Google.dk. This means the sample constitutes 1.18 % of the total population. Table 54 compares the sample and population on a number of behavioral and technical metrics.

		Sample	**K10 Total**
Behavior	Visits	496	42,025
	Unique Visitors	436 (87.90 %)	21,307 (50.70 %)
	Avg. Pageviews / Visit	4.95	4.64
	Avg. Time On Site	4 minutes 27 seconds	5 minutes 17 seconds
	Bounce Rate	39.11 %	49.83 %
	% New Visits	47.78 %	38.29 %
	New Thread	8.87 %	1.07 %
	New Post	2.22 %	2.43 %
	New Account	2.02 %	0.29 %
Browser	Internet Explorer	67.13 %	56.29 %
	Chrome	7.66 %	12.92 %
	Firefox	6.85 %	12.78 %
	Other	18.36 %	18.01 %
Platform	Mobile	12.97 %	12.93 %
Traffic Source	Search engines	100 %	78.91 %
	Referral	0 %	2.92%
	Direct	0 %	18.18 %

Table 54: Behavioral and technical metrics for the sample and overall k10 population

The most important insight from Table 54 is that the sample is reasonably representation of the total K10 population. When lacking demographic information, metrics such as browser and platform can be used as proxies for such information. The browser is also reasonably similar, although the sample is biased toward Microsoft Internet Explorer. According to a ComScore study, Internet Explorer users tend to be older and less affluent than Firefox users (Comscore 2009). The amount of mobile phone users is consistent across sample and population. This indicates that the sample and population have the same media consumption habits.

The sample is however quite different from the population on some metrics. The *visits to unique visitor ratio* is a lot higher in the sample than in the population. This is to be expected, as one of the aims of the experiment is to recruit new users. This data indicates that this has been done successfully.

The average time on site is more than one minute shorter in the sample. Although one would assume that time on site and bounce rate are correlated, only

40 % of the sample subjects bounced upon entering the site, compared to 50 % of the population. The sample also comprises more new visits than the population, 49 % compared to 39 %. An interesting aspect of this experiment is, thus, that it is an even mix of users who have and have not visited K10 before. The sample subjects created more new accounts than the population, 1.5 % compared to 0.3 %. This seems logical, as a larger proportion of the sample are new users.

The biggest difference noted between sample and population in the dataset is the amount of new threads created. 9 % of the sample subjects created a new thread, compared to only 1 % of the population. At the same time, the sample subjects are less likely to write posts in existing threads, as only 1.7 % of the experimental subjects add a post, compared to 2.4 % of the population.

The final notable difference is in traffic source. In the entire population, 78 % of the visitors enters K10 via search engines, 3 % are referred from another website, and 18 % type in K10.dk in the URL or use a bookmark from a previous visit. As all the experimental subjects are recruited through search engine advertisements, 100 % of the users come from search engines.

There are some differences between the subjects and the population. However, this is to be expected when the sample comprises many new users recruited via advertisement. These differences are not considered critical, and do not compromise the external validity of the experiment.

9.4.2.1. Sample Distribution

The reported results are based on two different datasets: CTR data from Google to K10, and behavior on K10.dk. The first dataset is much larger than the second, as it comprises people who clicked and did not click on the advertisements in Google. The second dataset only contains people who did click on the advertisement.

As the *first* part of this data is collected on Google.dk, the only available data is the data Google provides. Custom tracking is not possible on that domain. This data is only provided as an aggregate, which is problematic. However, as the data is binary, it is possible to reconstruct individual level data, as a user either clicks or does not click. The dataset is heavily zero-inflated, and an ANOVA test can therefore not be used. The Kolmogorov-Smirnov test of normality shows that $D(17,524) = 0.530$ $p < 0.000$, making it very unlikely that the data is normally distributed.

The *second* part of the data is from K10, and contains more details due to custom tracking. With a 40 % bounce rate, 40 % of the users never interact with the website, thereby creating a zero-inflated dataset.[245] As discussed further below, the bounce rate varies substantially between the four experimental groups. Bounce rates, thus, offer behavioral insights that make it inappropriate to re-move bouncing subjects from the data set. The Kolmogorov-Smirnov test shows that all six on-site dependent variables are highly significant, with $p < 0.000$, thereby *not* supporting the hypothesis that the data is normally distributed.

	Statistic	df	Sig.
Click	0.539	17524	0.000
Duration	0.298	460	0.000
Pages	0.261	460	0.000
New Account	0.537	460	0.000
New Post	0.535	460	0.000
New Thread	0.532	460	0.000
Bounce	0.397	460	0.000

Table 55: Kolmogorov-Smirnov test of normality for the seven dependent variables

In the first data set, *group 1* is n = 4,617, *group 2* is n = 4,516, *group 3* is n = 3,544, and *group 4* has n = 4,394. In the second data set, n = 181 for *group 1*, n = 84 for *group 2*, n = 73 for *group 3*, and N = 122 for *group 4*. This entails that all four groups have n > 30, and thus fulfill the central limit theorem.

---

[245] Please refer to Appendix 9 for histograms for the six variables.

Consequently, the ANOVA test can be used to analyze the data. ANOVA is an omnibus statistic, looking for at least one difference in the dataset. The test does not reveal between which groups the difference is significant. For pairwise post-hoc analysis, Tukey's HSD test is used.

9.4.3. Results

In the following sections, the effect of the intervention is measured on the seven dependent variables. Table 56 shows the raw data for each of the four groups across the seven dependent variables.

Group		CTR	Dura-tion	Pages	New Ac-count	New post	New Thread	Boun-ce	Con-tribu-tion
Group 1	Mean	0.039	0.002	3.94	0.02	0.03	0.12	0.44	0.149
	N	4802	181	181	181	181	181	181	181
	Std. Dev.	0.193	0.004	4.767	0.147	0.195	0.328	0.498	0.401
	Median		0	2	0	0	0	0	0
Group 2	Mean	0.019	0.004	4.92	0.05	0.01	0.08	0.29	0.095
	N	4602	84	84	84	84	84	84	84
	Std. Dev.	0.135	0.006	4.451	0.214	0.109	0.278	0.454	0.295
	Median		0	3	0	0	0	0	0
Group 3	Mean	0.019	0.004	6.4	0.01	0.03	0.1	0.32	0.123
	N	3614	73	73	73	73	73	73	73
	Std. Dev.	0.138	0.007	8.127	0.117	0.164	0.296	0.468	0.331
	Median		0	3	0	0	0	0	0
Group 4	Mean	0.025	0.002	4.34	0.01	0.02	0.07	0.43	0.090
	N	4506	122	122	122	122	122	122	122
	Std. Dev.	0.155	0.005	5.623	0.091	0.156	0.249	0.498	0.2876
	Median		0	2	0	0	0	0	0
Total	Mean	0.026	0.003	4.62	0.02	0.02	0.1	0.39	0.120
	N	1752	460	460	460	460	460	460	460
	Std. Dev.	0.159	0.005	5.649	0.146	0.167	0.294	0.489	0.3443
	Median		0	2	0	0	0	0	0

Table 56: Mean, median and standard deviations from the six on-site dependent variables

The results of the omnibus ANOVA test for the seven dependent variables are listed in Table 57. At a glance, there are significant differences on four dependent variables (CTR, duration, pageviews, and bounce rate) and no significant differences on the remaining three variables (new account, new posts, new threads).

Parameter		Sum of Squares	df	Mean Square	F	Sig.
CTR	Between Groups	1.166	3	0.389	15.505	0.000
	Within Groups	439.175	17520	0.025		
	Total	440.341	17523			
Duration	Between Groups	3152895.279	3	1050965.093	5.19	0.002
	Within Groups	92340810.07	456	202501.776		
	Total	95493705.35	459			
Pageviews	Between Groups	331.125	3	110.375	3.516	0.015
	Within Groups	14315.769	456	31.394		
	Total	14646.893	459			
New Account	Between Groups	0.083	3	0.028	1.307	0.272
	Within Groups	9.699	456	0.021		
	Total	9.783	459			
New Post	Between Groups	0.016	3	0.005	0.186	0.906
	Within Groups	12.721	456	0.028		
	Total	12.737	459			
New Thread	Between Groups	0.244	3	0.081	0.94	0.421
	Within Groups	39.547	456	0.087		
	Total	39.791	459			
Bounce rate	Between Groups	2.787	3	0.929	3.362	0.019
	Within Groups	126.004	456	0.276		
	Total	128.791	459			
Contribution	Between Groups	0.315	3	0.105	0.884	0.449
	Within Groups	54.109	456	0.119		
	Total	54.424	459			

Table 57: ANOVA test results of the seven dependent variables

9.4.3.1. Click Through Rate In Google

The first measurement is concerned with the CTR in Google on the four different ads. As Table 58 shows, the CTR varies quite a bit across the four groups. *Group 1* has the highest CTR of 3.85 %, followed by *group 4* with a CTR of 2.49 %. *Group 2* and *group 3* have a CTR of 1.87 % and 1.94 % respectively. The differences detected in the data are highly significant as F(1,439) = 15.05, p < 0.000

Group	Did Not Click	Did Click	Impressions	CTR	Avg. CPC	Cost (DKK)	Avg. position
Group 1	4617	185	4802	3.85 %	2.8	518.58	2.16
Group 2	4516	86	4602	1.87 %	2.81	241.91	2.28
Group 3	3544	70	3614	1.94 %	2.99	209.41	2.51
Group 4	4394	112	4506	2.49 %	2.99	335.15	2.37
Total	17071	453	17524	2.54 %	2.90	1305.05	2.33

Table 58: Results of AdWords campaign from the Google AdWords interface

Post hoc comparisons using the Tukey HSD test indicates that *group 1* (m = 0.038) and *group 2* (0.018) are significantly different, as p < 0.000 and Cohen's *d* = 0.12. Likewise the variance in *group 1* (m = 0.038) and *group 3* (0.019) is significant as p < 0.000 and Cohen's *d* = 0.11. The same is true for *group 1* (m = 0.038) and *group 4* (m = 0.025) as p < 0.000 and Cohen's *d* = 0.07. Thus, the *beneficial to self* advertisement has significantly higher CTR than all other groups. This suggests that advertisement that highlight a benefit to the subject make these more likely to click on the advertisement. This being said, the effect size is only small as Cohen's *d* range between 0.07 and 0.12.

The difference between *group 2* and *group 3* is not significant, p < 0.997 and Cohen's *d* = 0.01. Neither is the difference in the variance between *group 2* and *group 4* as p < 0.277 and Cohen's *d* = 0.05. Finally, the observed difference in CTR between *group 3* and *group 4* is not significant as p < 0.440 and Cohen's *d* = 0.04.

Contrary to the initial expectations, the *beneficial to self* group is outperforming the small *beneficial to small group* and *beneficial to large group* significantly. Although it was unexpected, it might be comprehensible. People searching for a keyword related to flex job might have questions they want to ask, rather than a

wish to share their knowledge. Search is mostly task-driven. Another surprising result is that none of the *beneficial to others* groups perform better than the control group (*group 4*). One possible explanation for this is that 47% of the subjects are unfamiliar with K10 at the time they see the advertisement, cf. Table 54.

9.4.3.2. New Posts

Only few subjects wrote a post on K10 throughout the experiment. The 460 subjects contributed 11 posts in total, entailing a mean of 0.02 posts and a median of 0 posts per subject. The minimum observation is 0 posts, and the maximum observation is 2 posts. *Group 1* and *group 3* perform slightly better with a mean of 0.3 posts, while *group 4* has 0.2 posts, and group 2 has 0.1 posts. The ANOVA test detects no significance in the data, as $F_{(0,13)} = 0.186$, $p < 0.906$.
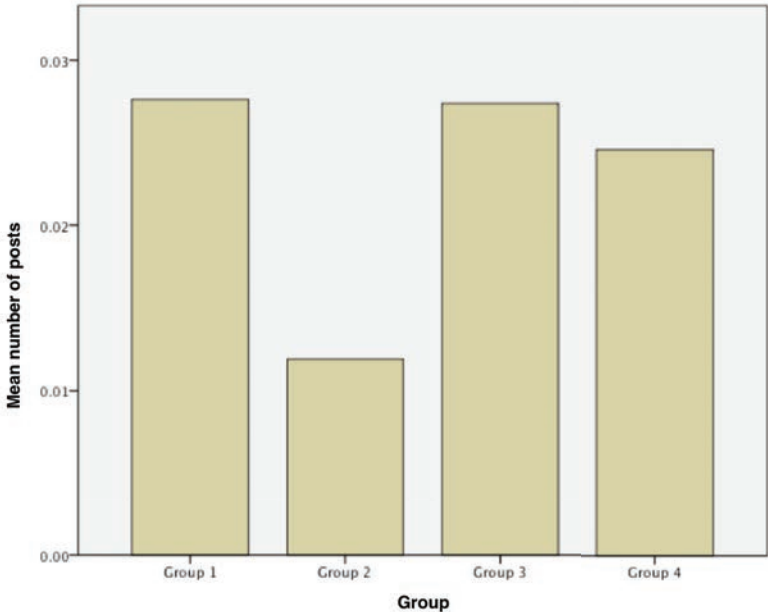
Figure 29: Mean number of posts in the four groups. The differences are non-significant

This experiment offers no statistical significant insights into the contribution rate of new users on the omnibus level. Consequently, it makes little sense to report the Tukey HSD post hoc test results and Cohen's *d* effect sizes, as no significant differences reside in the data set. In other words, the different levels of the independent variable – different priming of benefit – do not have any significant effect on the number of posts written.

9.4.3.3. New Threads

Contrary to the pre-experimental expectations, subjects were more likely to start new threads than to reply to existing ones. A total of 44 new threads were started by the subjects, with a mean of 0.10 thread per user and a median of 0 threads. *Group 1* performed best with 0.12 threads, followed by *group 3* with 0.10 threads, *group 2* with 0.08 new threads, and *group 4* with 0.07 new threads. The ANOVA test shows non-significant result with $F_{(0, 40)} = 0.940$, $p < 0.421$, and it can therefore not be concluded that these differences are caused by the experimental intervention. As there is no significance on the omnibus level, the pairwise post-hoc comparison and effect sizes are not reported.

Although the results are not significant, it is surprising that *group 1* subjects write more posts than *group 2*. It is, however, less surprising that *group 1* subjects start more threads than any of the other groups, as the ad copy highlights that the user can get help on early retirement pension issues. In sum, the different priming of benefit does not have any significant effect on the number of new threads the subjects start.

9.4.3.4. Contributions (Posts + Threads)

The lack of significance detected in the first two dependent variables might be the result of an insufficient amount of subjects. This explanation can partly be examined by computing a new variable comprising new threads and new posts. This variable is named *contribution*. Based on the two variables, the mean contribution is 0.12 per subject, with a minimum observation of 0 and a maximum observation of 3. *Group 1* has the highest mean with 0.19 contributions per sub-

ject, followed by *group 3* with 0.12 contributions. *Group 2* subjects made 0.10, and *group 4* has 0.09 contributions.

Aggregating these variables does not lead to any significant difference between the groups, as the ANOVA test shows $F_{(0, 54)} = 0.884$, $p < 0.449$. In sum, none of the four group constructs significantly influence the contribution level to K10. Thus, the experiment does not offer any significant insights into participatory behavior at K10.

As expected, it is hard to identify significant difference in explicit contributions. Therefore, additional variables were included in the measurement when the experiment was conducted.

9.4.3.5. New Account

More often than not, having a user account is a precondition for making a contribution to an online community. The number of account creations is also measured in this experiment. Of the 460 subjects, 10 users created an account. The mean account creation per user is 0.02. As this variable is a binary variable, the minimum observation is 0 accounts, and the maximum observation 1 account. The median is 0 accounts.
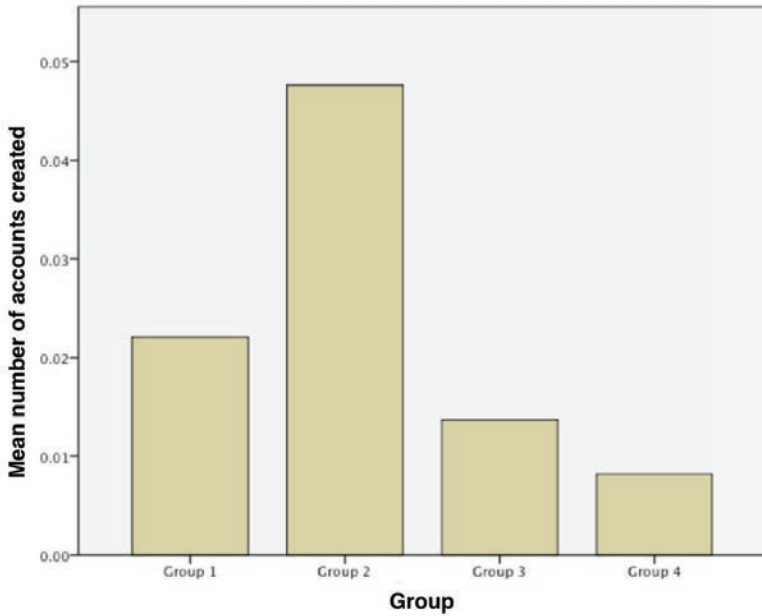
On this dependent variable, *group 2* outperformed the other groups. Group 2 created 0.05 accounts per user, *group 1* created 0.02 accounts, and *group 3* and *group 4* each has 0.01 account creation per user. According to the ANOVA test, these differences are not statistically significant as $F_{(0,10)} = 1.307$, $p < 0.272$. As the omnibus is non-significant, the Tukey HSD post-hoc comparisons are not reported. In other words, the different levels of the independent variable do not give any significant insights.

9.4.3.6. Pageviews

For all experimental subjects, the mean number of pageviews is 4.62 pages per subject, while the median is 2. The minimum observation is 1 page, and the maximum 36. Contrary to earlier reported variables, *group 1* does not perform better in terms of pageviews. *Group 1* subjects see 3.94 pages on average, *group 2* sees 4.92 pages, *group 3* visits 6.40 pages, and *group 4* views 4.34 pages. The ANOVA test finds this difference to be significant with $F_{(3)} = 3.516$, $p < 0.015$.
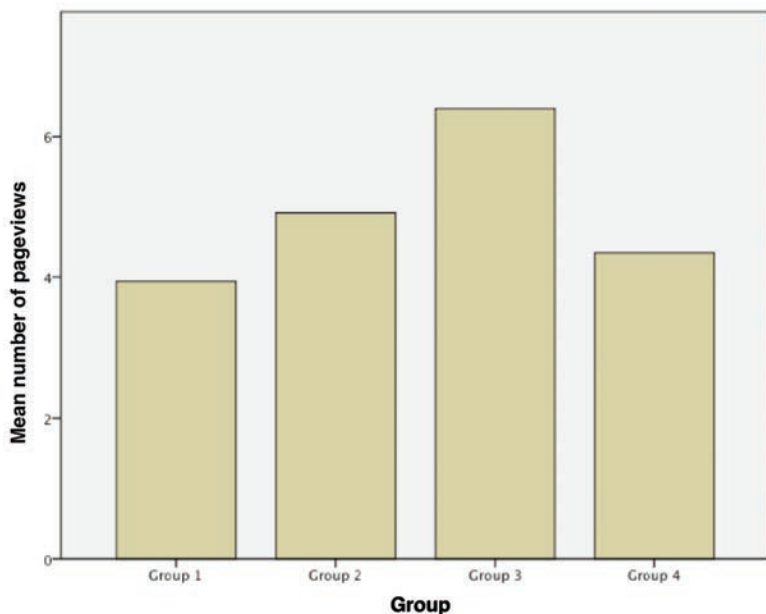
The pairwise Tukey HSD tests show which groups differ significantly. The only significant difference is between *group 1* (m = 3.94) and *group 3* (m = 6.4), as $p < 0.009$ and Cohen's $d = 0.11$. The difference between *group 3* (m = 6.4) and *group 4* (m = 4.34) is close to significant, but not quite as $p < 0.065$ and Cohen's $d = 0.06$.

It is particularly members of the large-group construct, *group 3*, who engage in the content on K10. Priming the ad copy so it highlights *benefit to other* thus seems to lead to significantly more engagement in terms of pageviews. The reported Cohen's d effect size indicates that the difference is only small. In other words, subjects in the *beneficial to large group* group construct visit significantly more pages than subjects in the *beneficial to self* construct.

9.4.3.7. Time on Site

Pageviews and time on site are expected to show similar trends, as number of pageviews logically needs to be correlated with time on site. It takes more time to view several pages than few pages.

The mean time on site for the subjects is 3 minutes and 55 seconds, and the median is 40 seconds. The shortest session is 0 seconds, and the longest 44 minutes and 57 seconds. *Group 1* subjects spend an average of 2 minutes and 42 seconds on K10. *Group 2* and *group 3* subjects spend more time; 5 minutes and 32 seconds, and 5 minutes and 54 seconds, respectively. *Group 4* subjects use K10 for 3 minutes and 55 seconds on average. This difference in time on site is highly significant, as the ANOVA test shows $F(3152, 9234) = 5,190$, $p < 0.002$.
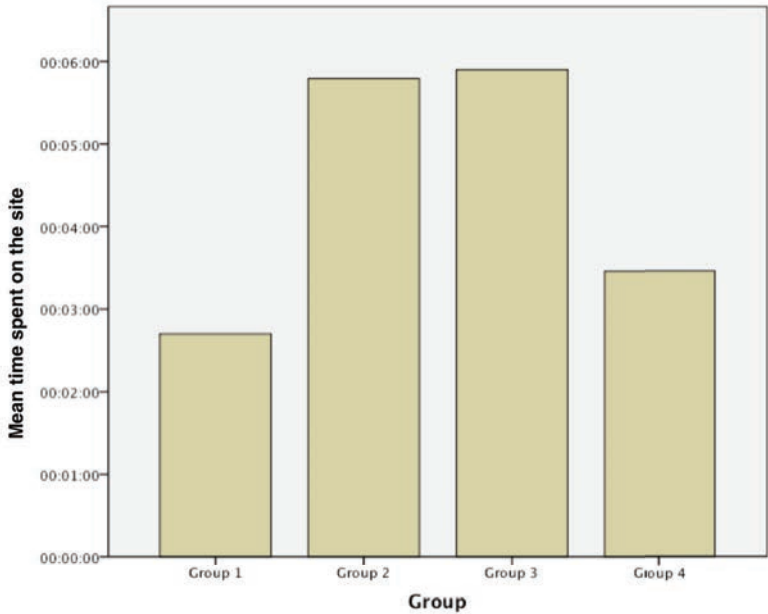
The observed difference between *group 1* (m = 3 minutes and 55 minutes) and *group 2* (5 minutes and 32 seconds) is significant as $p < 0.01$ and Cohen's *d* =

337

0.41. The same is true for the *group 1* and *group 3* (m = 5 minutes and 54 se-
conds) difference: $p < 0.012$ and Cohen's *d* = 0.37. The difference between *group
1* and *group 4* (m = 3 minutes and 55 seconds) is not significant as $p < 0.826$ and
Cohen's *d* = 0.04. In other words, the two *beneficial to others* groups are scoring
significantly higher in terms of time on site than the *beneficial to self* group. The
effect sizes in both groups are close to medium. There is no difference between
beneficial to self group and *no benefit* control group.

The difference between the two *beneficial to others* (*group 2* and *group 3*) is not
significant, $p < 1.00$ and Cohen's *d* = 0. There is no significant difference between
group 2 and *group 4,* $p < 0.125$ and Cohen's *d* = 0.24, or between *group 3* or *group
4* as $p < 0.123$ and Cohen's *d* = 0.22.

The engagement metrics align well with the initial research hypotheses. The sub-
jects exposed to the two *beneficial to other* groups spend much more time on K10
than subjects assigned to the individual group. No difference is measure between
any of the three experimental conditions and the control condition.

9.4.3.8. Bounce Rate

The last dependent variable measured in this experiment is the bounce rate. The
mean bounce rate across all groups is 39 %, with a median bounce rate of 0. As
this variable is a binary variable, the minimum observation is 0 bounces, and the
maximum observation 1 bounce. The mean bounce varies quite a bit between the
different groups. *Group 1* and *group 4* have higher bounce rates of 44 % and 43
% respectively. *Group 2* only has a 29 % bounce rate, and *group 3* has a 32 %
bounce rate. According to the ANOVA test, these differences are significant as
$F(3, 126) = 3.362$, $p < 0.019$.

Tukey HSD post-hoc test shows that the observed difference between *group 1*
(44 %) and *group 2* (29 %) is non-significant as $p < 0.111$ and Cohen's *d* = 0.33.
The same is true for the *group 1* (44 %) and *group 3* (32 %) difference: $p < 0.303$
and Cohen's *d* = 0.25. The difference between *group 1* and *group 4* (48 %) is not

significant as p < 0.906 and Cohen's *d* = 0.07. Neither is the bounce rate between *group 2* and *group 3* subjects as p < 0.985 and Cohen's *d* = 0.03.

The only significant difference in the data set is between *group 2* (m = 0.29) and *group 4* (m = 0.50) where p < 0.041 and Cohen's *d* = 0.34. This indicates that the effect size is small-medium. The bounce rate between *group 3* and *group 4* subjects do not differ significantly as p < 0.134 and Cohen's *d* = 0.14.
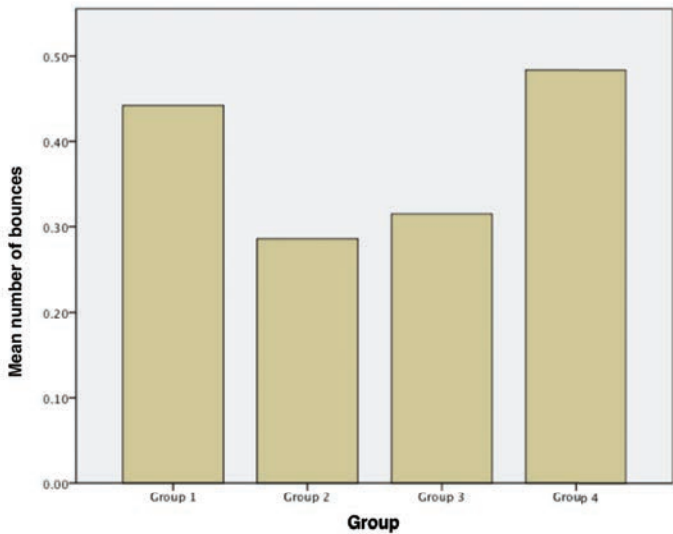
Figure 34: Mean bounce rate in each group. The differences are significant

In sum, the subjects exposed to the *no benefit group* tend to leave K10 immediately at a significantly higher rate than subjects exposed to the *beneficial to small group* variation.

9.4.4. Summary of Results

This experiment does not provide consistent insights across all dependent variables. The experiment only yields significant results on four parameters: clicks, pageviews, time on site, and bounce rate. The evidence does not give significant insights into the primary dependent variable, contributions. Being aware that a

contribution benefits others does not lead to significantly more contributions. Referencing the strong tradition of social identity theory, it is surprising that no significant effect is found across contributions. This might indicate a suboptimal theory operationalization. This is further described below.

The results are not consistent across all parameters. In terms of CTR from Google to K10, *group 1* is significantly better than the three other groups, although the effect size is only small. In terms of pageviews, *group 3* is performing significantly better than *group 1*, again with a small effect size. With respect to time on site, *group 2* and *group 3* are both significantly better than *group 1,* with a medium effect size. In terms of bounce rate, *group 3* performs significantly better than *group 4*, with a small effect size.

The null-hypothesis is partly rejected, as some significance is found in terms of CTR, pageviews, time on site and bounce rate. It is only partially rejected, as no variance is found with respect to posts, threads, and new accounts.

	Hypothesis	Evaluation
H_0	The group constructs have no effect in terms of engagement and contribution behavior.	Partially rejected
H_a	Subjects assigned to the *beneficial to others* groups will engage and contribute more than subjects assigned to *benefit self* and control group.	Partially verified
H_b	Subjects assigned to the *beneficial to small* group will engage and contribute more than subjects assigned to *beneficial to large* group.	Rejected

Table 59: Evaluation of the hypotheses in experiment 4

The H_a hypothesis is partially verified as the two *beneficial to others* groups were the only metrics on K10 that were significantly better than the *beneficial to self* group and control group. However, the *beneficial to self* outperformed the *beneficial to other* groups on CTR. In addition, there were no significant differences on three parameters. Hence the partial verification.The H_b hypothesis is rejected as there are no significant differences between the *beneficial to small* group and the *beneficial to large* group.

9.5. Limitations

In this section, three types limitations are discussed: internal validity, external validity, and construct validity.

9.5.1. Internal Validity

One limitation of this experiment is that it only measures behavior in the session immediately following the click from Google to K10. This study does *not* take future contributions into account. One possibility is that higher engagement metrics in the first K10 session leads to more contributions in the future. If this is the case, the *group 2* and *group 3* subjects will outperform *group 1* and *group 4* subjects in the long term. This is, however, merely a speculation.

One of the most interesting insights is that higher CTR in Google Ads not necessarily is related to the post-click behavior. It might be much more efficient to write advertisements that prime the desired behavior, than advertisements that drive a lot of clicks. However, getting too few clicks is also problematic as advertisement platforms will then show the ads less frequently, cf. the pilot study. A study measuring the behavior over a longer time span and multiple visitor sessions can help answer these questions.

A second issue in terms of internal validity is that the Google AdWords platform only gives few insights into the characteristics of subjects from the four groups. Although the subjects searched for the same keywords in order to see the advertisements, the different advertisements might attract diverse types of subjects rather than change the subjects through priming. Existing differences between the subjects from the three groups might thus be the real reason behind the different behavior. This does not, however, change the outcome of the experiment: that certain wording and priming leads to different outcomes.

9.5.2. External Validity

Google AdWords does not give many details about the sample subjects. It is therefore difficult to give any precise estimates about which, if any, populations

the sample can be generalized to. At least two different distinctions can be considered: those who did click vs. did not click on the advertisement on Google; and those who did search vs, those who did not search. Both distinctions can be seen as potential self-selection sampling issues, making it problematic to extrapolate the results from the experiments. The external validity of the experiments is further discussed in Section 10.3.1.

9.5.2.1. Did Click / Did Not Click

The first possible population that the experiment potentially can be extrapolated to is the population comprising all the people who searched for keywords relevant to flex job during the experimental period. As the experimental data shows, only 2.90 % of those exposed to an advertisement clicked on it. This means 97.10 % did not.

Although the experiment does not provide any data on this matter, systematic differences may exist between the clickers and non-clickers. For example, some searchers might be looking for more generic information about flex job, and not be interested in an online community. And there is no way to reach the people who did not click on the advertisements. Consequently, it is very difficult to know to which degrees the findings extend to all people searching for flex job.

9.5.2.2. Did Search / Did Not Search

The second distinction relevant to the external validity of the experiment is between people who did and did not search for keywords containing flex pension keywords. Lewis and Reiley (2010) question the external validity of using advertisements in search engines for conducting experiments:

> *"The population of people who sees a particular ad may be very different from the population who does not see an ad. For example, those people who see an ad for eTrade on the page of Google search results for the phrase 'online brokerage' are a very different*

population from those who do not see that ad (because they did not search for that phrase)."

<div align="right">*(Lewis & Reiley 2010: 1)*</div>

The point here is that people conducting a given search are different from those who do not. Furthermore, it is complicated to delimitate the "did not search" population, as this population comprises the entire country. This means no claims of validity in such a broad population are made.

9.5.3. Construct Validity

Significant differences are detected between all group combinations except *group 2* and *group 3*. This might indicate one of two things. On one hand it might indicate that the details of the group construct do not matter. Group similarity is of little concern to the subjects in this experiment. There might be some truth to this, as all flex job recipients implicitly are in a group comprising flex job recipients. The "like minded" construct and "others in Denmark" might in practice turn out to bear the same meaning.

On the other hand, the wording of the two groups might not be of sufficient quality to prime the different group constructs. This entails that the construct of this part of the theory is inadequate and causes low construct validity between *group 2* and *group 3*. Conducting further experiments with different wording as well as follow-up interviews can help identify the cause behind the lack of difference between the groups.

This leads to a more general critique of the construct validity: the research design does not give any indications of the degree to which (if any) the subjects experience group identity. This is a result of the very limited sample characteristics provided by Google AdWords. As shown in Table 54 above, more than 50% of the subjects have visited K10 before they clicked on a Google advertisement link. This means they have some familiarity with the community, which increases the likelihood of them feeling a shared identity. It is however a limitation of the

study that this identity is not measured. In order to make general claims about how social identity influences participation, it is necessary to measure to which degree the subjects share this identity. There are numerous scales that can help illuminate this issue. Again, such general theory falsification is outside the scope of this dissertation, but it is an interesting part for further research.

9.6. Implications for Open Government Community Design

The first implication for design is a methodological point: online advertisement platforms such as Google AdWords can be used to examine hypotheses efficiently. Advertisement platforms can be used to both recruit and randomly assign subjects. Thus, the major difference between this proxy and the ones used above (email, message, and survey) is that no prior users are needed. With a cost per click of less than DKK 3, the subject recruitment is very cheap. The major caveat of this proxy is that few details are known about the subjects. This makes it harder to detect nuisance variables.

Another implication for design is that one design pattern might not yield consistent effects across all variables. This means it is important to examine several dependent variables insofar they are critical to the community. And be aware that it might be necessary to make trade-offs between variables.

Another valuable implication for design is that words matter. There is always a first impression, whether or not it is on the community or somewhere else on the internet. This first impression can help prime for contribution, making it a mean to help overcome the participation challenge. This experiment indicates that social psychology can lead to self-fulfilling prophecies. If an online community is presented as something a user is supposed to benefit from, but not contribute to, she is likely to do just that. On the other hand, if a design expresses an assumption of pro-social individuals that all contribute, it will facilitate participation. At the very least such wording will help attract users who are interested in actively contributing.

The broader implications for open government, e.g. participation patterns, opportunity costs, and designing for niche audiences, are further discussed in section 10.2.

9.7. Chapter Summary

There are two strategies to building and sustaining an online community: motivation of existing members and outreach to new users. Whereas the first three experiments focus on the former, the fourth experiment focuses on outreach.

Through the use of advertisements via Google, it is concluded that people searching for flex job related keywords click significantly more on advertisements promising benefit to self rather than benefit to others. Interestingly enough, this higher click through rate does not translate into more engagement after entering K10, quite contrary. The subjects primed with a pro-social advertisement spend more time on K10 and see more pages than subjects primed with the individualistic oriented advertisement. Furthermore, the subjects primed with pro-social ad copies have significantly lower bounce rate than the control group. Although it was aimed to change participation in K10 by leveraging social identity theory, this chapter's main contribution is that priming for group constructs can affect engagement, but not participation.

Although several dependent variables were significantly changed by the individualistic vs. pro-social intervention, no significant difference was detected in terms of new account creations, posts, or threads.

Part 4

Discussion and Conclusions

Chapter 10: Discussion

10.1. Introduction

The four experiments reported in the previous chapters show that it is possible to alter and improve the participation in open government communities. As it is often the case in the social sciences, the results are more ambiguous than could be preferred. Only some results are significant and the effect sizes vary a lot. The aim of this chapter is to discuss how to best utilize the findings and contributions from this dissertation to improve open government communities' practice as well as further research.

The first part of the discussion is concerned with the implications for open government in practice. In this part of the discussion, focus is on altered participation patterns, opportunity costs, designing for niche audience, emotional reactions, and the use of participation goals.

Subsequently, the focus shifts to implications for research. It becomes clear that the biggest concerns are limited external and temporal validity. It is hard to know to which other open government communities and for how long the findings of this dissertation extend. The remainder of the chapter is dedicated to overcoming the challenge of reduced external and temporal validity by setting up a framework for lean experimentation. This framework is the main contribution of this chapter as it serves as a scalable and efficient solution to improving participation in open government communities in the future.

Through the lean experimentation framework it is possible to conduct experiments quickly and efficiently. Such experimentation makes it possible to overcome some of the external and temporal validity concerns and challenges through increased replication. It follows that the contribution of this dissertation extends beyond a set of findings to a methodological framework. Equally important, the lean experimentation framework helps answer the research question of this dissertation more thoroughly in the future.

Although this chapter maintains a distinction between research and practice, it is concluded with a discussion of how lean experimentation might be beneficial to both research and practice.

10.2. Implications for Open Government in Practice

The first part of the discussion focuses on the implications for open government communities in practice. The research question asked in this dissertation is: *Can social psychology theory help increase participation in open government communities?* The results from three of the four experiments suggest that it is possible to use social psychology theory to affect behavior in open government communities through design changes. Despite these encouraging findings, the design ideas tested here do not solve the participation problem in open government communities entirely. The results are too ambiguous and the effect sizes too low to solve this problem once and for all. But the research does facilitate the process.

The four experiments conducted and reported in this dissertation have five implications for practice: (1) participation patterns can be altered; (2) it might entail opportunity costs and suboptimal allocation of resources; it is important to consider (3) how to design for niche audiences; (4) how to cope with negative and positive reactions; and (5) whether or not to consider using participation goals.

10.2.1. How Participation Patterns Can Be Altered

Three of the four experiments show that it is possible to change participation through social psychology-based interventions. Providing users with social comparison information – or even just non-comparative participatory information – regarding other users within their community will make the otherwise low-contributing users contribute significantly more. Having users make a commitment about future contributions increase participation among those setting a high goal. And thanking users for their contributions make the users contribute more as well.

Altered participation is not necessarily the same as increased participation. But a more evenly distributed participation pattern might be highly desirable from an open government perspective. As described in Chapter 2, there is a debate over the participatory goal of open government. Some are concerned with the process and deliberative aspects, while others are concerned with the outcome and collaborative aspects. Experiment 1 shows that social comparison information can create convergence toward the mean: the most active users will do less, and the less active users will do more. This development is welcomed from a deliberative democratic perspective. The citizens and other stakeholders, who usually have a lot of voice and influence, will have less while the less vocal citizens will have more.

However, the experiment does not give any indications of how these altered participation patterns affect the contribution quality. It is a possibility that contribution quality comes at the expense of participation equality. If this is the case, proponents of the collaborative democracy perspective will not welcome the more egalitarian contribution pattern as it reduces the outcome quality. Further research needs to be conducted in this area. The experiments do not examine which of the two aspects of democracy is the right one to pursue. But it does show that different designs of open government communities can support either of the goals. The implication of this finding is that open government policy makers need to explicitly define their primary participation goals. Without making such explicit aims, it is impossible to design for it.

10.2.1.1. Social Engineering

Deliberately changing participation patterns in order to attain a goal is sometimes called social engineering.[246] To some, social engineering has a negative connotation. When governments engage in social engineering, critics are accusing the government of being authoritarian and paternalistic. But as explained in Chapter 2, no design is neutral. An online community design will always encourage certain actions while discouraging others.

---

[246] See for example (Richtel and Sack 2012).

This poses a risk for government owned communities, as they might be accused for social engineering when optimizing their community for certain goals. This is especially likely in libertarian countries, like the United States, where right-wing parties tend to accuse governments of social engineering (Cooper 2012).[247]

10.2.2. Opportunity Costs and Suboptimal Allocation of Resources

Citizens generally have limited time to participate in open government communities. In the short term, it might be a plus sum game as the internet is still evolving and more people are spending their time online. But in the long run it will be a zero sum game between online communities. Participation in one open government community entails opportunity costs among other online communities.

As it is possible to increase and decrease participation in open government communities through design, it becomes appropriate to make overarching choices about where citizen participation is most necessary. As noted by Schelling (1978), individual rationality does not always lead to collective rationality: Although, it is individually rational for each open government community to maximize participation in order to succeed, this kind of optimization might lead to suboptimal allocation on macro level. If some open government communities start implementing social psychology inspired design patterns that increase participation, they might cannibalize the participation from those open government communities that do not.

Open government communities optimizing for participation might not just cannibalize other online communities, but also offline civic participation. If citizens start spending more time on open government participation, there is a risk that less time will be spent on offline volunteering. It is important to consider if such changes are desirable. As described in Chapter 2, some online participation is

---

[247] For example, the Republican Party raised this issue in the 2012 election campaign (Cooper 2012).

characterized as *slacktivism*.[248] The risk is that high-value and high-impact of-fline participation is substituted with online participation with less value and impact.

Open government policy makers ought to treat participation as a scarce resource and define to which areas they should be allocated. In sum, when opportunity costs are included in the calculation, it might not always be desirable to increase participation in any given open government community, as it takes finite re-sources away from other online or offline areas.

10.2.3. Designing for Niche Audiences

The big social networks design for the average, mainstream citizen. Facebook and Twitter do not seem to pay particular attention to special needs groups. Nor do they need to in order to succeed in terms of growth, profitability, and sustain-ability. However, governments might not have this luxury. One obvious task for governments is to supply goods when the market fails certain groups. This can also be the case in terms of online communities, if the market does not address certain niche groups. An example is Danish Regions' health networks.[249] Danish Regions have created a number of communities for people with relatively rare health problems such as habitual abortion and modic changes. As these diseases are relative rare, people suffering from them do not have dedicated interest groups to provide open government services.

The experiments reported in chapter 6-9 show that designing for niche audienc-es, such as socioeconomically disadvantaged, can be challenging, as this group might react differently than other groups to a particular design pattern. Social comparison theory shows that people under stress react differently to social comparisons than people, who do not experience a similar level of distress. A

---

[248] Slacktivism refers to "activities that are easily performed, but they are considered more effective in making the participants feel good about themselves than to achieve the stated political goals." (Christensen 2011: 5).

[249] http://patientnetvaerk.sundhed.dk

good idea in a mainstream online community might be a bad idea in a niche community, or vice versa. Experiment 1 illustrates this difference, as many users were upset with the interventions. This is likely due to their unique conditions, which make them more likely to react negatively to social comparisons. This is, in other words, a matter of external validity. This is discussed thoroughly in section 10.3.1.

The point is that open government communities often need to be designed for special audiences. Solving the participation problem in these settings might require different tools than in mainstream online communities. Unfortunately, this makes it more complicated to design successful open government communities than non-governmental communities, due to the more limited tool set.

10.2.4. Positive and Negative Reactions

Some design practices are much more likely to make people happy while attaining the goal of increased participation.

10.2.4.1. Negative Reactions and Unintended Consequences

Design ideas leveraging social psychology theories can upset some users and have other unintended consequences. Experiment 1 demonstrates both.[250] Quite a large number of users are expressing their dissatisfaction with the emails containing social comparison information. Furthermore, although the effect was non-significant, Experiment 1 does shows an overall decrease in participation after sending out participatory information to two-thirds of all K10 users. The above-mean contributors participate less after the intervention. This indicates a risk of unintended consequences in social psychological derived design.

---

[250] The unintended consequences and negative reactions discovered in Experiment 1 might be the exception rather than the rule, as no negative reactions were detected in the other three experiments. This being said, the four experiments conducted in this dissertation do not offer conclusive evidence regarding the magnitude of unintended consequences. More experiments have to be conducted to identify this.

One effective way to discover such negative effects and their impact, is to conduct experiments. It is very likely that the negative reactions would have been even greater if the social comparison information had been implemented in full-scale in the community; e.g. through a leaderboard. In hindsight, it would have been advantageous to run a pilot study on a small sample of the K10 users to see how they reacted to the information before sending it out to thousands of users. Thus, experiments can be a way of detecting discontent before a full-scale implementation is initiated. However, it must be a field experiment, as laboratory experiments might fail to notice such effects.[251]

Many successful online communities such as Facebook, Instagram, and Google+ regularly make design changes that upset millions of users. Experiment 1 shows that open government communities have the same risk. This might be due to a poor match between theory and case, but it is often difficult to foresee these negative reactions. Open government communities, especially those owned by governments, probably need to be more careful with upsetting users than privately owned communities. These reasons are the special requirements for legitimacy in government, and the citizens' lack of option of exiting from government funded services.

On a higher level, this entails that not all social psychology theories are appropriate for online communities in general, and open government communities in particular. It also means that not all design practices are appropriate in open government communities. If social comparisons are considered to be inappropriate in open government communities - which is a fair assessment - then some often-used design practices should be avoided as well. This entails, for example, leaderboards. These play a significant role in successful communities such as Foursquare and Amazon. Having a more limited tool set when developing communities in the governmental sphere makes it harder to succeed.

---

[251] This seems to be the case in Chen et al. (2010), who did not report any upset users in a similar study conducted in an online laboratory.

10.2.4.2. Positive Reactions

Although some social psychology inspired design ideas can upset users, other design ideas leveraging social psychology theories can make users happy. Experiment 3 serves as an example, due to the positive reactions that the appreciative emails generated. It might be efficient for government to start thanking people for their contribution. Although the research conducted here offers few insights into the external validity of this finding, it is possible that thanking citizens can foster many kinds of online and offline citizen contributions.

Interestingly enough, there is a thin line between the different design patterns here. Being thanked privately makes users happy. However, if this knowledge becomes public, it might have a negative impact among other users. For if it is public knowledge that some users are being thanked, it might serve as social comparative knowledge, and thus cause the negative reactions detected in experiment 1.

10.2.5. The Use of Goals

As experiment 2 demonstrates, setting up goals can be a good way to get users to contribute more. Experiment 2 shows that citizens and other stakeholders can be asked to make a commitment about future participation. By anchoring the commitment possibilities higher than the status quo, it is possible to increase participation. This type of intervention can be applied selectively to get certain relatively passive segments of the citizenry to participate more.

This is a very direct way for governments to ask citizens to get involved, and it will be explicit that government needs and relies on the citizens. It is quite rare that government set goals for someone other than itself. But goals can possibly be applied many places: governments and citizens can set goals for participation in online communities, set economic goals for society, or set goals for electoral turnout.

Using goals comes with one risk though. And as demonstrated by Drenner, Sen, and Terveen (2008), goals can serve as barriers of entry for some users even though it increases the overall participation. This mean the increased participation comes at the expense of more equal participation, as many users will fail to overcome the barrier of entry, or be out off by its very existence. Such skewed participation is undesirable the deliberative democracy perspective.

10.3. Implications for Information System Research

As described above, the external and temporal validity concerns are significant obstacles when seeking to develop open government research further. These challenges are discussed in the following sections.

10.3.1. Validity Concerns

As described in Section 5.9.1, six types of validity are relevant for this dissertation: internal validity, external validity, ecological validity, population validity, temporal validity, and construct validity. The validity of each experiment is addressed in Chapter 6-9. Based on these assessments, it is clear that especially external validity and temporal validity have implications for further open government research.

10.3.1.1. External Validity

On a high level, K10 is an interesting case because the users differ from the average internet user. But it does make the external validity more limited. Two different types of external validity issues are relevant to assess here. The *first* one is to which extend the results extrapolate to all Danish citizens on either early retirement pension or the flexjob program. As described in section 3.3.3.2, the K10 community (n = 9,000) seems to have certain biases compared to the entire group of early retirement pensioners and flexjobbers (N = 296,241). Compared to all flexjobbers and early retirement pensioners, K10 has an overrepresentation of the 25-44 year-olds and; and underrepresentation of 55-65 year olds. Therefore, the results of this analysis cannot necessarily be generalized to the entire population of flexjobbers and early retirement pensioners.

The *second* and more important type of external validity is to which degree the findings from the four experiments extrapolate to other open government communities. This dissertation only offers one case: K10. Although the insights would have been fruitful, this dissertation does not aim to test external validity of one theory or design pattern across several online communities. Again, making such comparisons is not necessarily the purpose of individual research but ought to be created in the overall research corpus.

Thus, the experiments do not offer any conclusive insights into the external validity of research into online communities. Online communities are very diverse as they exist across sectors with different purposes. This diversity makes extrapolation problematic on several dimensions. The external validity between governmental and non-governmental communities is assessed in Chapter 3, where it is discussed in terms of recruitment/retention balance, contribution type, cost of contribution, legal conditions, and organizational culture. In addition to these, it is relevant to discuss external validity in terms of cultural differences and demographics.

10.3.1.1.1. Cultural Differences: East/West

Social psychologist (e.g. Heine 2010) and HCI researchers (e.g. Clemmensen 2006) are concerned with cultural differences. As argued by Clemmensen (2006), cultural differences ought to be considered when using social psychology in HCI and open government contexts.

The most important cultural distinction in open government community research is probably the one between East and West. The social psychology literature acknowledge the existence of important cultural differences but still lack data to accurately map them (Heine and Norenzayan 2006). Current research indicates that notable cultural differences include perception of fairness, attention, and the need for high self-esteem (Heine 2010). These cultural differences might impact users' motivation to participate.

According to Macintosh et al. (2009), a technology is never a stable construction, and cannot be classified as participatory or exclusive. They argue that "technologies are neither inherently participatory, nor exclusive, but depend upon cultural practices and policy contestations" (Macintosh et al. 2009 6). Thus, a system can be participatory in one context, but not in another. Examining such differences by replicating studies across cultures is necessary over time in order to examine and improve the cross-cultural validity.

The point is that neither social psychology nor open government literature is culturally neutral. Although open government is a global phenomenon, it is currently unknown to which extend findings from one culture extrapolates to the other. Citizens and other stakeholders might very well participate for different reasons. Investigating such differences lie outside the scope of this dissertation. At the same time it is acknowledged as an important limitation of the external validity of this research, as the findings are limited to Western cultures.

10.3.1.1.2. Socioeconomically Advantaged and Disadvantaged

As described in section 3.3.3, the K10 demographics share a number of characteristics that are different from the average Danish citizen. The k10 users are socioeconomically disadvantaged compared to the Danish citizenry at large. It is likely that socioeconomically advantaged and disadvantaged people will react differently to some social psychology inspired design patterns. This weakens the external validity to open government communities comprising more advantaged members.

10.3.1.2. Temporal Validity

Online communities are never static. Size, demographics, group interactions, and context change over time and often rapidly. It is virtually impossible to make precise predictions about how an open government community develops over the coming years. Some users will leave while others join. Relevant policies about flex job and early retirement pension might change, thereby making K10 obsolete or increasingly relevant. New competitors might arise and others dis-

solve. New technologies will be invented in other communities, setting expectations about how K10 ought to function. The point is that the temporal stability is low compared to many other social science realms.

The internet arguably moves faster than science is used to, increasing the need for frequent replication. At this point it is unknown after how frequently an experiment needs to be replicated in order to find out if the findings still apply (Davenport 2009). There is thus currently no knowledge about how long research findings, like those from this dissertation, are valid. One indicator might be that online communities are changing design and features often, entailing that the lifespan of features and design patterns is relatively short.[252]

As described in Chapter 4, the temporal stability is likely to be greater when focusing on humans rather than technology. By relying on social psychology theory it is assumed that the temporal stability will be greater as the research and design focuses on fundamental human behavior.

It is concerning that the temporal validity of this research strand is largely unknown. It is impossible for online community developers and managers to know how old research they can rely on, and it is equally hard to know when it is the right time to replicate previous studies. As noted by Varian (2007):

> *"The ability to experiment easily is a critical factor for Web-based applications. The online world is never static. There is a constant flow of new users, new products and new technologies. Being able to figure out quickly what works and what doesn't can mean the difference between survival and extinction".*
>
> (Varian 2007 quoted in Kohavi et al. 2009: 1777)

---

[252] For example, Facebook regularly makes radical design changes (Quora 2013).

10.3.1.3. Why External and Temporal Validity Matters

The four experiments reported in this dissertation yield insights into which design features lead to changed behaviors at K10. Even when assuming these findings are objectively true and 100 % accurate, it would still be risky to implement them into other communities without further experimentation. A given design pattern may have the desired outcome in community X, but not when it is implemented in community Y. Problems often arise when a design pattern is copied without a proper understanding of *why* it works. In information science, this potential pitfall is known as cargo cult programming. The point is that copying a design pattern can go wrong when it is not tested on users before implementation.

An alternative to making assumptions about external validity is to test the feature or function in the new context before implementing it (Carroll and Kellogg 1989). It is therefore appropriate to test a theory, hypothesis, or assumption in a specific setting before implementing it. The same applies to temporal validity. Although implementing a feature in T_0 increases the performance on one or more important dependent variables, the same feature might have a different effect in T_1 or T_2. This means a feature needs to be evaluated regularly. An additional benefit to such thorough testing and experimentation is increased knowledge about the boundaries, effect sizes, and validity of the underlying theory, Creating a framework that enables quick and low-cost replications will be immensely helpful. Such framework is proposed in section 10.4.2.

10.3.2. Experimentation in Online Settings

Experimentation in open government research is rare; and there is little evidence that suggests that the methodology is often used in open government practice. This is not particularly surprising as it is easier and cheaper to refrain from conducting experiments; at least in the short term. In the long run, resources and democratic potential are wasted because suboptimal products are developed. In the final part of this dissertation I seek to use the experiences and

findings from the dissertation work to propose a framework that can further experimentation in research and practice.

10.3.2.1. Shared Characteristics of the Four Experiments

As previously noted, experimentation is only one of many research options available in open government research. Experimental methodology is not superior to other methodologies and cannot help answer *all* types of research questions (McGrath 1995). Like all other methodologies, experiments come with advantages and disadvantages. Seen in the light of the omnipresent participation challenge to open government communities, one question that needs to be addressed is how a different design changes participation. Since such questions are concerned with the causality of how theoretical constructs affect behavior, experimentation is a good methodology to employ.

Four experiments are conducted and reported in this dissertation. These experiments share four characteristics:[253] *First*, all four experiments fulfill the requirements of being true experiments, as the subjects have an equal chance of being assigned to each group. This optimizes the chance of discovering valid causal findings, as the subjects are randomly assigned to the groups.

Secondly, the experiments use constructs derived from theory. This helps generate better design ideas and potentially increases the external validity, as the theories are concerned with human behavior rather than specific technologies.

Thirdly, the experiments are conducted in the field. This entails that the experiments are conducted in a real open government community on real users. This is

---

[253] These shared characteristics are partly a result of deliberate choices, accounted for in Chapter 5, and partly of the conditions under which the dissertation research is conducted. The explicit choices are primarily concerned with the experimental design and the theory-driven hypothesis testing design. As this dissertation is not part of a larger research project, the resources are scarce. This has made it necessary to make the experiments relatively simple, quick, and inexpensive.

advantageous as it increases the experimental realism and, thus, the internal validity.

Fourthly, the experiments rely on proxies. As described in Chapter 5, most online experimentation is conducted via websites. Such experimental design requires programming, and can thus be rather long and expensive. In comparison, proxy experimentation holds two advantages: it is quick and associated with fairly low costs. All four experiments are designed, conducted, and analyzed by one person.[254] The experimental process is quick, and each experiment is designed, built, run, and analyzed in 1-4 months.

10.3.2.2. Deriving Knowledge from the Experiments

Experiments based on the four common traits mentioned above have proved to be an effective way of testing design and ideas derived from theories. The approach used in this dissertation is useful beyond the usage here, e.g. in open government research and practice more generally. This section will elaborate on this methodology and discuss possibilities and limitations.

As noted in Chapter 5, the research based on experimental method primarily focuses on offline experiments. Cámara and Kobsa (2009) note the following research gap: "To the extent of our knowledge, no research has so far been reported on treating online test design and implementation in a systematic manner" (Cámara and Kobsa 2009, 18). The research that has been conducted, although unsystematic, mainly focuses on A/B/n and multivariate website experimentation (Kohavi et al. 2009a; Crook et al. 2009). To address this gap, I rely on the experiences from the experiments conducted in this dissertation and propose a framework for experimentation.

The experimental process has been extensively covered in Chapter 5 and sporadically in Chapter 6-9. The proposed framework, lean experimentation, builds on this work and focuses on online environments such as open government com-

---

[254] With the exception of SPSS statistical package, no expensive software is being used.

munities. Thus, the focus in this chapter is on open government communities, but there is no reason why the framework cannot be used on non-governmental online communities as well.

10.4. Lean Experimentation

The following sections describe lean experimentation. First the roots of lean experimentation are described, and then the framework is presented and explained. Finally the limitations of the framework are discussed.

Lean experimentation is defined as *an early stage field experiment testing theories or assumptions on real users before developing and implementing features.*

10.4.1.1. Theoretical Roots

The name lean experimentation[255] is chosen, as it is based on the same philosophy and approach as the *lean startup* and *lean analytics* methodologies (Ries 2011; Croll & Yoskovitz, 2013). In addition to lean startup, the lean experimentation framework has roots in several disciplines, including rapid prototyping, innovation, and participatory design. These are briefly described in the following sections.

Lean startup is a methodology that helps evaluate the quality of ideas in areas of uncertainty. The essence of lean startup is to accelerate validated learning and avoid waste. These goals are pursued through experimental examination of assumptions and hypotheses as early on as possible.[256] As the name indicates, the

---

[255] The term is possibly similar to "enlightened experimentation" and "rapid experimentation", as used by Thomke (2001), although he does not define the terms. Similarly, Anderson and Simester (2011) recommend keeping experiments simple and finding easy ways to execute them and starting with proof-of-concept tests.

[256] Lean startup is inspired by the lean processes first used by Toyota in the 1980ies. However, producing physical goods is significantly different from producing software. What Reis keeps from the Toyota Lean process is primarily the aim to reduce waste in production. Waste is here understood as building products and features nobody wants.

philosophy is aimed at startups. Entrepreneurship exists in many large organizations across all sectors, including government (Ries 2011). Ries' lean startup philosophy has a strong commitment to validated learning, which aligns well with the work in this dissertation. Ries proposes that "learning can be validated scientifically by running frequent experiments" (Ries 2011: 9). From the lean startup movement, lean experimentation adopts the belief in scientific testing of assumptions and theories on real users, and that experiments ought to be conducted early on.

Innovation can be defined as "the multi-stage process whereby organizations transform ideas into new/improved products, service or processes, in order to advance, compete and differentiate themselves successfully in their marketplace" (Baregheh, Rowley, and Sambrook 2009: 1334). In innovation activities, a significant amount of time and costs are spent on experimentation (Thomke 1998). Thomke (2001) argues that innovation and experimentation are closely intertwined: "Experimentation lies at the heart of every company's ability to innovate. In other words, the systematic testing of ideas is what enables companies to create and refine their products" (Thomke 2001: 67). Improving these experimental processes is thus a potentially valuable activity. Successful open government relies on successful innovation and/or adoption of successful design patterns from other contexts. To anticipate the effect of a new design feature, it can undergo experimental testing similar to those practices used in the innovation literature.

> *"The ability to experiment quickly is integral to innovation: as developers conceive of a multitude of diverse ideas, experiments can provide the rapid feedback necessary to shape those ideas by reinforcing, modifying, or complementing existing knowledge."*
>
> (Thomke 2001: 69)

From the innovation literature, lean experimentation uses the belief that experimentation and product innovation are aligned. In the context of open government communities, experiments can be used to innovate design and thereby help

363

overcome the participation challenge. And the process of innovation shows the importance of weeding bad ideas from good.

Rapid prototyping is defined by Chua, Leong, and Lim (2010) as "An approximation of a product (or system) or its components in some form for a definite purpose in its implementation" (Chua, Leong, and Lim 2010: 2).[257] The purpose of prototyping is to present early-stage iterations of products to users in order to get feedback on the product itself and potential market (Chua, Leong, and Lim 2010). From rapid prototyping, lean experimentation draws on the concept of presenting early-stage iterations of a concept to users in order to get valuable feedback. The main difference between the two is that rapid prototyping is mainly concerned with physical products,[258] while lean experimentation is concerned with web development and online communities.

Participatory design "explore[s] conditions for user participation in the design and introduction of computer-based systems at work" (Kensing and Blomberg 1998: 167).[259] Participatory design is concerned with the users' interpretation of an artifact: "Participatory design's object of study is the *tacit knowledge* developed and used by those who work with technologies" (Spinuzzi 2005: 165). As explained by Tohidi et al. (2006) lo-fi prototyping is related to participatory design, as users get a chance to influence fundamental aspects of the design. With the risk of stretching the participatory design too far, it can be argued that lean experimentation implicitly is connected to participatory design. Although the users might not know they serve as experimental subjects, their behavior is used to inform and guide the design in important ways.

---

[257] A prototype is usually a physical object, but does not necessarily have to be. Prototypes are usually fully functional and contain full-scale functions (Chua, Leong, and Lim 2010).

[258] It can be used in software development as well, but it is rare (Thomke 1998).

[259] Participatory design has roots in Scandinavia. Based on Marxism, it sought to democratize the workplace by allowing "workers to determine the shape and scope of new technologies introduced into the workplace" (Spinuzzi 2005: 164). Although it was inspired by Marxism, these thoughts are largely left out of contemporary PD (Spinuzzi 2005).

Since lean experimentation does not ask for user feedback directly, but rather leverages their behavior implicitly, important aspects of participatory design are excluded from lean experimentation. This prioritization is made based on the belief that users do not always know what they want, entailing that data stemming from observed behavior is more valid than self-reported data.[260] Furthermore, relying on data from the field rather than the laboratory increases the scalability significantly.

10.4.2. The Process of Lean Experimentation

The goal of lean experimentation is to generate insights and knowledge about how a feature or product will work before programming it. Thus, the more insight and knowledge that can be generated in little time at lower costs, the better the lean experiment is. The lean experimentation framework builds on Davenport (2009)'s *Test and Learn wheel*.[261] This means that lean experimentation relies on a well-tested experimental framework. The contribution of lean experimentation is a reconstruction of the steps in the framework due to a higher emphasis on collecting knowledge early in the development process. This is accomplished by creating efficient experiments that rely on proxies.

---

[260] As noted in section 10.4.3.2, there are other reasons why self-reported data can be problematic. Equally important is discrepancies between self-reported behavior and observed behavior.

[261] According to Davenport (2009), his wheel is adapted from Applied Predictive Technologies.
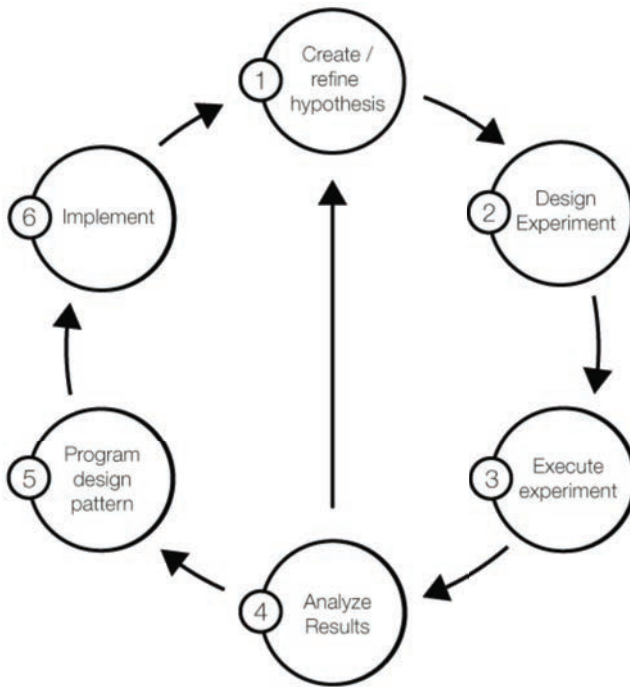
Figure 35: Lean experimentation framework, adapted from Davenport (2009)

Lean experimentation goes through the following six steps: (1) select theory or assumption; (2) design the experiment (often through proxy); (3) execute the experiment; (4) evaluate results - good ideas can be turned into code while bad ideas can be discarded; (5) turn idea into code; (6) implement the design.

10.4.2.1. Step 1: Assumption Selection

In the short history of online experiments, many elements have been tested. Some findings are more valuable than others. In a now infamous experiment, Google could not decide between various shades of blue on their website and ended up testing the advertisement click through rate of 41 shades of blue (Holson 2009). Although Google presumably found the right shade of blue, the overall knowledge and insights derived from such experiment is low.

Knowledge about fundamental human behavior in a defined context is more valuable than small aesthetic changes. To generate useful and externally valid knowledge, design patterns ought to be based on theory and hypotheses. Such theory-based approach is cumulative and will help identify the boundaries of the theories over time.

As described in section 4.2.3, there are several ways to select one theory over another. When relying on social science literature, such choices have to be made constantly due to the many mid-level theories. According to Flyvbjerg (2001), theory selection ought to be context-dependent in order to increase practical relevancy. In this dissertation, two criteria are used to select theories: level of abstraction match and case fit.[262] Based on the outcome of the four experiments, these criteria seem to have generated relevant hypotheses that can increase participation in the examined case. As explained previously in this chapter, the results are more ambiguous than one could have hoped for. It is possible that other theory selection criteria could have produced better results. That being said, ambiguous results are to be expected in the social sciences.

10.4.2.2. Step 2: Design Experiments

As explained in section 5.8.2, true experiments are preferred to pseudo experiments as they can make stronger causal claims. For an experiment to be a "true experiment", three conditions need to be fulfilled: (a) control over variables, (b) use control groups, and (c) use random assignment of subjects.

The most important premise in lean experimentation is to conduct experiments before programming the feature. It is therefore often more efficient to design the experiments through proxies instead of programming it right away. In the following sections, a distinction between website experiments and proxy experiments is used. *Website experimentation* is the process currently used in practice. It relies on coding different design patterns and features on a website, and subsequently testing different treatments against each other. This often happens via

---

[262] These criteria are derived from Arazy et al. (2010).

software such as Wingify Visual Website Optimizer, and Google Website Optimizer. *Proxy experimentation* minimizes the coding by relying on the most efficient and attainable proxy. Sometimes this means running experiments on websites, but this is only one of many possible ways to run experiments.

Both website experiments and lean experiments are conducted on real community members. If it is a large community, only a subset of users is included in the study. But many open government communities are small, and experiments thus include all users. The difference between the two types of experimentation is how the independent variable is manipulated. In website experiments, it is manipulated through different versions of a webpage. In lean experiments, it is manipulated through proxies.

The existing research paradigm on experimentation in online communities focuses on the experimentation of different website variations on an existing website (e.g. Kohavi et al. 2009a; Kohavi et al. 2009b; Crook 2009; Brodovsky and Rosset 2011). This approach is great for testing some website elements, but suboptimal for other elements. When discussing the difference between website and proxy experiments, three parameters are relevant: costs, required buy-in, and type of change.

Costs are crucial in experiments. Making experimentation cheap is critical to increase the adaptation of the methodology. Thomke (1998) argues that the cheaper it is to experiment, the more likely it is that experiments will be used in product development cycles, and the higher the experimental efficiency will be. Thus, by developing a cheaper framework, the likelihood of adoption is increased.[263] Figure 35 shows how experimental design comprises six steps: (1) creating hypothesis, (2) designing experiment, (3) executing experiment, (4) an-

---

[263] Web experiments often have a positive ROI compared to no testing at all. But as noted by Camara and Cobsa (2009), web experiments are often costly in terms of building the necessary infrastructure and development efforts for each test. An advantage of lean experimentation through proxies is that it is faster and less expensive. No new design features have to be coded, as the experimental manipulation is implemented by sending different variations of an email.

alyzing results, (5) turn concept into code, and (6) implement the feature. Lean experimentation is a low cost experimental framework because it reduces the amount of coding significantly.

The *required buy-in* from community owner is higher in website experimentation than in proxy experimentation. Changing a website entails larger risks for the organization than sending out emails to users or awards. A problem with relying on live experimentation in practice is that "Businesses are likely to be careful about the design of experiments and hesitant to deploy anything that sours their users' interaction with a system" (Oktay, Taylor, and Jensen 2010: 2). Conducting lean experiments does not only help overcome buy-in from decision-makers, but also from developers. Many organizations might not be willing to make big changes to existing information systems, unless they know it is a good idea. Running a proxy test can help demonstrate this, and is a more valid solution than copying ideas and features from elsewhere. Although a proxy experiment can backlash, it is likely to be a smaller backlash compared to a full-scale implementation (Davenport 2009).

Type of change. There is not just one way to conduct a given experiment (Thomke 1998). Web experimentation is the better alternative when testing content and layout changes such as call to action, copywriting, images, and similar elements. Website experimentation is an appropriate choice here for two reasons. *Firstly*, it is cheap to create a new page in the community insofar it is not too different from the existing one. *Secondly*, such elements exist independently from other pages in the community. This means it is easier to hold other variables constant as well as avoid confusing the users.

It is more complicated to test features such as points, rankings, mentor programs, etc, as they have to be integrated several places in the community. It is hard to isolate the different conditions, and there is a risk confusing the users. Furthermore, developing new features is often an expensive process, as these either have to be invented or at least copied from other communities and integrated into the code base. This process often takes months or years and a great

amount of resources. In these cases, it is worth testing the idea behind the feature through a proxy.

10.4.2.3. Step 3: Execution

The execution in lean experimentation is done in the wild. As described in Chapter 5, the internet offers a plethora of opportunities for large-scale experiments (op cit.). The internet makes it possible to scale experiments better, and the marginal costs of adding more subjects are low. This dissertation relies on field experiment rather than laboratory experiments. By testing in a natural setting, the full scale of user reactions are discovered. Experiment 1 serves as a good example. The relatively high number of upset emails strongly indicates that some K10 users dislike the use of social comparison information. Such reactions would be harder to detect in a laboratory experiment as the natural context comprising relations to the community is missing. Similarly, Experiment 3 would not be as realistic if random people were selected to be thanked.

Testing in the field also helps overcome the problem of synergistic effects. Synergistic effects occur when the sum of interventions implemented at the same time yield different results than the interventions do when implemented one at a time. The effect can take place when an idea is translated from a simple environment into a complex one. Synergistic effects often remain hidden in laboratory experiments, and that potentially has severe consequences for online communities implementing the ideas. By testing in the field these interaction effects are integral parts of the experiment, even though the independent variable is isolated in the intervention. The important thing in this context is that the dependent variable is measured in a natural field environment.

10.4.2.4. Step 4: Evaluation

The fourth step in the lean experimentation process is evaluation of the experiment. As mentioned above, the lean experimentation framework draws on several design traditions such as lean startup, rapid prototyping, and innovation literature. All recommend multiple iterations of an experiment throughout the

process, although a single iteration sometimes is sufficient (Thomke 1998; Ries 2011).

However, lean experimentation is not conducting experiments on real features, but only proxies of these features. This means that a single experimental iteration sometimes is sufficient to make a qualified decision about whether or not to turn the theory or assumption into design. Iterations are always helpful, but the marginal experimental efficiency is declining because the amount of new knowledge decreases per iteration.

Furthermore, multiple iterations of the same assumption can lead to incrementalism instead of radical innovation. Insofar multiple iterations are used, each step ought to evolve around different theoretical constructs, or even different theories and assumptions aiming to improve one goal. This is one of the reasons why several different social psychology theories are used in this dissertation.

Thomke (1998) recommends experimental diversity due to diminishing returns on iterative experimentation with the same methodology. Thus, if the same feature is tested in multiple iterations, it will probably generate more valuable insights if different experimental designs are used. Thus, due to the diminishing returns, different kinds of lean experiments should be tried over time. This is feasible, because lean experimentation is not one experimental design, but rather an approach to designing tests.

10.4.2.5. Step 5: Development

The basic premise of the lean philosophy is that most new products and features fail (Reiss 2011). This anticipation aligns well with online experiments, where it is difficult to predict the outcome of design ideas (Kohavi et al. 2009).[264] As noted by Cámara and Kobsa (2009) "a very high percentage of the tested modifica-

---

[264] Humans tend to look for hidden patterns, even where there are none. In an experiment where subjects were to guess random colors, humans lost to rats because of this (Mlodinow 2008 referred to in Kohavi et al. 2009).

tions are usually discarded since they do not improve the site performance. As a consequence, a lot of effort is lost in the process" (Cámara and Kobsa 2009: 18). According to Kaushik (2006), "80% of the time you/we are wrong about what a customer wants." (Quoted in Kohavi 2009b: 8). Internal experiments in Microsoft show that only around 1/3 of all experiments produce an improvement. Around 1/3 yield no significant effect, while the last 1/3 perform worse than before. Similarly, Netflix considers 90 % of what they do to be wrong (Kohavi 2009b). Amazon is more optimistic and assumes a success rate of just under 50 % (Kohavi et al. 2009). What seems like a good idea to the researcher or designer might not be what the users want. What works in one context might not work in another. And what works today might not work later on.

The biggest problem with the current website experimentation paradigm (e.g. Kohavi et al. 2009; Kaushik 2009) is that ideas have to be turned into code prior to testing. This existing paradigm approach can help avoid implementing features with a negative impact on KPIs but it cannot save development costs. In this regard, Thomke (1998) uses the term experimental efficiency, which he defines as "the economic value of information learned during an experimental cycle, divided by the cost of conducting the cycle" (Thomke 1998: 745). The earlier the experiment is conducted, the more valuable the knowledge becomes, and the more efficient the experiment is.

Tohidi et al. (2006) make a distinction between "getting the design right" and "getting the right design". The former is concerned with the details of a particular design while the latter is more fundamental and entails choosing between very different designs. In lean experimentation, the aim is to test the idea *before* coding, not after. The aim of lean experimentation is, thus, to get the right design, and not just getting the design right.

The methodology used in this dissertation is not necessarily the final implementation of ideas, although it can be. For example, emails can be used as a one-off proxy for measuring the reaction to a hypothesis. But emails can also be used to implement the idea subsequently. In experiment 1, emails are used to send out

social comparison information. If a decision is made to provide social comparison information to some user groups on a regular basis, K10 can either show this information on the user's profile, or send a monthly email to the users.

10.4.2.6. Step 6: Implementation

If a feature is programmed into the online community, the users will be exposed to it every time they visit the website. Therefore, if the experiment provides the desired result, it is easy to use the experimental version in the future, as it has already been developed and tested. The effect of an email is likely to decline fairly quickly (Beenen et al. 2005) since users do not see the experimental change every time they interact with the website. This limited effect time of emails is an advantage when an experiment does not yield the desired outcome. The effect can be tested before any potentially expensive coding is done, thereby saving resources in case of a failed experiment.

In sum, using proxies in experimentation is most often *not* an alternative to coding in the long run. However, it helps avoid programming features and products that do not support community goals.

10.4.3. Advantages and Limitations of Lean Experimentation

As Table 60 shows, the lean experimentation framework offers different advantages and limitations. The advantages include examining causality, testing competitive design, examine actual behavior, influence on decisions, and economic benefits. The disadvantages include fit between task and experimentation, limitations on available metrics, risk of confusing users, incrementalism, subject recruitment, lack of qualitative depth, measures short term effects, and low quality proxies.

Benefits	Limitations
Examine causation	Limited to available metrics
Observed behavior	Lack of Qualitative depth
Influence on decisions	Incrementalism
Economic benefits	Subject recruitment
	Confuses users
	Measures short term effects

Table 60: Advantages and disadvantages of online experimental methodology

10.4.3.1. Examine Causation

Most online communities are developed incrementally. New design patterns and features are added periodically while existing ones are removed. Without isolating variables, it is impossible to know if the changes are facilitating or hindering the desired behavior. The research question and statistical hypotheses in this dissertation are concerned with determining causality: How does a theory-based design affect behavior? Can a certain operationalization of social psychology theory increase participation? When examining causal claims, experimental methodology is considered to be the most precise (McGrath 1995; Sani et al. 2006; Gilovich et al. 2010).

10.4.3.2. Observed Behavior, Not Self-Reported

There are several ways to acquire knowledge about people. One approach is to ask them, another is to observe them (McGrath 1995). Social psychology theorists have a long history of using experiments. Nonetheless, the majority of the research reported in Chapter 5 relies on self-reported data collected through surveys and linguistic analyses. This means that users are asked about their intentions and thoughts.

Taking the subjects' words at face value is often problematic. Social psychologists argue that we cannot always trust what people say. Self-reported data is often deceptive, whether it is intentional or not (McDermott 2002). There are often significant discrepancies between self-reported behavior and actual behavior, but by observing actual behavior, it is possible to eliminate self-reporting errors (Nickerson 2007). When findings rely solely on self-reported data, and not

on behavioral data, there is a risk of bias and even misleading results. For example, based on meta-analytical reviews of social loafing literature, Karau and Williams (1993) find plenty of evidence for social loafing in the behavioral data, but none in the self-reported data. This indicates that people either do not realize that they are engaging in social loafing or they are unwilling to report it.

This problem also exists in practice, and the same concern is echoed in innovation literature. According to Ries (2011) most users do not know what they want. This is why experimentation works better than surveys: "We must learn what customers really want, not what they say they want or what we think they should want" (Ries 2011: 38).[265]

To some, it is difficult to observe behavior. For example, Flavian and Guinaliu argue "due to the difficulties associated with measuring real behaviors, we concentrate on intentions to participate in an online travel network in the long term" (Flavian and Guinaliu 2010: 899). However, as argued in Chapter 5, it is easier and cheaper to set up large-scale online field experiments than ever before (Kaushik 2009). Measuring real behavior might be more difficult than surveying users, but it is by no means impossible in online settings (Reips 2007).

10.4.3.3. Influence on Decisions
In website experimentation, new ideas are often coded before they are tested. This implies a stronger commitment to the idea, as an upfront investment is made in order to turn the idea into code. It is often hard to accept that the development and the many allocated resources have led to no improvement and rendered themselves useless. This might particularly be a problem in government agencies where there is high degree of risk aversion (op cit).

---

[265] For example, based on customer inputs from a survey, WalMart spent millions of dollars making their aisles less cluttered. After the initial costs, the real costs started to occur as it turned out that the customers were wrong about their own needs. It ended up costing WalMart approximately $1-2 billion in lost sales (Hurst 2011).

According to Tohidi et al. (2006), the data and insights that can be generated from lo-fi and hi-fi prototyping are comparable. However, getting the data early is much more valuable than getting it late. It is therefore meaningful to create several lo-fi prototypes early on. The information about the idea is more likely to make an impact on decision-makers if it is tested before expensive commitments are made. Furthermore, trying out different design ideas lead to more informed and open decisions (Tohidi et al. 2006). [266]

10.4.3.4. Economic Benefits

Throughout the history of HCI and egovernment, costs have always been a concern (Carroll 2003). Experimental methodology has traditionally been an expensive venture, but technological developments have made it cheaper. Experimental modes are at least partly driven by technological development (Kohavi et al. 2009b). Thus, experiments are likely to become cheaper over time; especially in online environments.

One important aim of experimental design is to avoid pursuing ideas that have no or negative value. By identifying needs and problems upstream, it is possible to prevent waste downstream (Thomke 2001). The earlier in the process an experiment is conducted, the more valuable it is, because it enables an early and thereby cost effective refocus if need be: "experiments are particularly desirable when they are performed early on so that unfavorable options can be eliminated quickly and people can refocus their efforts on more promising alternatives."

---

[266] Another argument for early experimentation is that people tend to give more honest and critical feedback when they think everything can be changed (Tohidi et al. 2006). For example, one type of critique is given in the early stages of development than the day before the product or feature is scheduled to launch. Much more can be changed early in the process than later on. Tohidi et al. (2006) find that users are more likely to give honest and critical feedback when they are shown multiple designs instead of a single one. A possible explanation for this is that users are less concerned about disappointing designers when they have yet to commit to a single design. One problem with high fidelity prototyping such as website experimentation is that the product/feature seems almost ready to launch. This gives opponents the impression that only small changes can be made (Thomke 2001).

(Thomke 2001: 71).[267] More specifically Kohavi et al. (2009) sketch out four scenarios where experiments offer different kind of value.

Scenario	Value
Experiment shows idea is as good as expected	Experiment adds little value
Experiment shows idea is even better than expected. A breakthrough.	Experiment adds value by showing more resources ought to be devoted to this.
Experiment shows idea is not as good as thought, as it hurts important metrics.	Experiment adds value by not making the website worse.
Experiment shows idea does not perform neither, nor worse than status quo.	Experiment saves deployment costs. If the experiment is conducted upstream, development costs will always be saved.

Table 61: When and how experiments add value. Based on (Kohavi et al. 2009)

Table 61 shows that early experiments add little value when an idea lives up to the expectations. The real potential lies in identifying the excellent, mediocre, and bad ideas. This helps devote either more or less resources to the different initiatives.

10.4.4. Limitations Of Lean Experimentation

Although there are several advantages to using lean experimentation in the work with online communities and open government, the methodology has its shortcomings. The disadvantages are fit between task and experimentation, limited to available metrics, lack of qualitative depth, confuses users, not finding optimal solutions, subject recruitment, and measures short term effects.

10.4.4.1. Limited to Available Metrics

As described in Chapter 5, experimental methodology is usually limited to quantitative variables (Davenport 2009). Although the independent variable can be qualitative, the dependent variable is usually measured quantitatively. This sets limitations about what can meaningfully be measured.

---

[267] As noted by Kohavi et al. (2009), "Think of how much effort can be saved by building an inexpensive prototype and discovering that you do not want to build the production feature at all!" (Kohavi et al. 2009: 8).

In this dissertation, this entails that participation primarily is measured in terms of counts: how many times a user writes a post on K10. This is a meaningful measurement as there is a participation challenge in most open government communities. But this measurement does not tell the full story, because it contains no information about the quality of the posts. However, no one has yet found reliable quantitative metrics of content quality (Farina et al. 2010).

This issue is particular problematic in academia as a wider range of outcomes and dependent variables are measured here.[268] Being able to measure qualitative outcomes is important in open government; especially when one applies a collaborative democracy perspective.[269] As described in section 2.3.3, there are no reliable way to quantitatively measure participation quality at this point.

10.4.4.2. Lack of Qualitative Depth

The experiments in this dissertation mainly focus on number of contributions. Other recurrent dependent variables include click through rate and conversion rate. In its current state, web experimentation is largely limited to quantitative metrics. It is possible to measure where users click, how many users click on desired links, how many posts they write, how long the posts are etc. But these variables do not offer qualitative explanation of the observations (Kohavi et al. 2009). In a broader perspective, experimentation ought to be supplemented with qualitative methods such as interviews and open-ended user surveys (Davenport 2009). However, triangulation might be unattainable or impractical in terms of each experiment. When running natural field experiments, the users are not asked to participate in the experiments, thereby making it hard to make logical follow up interviews and data collection.

---

[268] In the industry, it is often sufficient to know how quantitative metrics perform.

[269] For more on collaborative democracy, please refer to section 2.3.1.

10.4.4.3. Confuses Users

Experimentation means that different variations of the same thing exist. In optimal conditions, one user is only assigned to one condition. But perfectly controlled subject assignment is hard to maintain in some experimental setups, particularly when manipulating the independent variable on the website. When users access multiple computers or decline cookies, the same user can be exposed to multiple treatments during the course of the experiment. This potentially causes confusion and dissatisfaction among the community members.

Conducting experiments entails doing something differently. Change is necessary in order to adapt to the norms and expectations on the internet. As demonstrated by Wohn et al. (2012), much internet behavior is driven by habit. Changes interfere with workflows and habits, and therefore constitute a threat to the usage of an open government community.

10.4.4.4. Not Finding Optimal Outcome

As described above, one advantage of lean experimentation is that different ideas and designs can be compared directly. But testing different ideas does not reveal optimal solutions, merely the best designs currently known (Thomke 1998). Thus, the condition performing best in any given test is not necessarily the optimal outcome, but merely the best outcome of the included possibilities. However, no methodology can optimize beyond the inputs going into the frameworks and models.

10.4.4.5. Subject Recruitment

Recruiting subjects *per se* is fairly easy in lean experimentation. However, in some contexts the initial population is small. This is often the case in B2B or G2G contexts, as there are fewer users than in B2C or G2C (Anderson and Simester 2011).

Countries with small populations and niches with few people in it are more susceptible to this issue. This entails that it is more complicated to conduct experi-

ments in Danish open government communities than in corresponding US communities. Furthermore, some open government communities evolve around niches with very small populations. For example, Danish Regions operate a number of health related online communities in Denmark evolving around illnesses such as child incontinence, recurrent miscarriages, and lymph and chronic edema. As some of these illnesses are quite rare, the populations are too small for statistically valid experiments.

10.4.4.6. Measures Short Term Effects

The time frame used in the experiments in this dissertation is usually four weeks per experiment: two weeks before the intervention, and two weeks after. This is a normal time frame for such experiments. One reason for this short time span is that experiments grasp immediate reactions and short term behavior more easily (Anderson and Simester 2011). One risk to this approach is that short-term behavior often differs from long term behavior (Kohavi et al. 2009). This means that the experimental findings cannot be generalized across time. More research such as replication and longitudinal studies are needed to account for the existence and size of this critique.

A related issue is the novelty bias, entailing that people react differently when exposed to something new (Kohavi et al. 2009). After some time, the effect will wither, simply because it is not new anymore. Regular replication and evaluation can help mitigate this issue.

10.5. Research and Practice

This chapter is segmented into implications for practice and implications for research. Despite this distinction, research and practice do not exist in different vacuums. Issues, findings, and methodology can be used in both research and practice.

The concerns about external and temporal validity apply to researchers and professionals alike. Both groups need to know how far research findings extend over

time and place. But at this point, no research documents to which extend findings from online community X extend to online community Y. It is therefore desirable, and often necessary, to test a design pattern when copying it from another context. Without such knowledge, researchers will draw wrong conclusions, and open government professionals will create communities that are sub-optimally designed, and thus in risk of failing.

Lean experimentation can be used in both research and practice. The idea of applying experimental methodology *in practice* is not new. 90 years ago, Claude Hopkins said "Almost any question can be answered cheaply, quickly and finally, by a test campaign. And that's the way to answer them – not by arguments around a table" (Hopkins 1923 quoted in Kohavi et al. 2009: 177). Thus, it is not the use of experimentation that is new, but the technology (Davenport 2009).

Experimentation is becoming increasingly popular in private companies and other professional settings. On a high level, Kim, Barua, and Whinston (2002) argue that the increased usage of experimentation is "allowing hypotheses to be tested while not sacrificing parallelism to the real business world" (Kim et al. 2002: 216). Therefore, experimentation "needs to come out of the laboratory and into the boardroom" (Davenport 2009: 8). Lean experimentation aspires to contribute to the introduction of online experimentation into open government.

As described in Chapter 5, the lean experimentation process has to some extend already been used *in research*. More specifically, several proxies have successfully been used as both dependent and independent variables.[270] Grimmelikhuijsen (2012) uses four different websites as a proxy. Beenen et al. (2005), Chen et al. (2010), Brajer and Gill (2010), Keane (2010), Butler and Broockman (2011), Bosch, Carnero, and Farre (2010), Baldini and Federici (2011) and Violi et al. (2011) all rely on emails to expose subjects to different experimental conditions. Gaines et al. (2006), Sniderman and Piazza (1993) and Katz (2011) use survey to make controlled experiments. Finally, Lewis and Reiley (2010), Guerini et al.

---

[270] Please refer to Chapter 3 and 5 for a review of this research.

(2010), Chan (2011), and Dow et al. (2010) use paid advertisements in experiments. The point is that online experimental research in relying on proxies is not new, but it lacks a systematic framework. As described in section 5.9, online experimental methodology is still in an early stage. Little research explicitly examines and discusses the possibilities and limitations of experiments in online environments compared to offline. Furthermore, much experimental research takes a long time to conduct and report, and is therefore less relevant when published. For example, the experiments in Chen et al. (2010)[271] were conducted in 2006 and published in 2010. The hope and aspiration is that lean experimentation can serve as a common framework and further the production of more current knowledge.

10.6. Chapter Summary

The experiments conducted and reported in this dissertation can be translated into a number of implications for research and practice.

In order to help solve the open government participation problem *in practice*, five points are discussed. *First*, how participation patterns can be altered, and under which conditions it is desirable to change it. *Secondly*, related to this is opportunity costs and suboptimal allocation of resources, as voluntary participation to some extend is a zero-sum game. Increasing participation in one online community potentially decreases participation in another. *Thirdly*, K10 represents a niche audience quite different to the mainstream internet user, making it relevant to assess how different groups and communities will react differently to identical design patterns. *Fourthly*, throughout the four experiments, different kinds of emotions are detected. It is discussed how to cope with both negative and positive reactions in open government research. *Finally*, the use of goals is discussed.

---

[271] This paper is reviewed in section 4.3.

The empirical research conducted in this dissertation comes with a number of challenges to *research* in terms of external and temporal validity. Seeking to address these concerns, an experimental framework named lean experimentation is introduced. As it potentially helps solve the participation challenge in an efficient and scalable manner, this framework is the main contribution of this chapter, and one of the main contributions of this dissertation. Lean experimentation goes through the following six steps: (1) select theory or assumption; (2) design the experiment (often through proxy); (3) execute the experiment; (4) evaluate results - good ideas can be turned into code while bad ideas can be discarded; (5) turn idea into code; (6) implement the design. The framework is based on the experiences from this dissertation and has yet to be directly applied in research and practice.

Chapter 11: Conclusions

11.1. Introduction

In this dissertation, the following research question is asked: *Can social psychology theory help increase participation in open government communities?* The aim of this chapter is to answer this research question by leveraging the findings and insights from the previous 10 chapters.

The chapter progresses the following way: First, the substantive, theoretical, and methodological findings from the four experiments are reported. These three types of knowledge from the four experiments are summed up in Table 62. Combined, these insights provided firm ground for answering the research question. The chapter, and dissertation, is concluded with a number of policy implications and call for further research.

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Substantive results				
Findings	• Providing information about participation affects behavior • Surprisingly, no significant difference between comparative and non-comparative information is detected	• Users assigning themselves an ambitious goal increases subsequent participation • Only significant effect for subjects committing to a higher goal	• Thanking subjects increase subsequent participation and make users happy • As a medium, internal message system is more effective than email	• Different depend variables give different answers • Group identity does not affect subjects' participation. But it does affect engagement
Effect sizes	• Medium-large	• Large	• Medium	• Small
Internal validity	• General good internal validity. • Only concern is that sending emails per se might have an effect	• Significant change in control group, indicating potential spurious variable	• Potential carryover effect. Compromises effect size but not logic	• The two-week time span might have been too short • The different advertisements might attract different people to begin with
External validity	• Successful partly replication of findings in Chen et al. (2010) indicates good external validity • K10 users are younger than general early retirement pension and flex job recipients • Unknown to which degree these findings apply to other open government communities	• The compromised internal validity means that one has to be careful with extrapolation • K10 users are younger than general early retirement pension and flex job recipients • Unknown to which degree these findings apply to other open government communities	• Non-random sampling makes it challenging to extrapolate • K10 users are younger than general early retirement pension and flex job recipients • Unknown to which degree these findings apply to other open government communities	• K10 users are younger than general early retirement pension and flex job recipients • Unknown to which degree these findings apply to other open government communities
Construct validity	• Number of posts might not	• Lack of feedback after the	• Possibly unclear who is	• The difference between

	be most salient comparison metric for the subjects as it lacks qualitative depth	goal is set potentially explains few significant results	thanking the subjects	beneficial to small group condition might not be sufficiently different to beneficial to large group condition
Implications for design	• Sending participatory information alters participation patterns • No significant difference between comparative and non-comparative information • Participatory information might upset users • Social comparison theory can be implemented through stats on the website, leaderboards or regular emails.	• Simply asking users how much they will contribute might increase participation • Goal-setting can be implemented in several ways: in sign-up process, in surveys and through direct outreach	• Recognition can be implemented in different ways including automated recognition, peer-to-peer recognition and expert recognition • Expressing gratitude makes users feel happy	• First impressions matter. Either because it affects the subjects' subsequent behavior or because it attracts certain types of subjects • Advertisement platforms are useful for recruiting and assigning subjects
Theoretical insights				
Applied Theory	• Social comparison theory	• Goal-setting theory	• Self-efficacy theory	• Social identity theory
Methodological insights				
Experimental design	• Split-plot design	• Split-plot design	• Within-subject design	• Between-subject design
Experimental proxy	• Email	• Survey	• Email & message	• Advertisement
Number of subjects	• n = 3,959	• n = 84 (n = 4,081 including non-respondent control group)	• n = 29	• n = 453

Table 62: Summary of the substantial, theoretical, and methodological results from the four experiments

386

11.2. Substantive Results

In this section, four aspects of the results are described: the substantive findings, effects sizes, validity, and the implications for design.

11.2.1. Findings

11.2.1.1. Experiment 1

The results of experiment 1 suggest that participatory information can lead to improved participation among the users who usually do not participate a lot. An interesting aspect of this finding is that the comparative information is not significantly different from the non-comparative information. In fact, the p-values are almost similar for the below mean users in *group 1* (as T = 0, z = -2.226, p < 0.026, r = 0.909) and *group 2* (T = 50.50, z = -2.272, p < 0.023, r = 0.474), although the effect size is larger in the former. This might indicate that non-comparative participation information can be as effective as comparative information. A possible explanation is that subjects can derive comparative insights from seemingly non-comparative information.

Although the intervention is non-significant for the subjects that contributed above the means in the pre-intervention period, they did contribute less after the experiment than they did before. This results in a decline of posts at the aggregate level from 1,195 to 1,111 posts. The overall drop, which is not statistically significant (in *group 1*, T = 36.50, z = -1.006, p < 0.315, r = 0.269, and in *group 2,* T = 15, z = -1.277, p < 0.202, r = 0.404), is a result of a skewed participation pattern where a few users are doing a lion share of the contributions. If the participation had been more even, an increase among the below-means contributors would more than offset a decline among the above-means contributors.

The similar effect of the comparative and non-comparative information can also be interpreted in another way: the construct of the social comparison theory does not yield the effects expected based on the theory. The significant effect in some subject segments might simply be due to other factors at play, e.g. receiving

an email. However, this explanation cannot account for why only the below-means subjects react to the intervention.

	Hypothesis	Verdict (all subjects)	Verdict (all participatory subjects)
H_o	Receiving social comparison information has no effect on subsequent participation	Rejected	Rejected
H_a	Receiving social comparison information has an effect on subsequent participation	Rejected	Rejected
H_{a1}	Receiving social comparison information reduces the contribution level among top contributors	Rejected	Rejected
H_{a2}	Receiving social comparison information increases the contribution level among non-top contributors.	Rejected	Verified
H_b	Receiving non-comparative information has an effect on subsequent participation, but smaller than comparative information.	Rejected	Rejected
H_{b1}	Receiving non-comparative information increases the contribution level among top contributors	Rejected	Rejected
H_{b2}	Receiving non-comparative comparison information increases the contribution level among non-top contributors.	Confirmed	Verified

Table 63: Evaluation of the seven research hypotheses in experiment 1

As Table 63 shows, the null hypothesis is rejected. Two alternative hypotheses are verified while another four alternative hypotheses are rejected. The effect sizes vary between low and medium-large. The social comparison based intervention has the expected effect on *some* subject segments, but not all. Social comparison theory can therefore only account for *some behavior* in *some segments* of the K10 users. This shows on a high level that different users and segments behave differently and that segmentation is necessary.

11.2.1.2. Experiment 2

Based on goal-setting theory, it is expected that the subjects' behavior reflect self-assigned goals. In this experiment, the 84 subjects in the experimental condition make an implicit commitment to future participation level. The intervention and commitments are made through a survey.

On the overarching level, there is no significant effect of the intervention measured on post-intervention participation compared to pre-intervention participation (t(84) -2.830, p < 0.006, Cohen's d = 0.44). However, taking the subjects' goals into consideration yields better insights. The subjects setting a *higher goal* are contributing significantly more after setting the goal (T = 26.50, Z = -2.934, p < 0.003, r = 0.623). Despite a high effect size, no significant correlation (Spearman's Rho, r = (21) 0.606, p < 0.116) can be detected between the two variables. The subjects setting a *status-quo* goal do not have significantly different participation patterns after the experimental intervention (T = 29, Z = -0.357, p < 0.721, r = -0.09). This is to be expected. The statistically significant Spearman's correlation (r = (14) 0.614, p < 0.015) shows, however, that goal is a reasonable predictor of subsequent participation.

Contrary to the pre-intervention expectations, the *lower goal* subjects increased their participation. However, the change is non-significant (T = 14, Z = -0.567 p < 0.057, r = -0.15). The Spearman's correlation (r = (13) 0.846, p < 0.009) does show a strong correlation, which is statistical significant, between goal and participation, indicating that the two variables are associated.

	Hypothesis	Evaluation
H_0	Setting a participation goal has no effect on subsequent participation	Rejected
H_a	Setting a participation goal makes subjects strive to achieve the goal	Partially verified

Table 64: Evaluation of the research hypotheses in experiment 2

These results entail that the null-hypothesis is rejected, as the high goal subjects contribute significantly more post-intervention. As shown in Table 64, the alternative hypothesis (H_a) is only partly verified for two reasons: no significant change is observed in the low goal segment, and no significant correlation is found between goal and participation for subjects setting a high goal. More problematic, there is a significant change in one of the control groups (in *group 2* comprising subjects who answer a survey but are not asked to set a participation goal). These subjects contribute significantly more in the two weeks after the

intervention than in the two weeks before. This questions the overall validity of the experiment.

Similar to the first experiment, the theory-driven intervention succeeds in finding significant effects among some segments of the users. This supports the high-level finding in experiment 1: that social psychology inspired design changes have some explanatory and predictive power, but are far from perfect.

11.2.1.3. Experiment 3

According to self-efficacy theory, gratitude and appreciation are important reinforcement mechanisms. In the third experiment, the 29 subjects are thanked for their previous contributions to K10. The subjects' post-intervention behavior is significantly different when they are being thanked than it is when they are not receiving this gesture ($t(22) = 2.186$, $p < 0.040$, Cohen's $d = 0.454$). After being thanked, the subjects contribute significantly more with a medium-large effect size. This means that the theoretical claims are backed up by this experiment. Hence, thanking K10 users for their efforts has an impact on the users' subsequent behavior.

The results are not highly significant as the probability level is $p < 0.04$, entailing a 4 % risk that the variance is caused by nuisance variables and not the independent variable. The effect size is medium-high, which might indicate that the relatively high p-value is a result of the low sample size. The carryover effect, that always constitutes a risk in within subject design, might be another possible explanation for the high p-value.

	Hypothesis	Evaluation
H_0	Knowledge about users gratitude for the subjects' previous contributions does not affect the subjects' future contributions	Rejected
H_a	Knowledge about users gratitude for the subjects' previous contributions affects the subjects' future contributions	Verified

Table 65: Evaluation of the research hypotheses in experiment 3

The qualitative responses show that expressing gratitude can be understood in terms of self-efficacy as well as generalized reciprocity. One of the subjects mentions the satisfaction of giving back to K10. On a higher level, this message underscores that participation motivators are complex and multidimensional, and cannot be limited to a single variable.

Contrary to the first two experiments reported in this dissertation, the null hypothesis is clearly rejected and the alternative hypothesis clearly confirmed, cf. Table 65. This means that the theory applied here offers relatively sound explanatory and predictive insights into the open government community to which it is applied.

11.2.1.4. Experiment 4

Social identity theory contends that people seek to enhance the groups they belong to in order to enhance their own self-image. The aim of the fourth experiment is to examine if knowledge about the beneficiary of a contribution affects participation. This is done by highlighting different types of beneficial scenarios: a contribution can be *beneficial to self*, be *beneficial to a small group*, be *beneficial to a large group*, or it can be *beneficial to no one*.

The experiment fails to detect any significance in terms of number of posts ($F_{(0,13)} = 0.186$, $p < 0.906$), threads ($F_{(0, 40)} = 0.940$, $p < 0.421$), new accounts ($F_{(0,10)} = 1.307$, $p < 0.272$). But other included dependent variables, CTR ($F_{(1,439)} = 15.05$, $p < 0.000$), pageviews ($F_{(3)} = 3.516$, $p < 0.015$), time on site ($F_{(3152, 9234)} = 5,190$, $p < 0.002$), and bounce rate ($F_{(3, 126)} = 3.362$, $p <$

0.019) are significantly different across the four groups. Because of these very mixed results, the null hypothesis is only partly rejected, cf. Table 66.

	Hypothesis	Evaluation
H_o	The group constructs have no effect in terms of engagement and contribution behavior.	Partially rejected
H_a	Subjects assigned to the *beneficial to others* groups will engage and contribute more than subjects assigned to *benefit self* and control group.	Partially verified
H_b	Subjects assigned to the *beneficial to small* group will engage and contribute more than subjects assigned to *beneficial to large* group.	Rejected

Table 66: Evaluation of the research hypotheses in experiment 4

The H_a hypothesis is partly verified as the two *beneficial to others groups* perform significantly better than the *beneficial to self group* and *control group* in terms of time on site and number of pageviews. However, the *beneficial to self* group outperforms the *beneficial to others groups* with respect to CTR. There are no significant differences on other three parameters. The H_b hypothesis is rejected, as there are no statistically significant differences between the *beneficial to similar others group* and the *beneficial to others group*.

Although this experiment is the *odd man out* of the four experiments, because it targets new K10 users as well as existing users, the findings echo experiment 1 and 2. Some significance is found in certain groups. But the differences are not as ubiquitous as one could desire.

11.2.2. Effect Sizes

As noted above, the four experiments have not resulted in pervasive and consistent results across all dependent variables. The same pattern applies to effect sizes. The effect sizes vary within and between the experiments. In experiment 1, the effect sizes are *medium to large*. Experiment 2 has *large* effect sizes while the effect sizes in experiment 3 are *medium*. Finally, the effect sizes in experiment 4 are *small*.

It is not surprising that different experiments have different effect sizes. After all, operationalizing different theories will give different results. The variation in effect sizes shows that it is challenging to use social science theories in design, and that different people react differently to the same change. Encouragingly, these effect sizes show that social sciences can be used to make a large impacts in terms of participation.

11.2.3. Validity

Three types of validity are constantly evaluated in this dissertation: internal validity, construct validity, and external validity.

11.2.3.1. Internal Validity

Internal validity is the most important concern in experimental methodology. The internal validity of the experiments is generally satisfactory. The internal validity of experiment 1 is good. The aim of the experiment is to see if the content of different emails affect subsequent behavior. The only real concern here is that receiving emails *per se* might have an effect on the subjects. This entails that the detected changes potentially are caused both by the content and the delivery methodology.

The internal validity in experiment 2 is the most problematic of the four experiments. The reason is significant changes in one of the control groups (more specifically in *group 2*, who answered the survey but were not asked to set a goal). This indicates an interacting spurious variable that is not accounted for. It seems likely that answering a survey *per se* has an effect on subsequent behavior, as no significant changes are detected in the second control group, comprising K10 users who *did not* answer the survey.

The internal validity of experiment 3 is potentially suffering from a carryover effect. The carryover effect might occur because the time between the two experimental rounds should have been longer. It is important to notice that this car-

ryover effect does *not* compromise the logic of the experiment, but potentially reduces the effect sizes.

It is not only experiment 3 that would have benefitted from a longer time span. Experiment 4 is conducted over a two-week period. But this timespan might be too short to detect long-term effects of the intervention, and thus account for the general lack of significant effects in terms of contribution level. Moreover, the internal validity might be weakened by the fact that different advertisements might attract different types of subjects rather than change the subjects through priming. Existing differences between the subjects from the three groups might thus be the real reason behind the different behavior.

11.2.3.2. Construct Validity

When assessing the construct validity of the experiments of this dissertation, it is important to keep the overall research question in mind: the aim is to contribute to solving a practical problem - participation challenge - not improve social psychology theories. If the latter was the purpose, the construct validity would be unsatisfactory because the strength of the theoretical construct is not measured.

As described in Chapter 2, the dependent variable, post count, can be contested because it is too simple and does not grasp the quality of contributions. The ideal dependent variable would be a variable that grasps post quantity and quality. Alas, no such variable exist at this point without a tremendous amount of manual coding and subjective assessments (op cit.). As this dissertation mainly is concerned with the participation challenge, participation quantity is deemed more important than participation quality. But this choice is a limitation in terms of construct validity. It is discussed further in section 11.5 how to overcome this limitation.

At a general level, the construct validity is acceptable. In experiment 1, the most problematic aspect is that the number of posts might not be most salient comparison metric. The experiment might have had different effect sizes with quali-

tative metrics. In experiment 2, no feedback is given after the goal is set. According to goal-setting theory, ongoing feedback is positively correlated with effort and persistence (op cit.). Lack of feedback after the goal is set potentially explains few significant results.

In experiment 3, the construct validity is generally high. The only minor concern is that it might be unclear who is thanking the subjects. The message can be understood as an expression of gratitude from the owner, from peers, or a combination of both. Finally, the construct validity in experiment 4 had been better if the differences between *beneficial to small group* condition and *beneficial to large group* were bigger.

11.2.3.3. External Validity

Of the three types of validity, the weakest one in this dissertation is arguably external validity. K10 is used as the experimental environment, in which the four experiments are conducted. As described in Chapter 3, K10 is considered to be an open government community, and therefore a suitable environment for examining possible solutions to the participation challenge. The aim is not to create knowledge that can be generalized to all open government communities. From a *technical* point of view, there is nothing that makes K10 unique from other open government communities. The basic functionality and interface share more commonalities than differences to the average online community – whether governmental or non-governmental. From this perspective the external validity is expected to be good. Despite these technical similarities, it is too early to tell if different open government communities are sufficiently alike for such extrapolation in terms of *population* and *purpose*. This entails that the results stemming from the K10 experiments should not be immediately empirically generalized to all open government communities.

Moreover, the aim is not to make empirical generalizations to the entire population in Denmark. Due to the unique characteristics of K10 users (most of them are on government benefit programs for people with severe illnesses), it is hard

to know how other segments of the Danish citizenry would react to the same stimuli, let alone people from other countries and cultures. Finally, it is not the aim to generalize to all early retirement pensioners and flex job recipients. As further described in Chapter 3, K10 users are younger than general early retirement pension and flex job recipients. Therefore, the aim is rather to create knowledge that allows analytical generalizations to the open government research community and practice.

Experiment 1 is successful in partly replicating the findings in Chen et al. (2010), which indicates that the external validity is good. The compromised internal validity of experiment 2 means that one has to be careful with, or completely avoid, extrapolating to other communities. The non-random sampling used in experiment 3 also means that extrapolation has to be made with caution. Finally, the scarce details on the subjects in experiment 4 also warrants caution with extrapolation.

11.2.4. Implications for Open Government Communities Design

The findings from the four experiments yield substantial implications for practice that can be used to improve open government community design. Although the HCI community is sometimes criticized for being too focused on implications for design,[272] these implications constitute an important part of contributions in the HCI community.

According to the results of experiment 1, it is possible to alter participation patterns by exposing users to participatory information. Surprisingly, no significant differences between comparative and non-comparative information are detected. Many communities already publish information about their users' participation pattern. For example, Twitter shows the user's number of Tweets on the user's profile, cf. Screenshot 15

---

[272] This critique is discussed further below in this section

Similarly, Foursquare shows the number of check-ins at the user's profile, cf. Screenshot 16.

The point is that such participatory information seems to affect participation pattern, when it is made salient. It is, however, critical to note that this information *alters* participation patterns, but it does not necessarily *increase* participation. This experiment – as well as Chen et al. (2010) – shows that this type of information creates participation convergence toward the mean: the very active contributors contribute less while the less active contributors contribute more. Due to a much skewed participation pattern at K10, the overall effect was negative: the increased participation among the less active contributors could not offset the participation decrease among the highly active contributors. It is therefore critical to examine the participation pattern *before* making participation information more salient.

Experiment 2 shows that simply asking users how much they will contribute might increase subsequent participation. More specifically, this works if one manages to make the user set a higher goal than their previous participation ef-

fort suggests. As the goal-setting theory suggests, the effect of the goal depends on a number of factors concerning the goal itself, and the context in which the goal is set. Goal-setting can be implemented in several ways. *Firstly*, during the sign-up process new users can be asked how much they plan to contribute. As seen in Screenshot 17, Unibet uses this approach, when asking how much new customers plan to bet for each month. This design pattern is a less coercive variation of Drenner et al.'s (2008) study , where the subjects were required to perform a certain amount of participation.

Screenshot 17: Upon registration, Unibet asks new users how much they plan to bet each month.

A s*econd* way to use goal-setting is through surveys, as it is done in this experiment. One caveat with this design pattern is that it does not scale very well: users do not want to answer the same surveys on a regular basis. A *third* way to use goal-setting is through direct outreach to designated community users. Another insight from this experiment is that it is important to consider ongoing feedback channels. According to goal-setting theory, such feedback will result in increased persistence and effort.

Experiment 3 shows that recognizing valuable community members can lead to increased participation. In this experiment, the users are thanked privately. Whether this is more or less effective than public recognition is not examined in this experiment. However, one can speculate that public recognition of some users will have an effect on the users who are not being thanked. It might be positive if it gives other users something to strive for, but negative if it leaves them disappointed.

Recognition can be implemented in various ways. Some online communities offer automated means of recognition. In this way of designing for appreciation, certain behavior leads to certain forms of recognition. Foursquare serves as an example of this practice, as *a priori* defined check-in behavior unlocks badges.

Member recognition can also be accomplished through peer to peer recognition, where users can thank each other manually. Several big online communities use this design pattern. An example is the aforementioned Barnstar system that Wikipedia created to recognize members. A third way to integrate gratitude into design is through expert recognition. Moz has implemented a feature in their Q&A, where staff can endorse an answer written by a user. The result is editorial approval of both the answer and the author.

Experiment 4 shows that first impressions and priming matter. The way a community describes itself around the web affects subsequent usage of the community. The wording can either affect the subjects' subsequent behavior or attracts certain types of subjects, displaying a certain behavior. An example is the

crowdsourcing platform, Kickstarter,[273] where new users are urged to help fund
and follow projects as soon as they enter the site.

In order to utilize the finding from experiment 4, open government communities
ought to test different kinds of community descriptions and see which yielded
the most desirable results in terms of engagement and participation. The de-
scription that generates the most desirable result ought to be used around the
web, e.g. in advertisements, on social media pages, etc. It should also be visible
on the community itself, e.g. as a tagline.

11.2.4.1. Limitations of Design Implications

The design suggestions above suffer from two limitations, which largely is a re-
sult of the American approach applied here: the suggestions are too focused on
problem-solving and rely on a simple quantitative dependent variable.

First, as described in chapter 1, HCI has always been a problem-driven discipline,
aiming to evaluate and/or improve design (Dourish 2006). The research ques-
tion asked in this dissertation is concerned with better design of open govern-
ment communities. One outcome of the dissertation is the design suggestions in
the previous sections.

---

[273] http://kickstarter.com

There is a tendency to use HCI research to solve specific design problems and challenges, rather than trying to understand broader societal issues (Carroll 2003). Researching to improve the design is a very limited way of doing research, and it constitutes one of the limitations of the current HCI field (Dourish 2006).[274] Instead of conforming and focus on better design, one could follow the Scandinavian IS tradition and argue that academic HCI research ought to be more concerned with other social outcomes (Spinuzzi; 2002; Dourish 2006). This dissertation is subject to this critique, as it aims to increase participation to K10 without being overly concerned with the broader issues of open government. This is a typical trait of the "hard" American research approach used here. As described in Chapter 5, an alternative is the "soft" Scandinavian approach, which applies a critical and reflective approach to IS research. Although it is outside the scope of this dissertation, examining open government through reflective and critical perspectives is important and necessary.

Secondly, as described earlier in this chapter, a simple quantitative dependent variable is used. This research does not examine if the social psychology inspired design patterns change more than the quantity. This is a severe limitation, and it is further discussed in section 11.5.

11.3. Theoretical Insights

The experiments show that design inspired by social psychology theory can make an impact. The theory constructs used in this dissertation are thus able to help answer the research question: *Can social psychology theory help increase participation in open government communities?*

In this dissertation, four social psychology theories are selected and operationalized into experiments. The four theories are social comparison theory, goal-setting theory, self-efficacy theory, and social identity theory. The selection and operationalization process comes with several challenges. *Firstly*, it is challeng-

---

[274] Dourish (2006) argues that the focus on implications for design "creates a status hierarchy in which engineering demands tend to override social ones" (Dourish 2006:533).

ing to select one of the many middle-range theories that seek to explain similar phenomena. Many different theories can be used on the same subject matter. In this dissertation, level of abstraction match and case fit is used as criteria for theory selection. It would, however, be fruitful to develop theory selection criteria further.

Secondly, the theory operationalizing also comes with challenges. As the purpose is improvement of practice rather than theory evaluation, the operationalization can be less rigorous. It is not necessary to apply multiple constructs of each theory, nor is it necessary to measure the strength of the intervention on standardized scales. What is necessary is the understanding of how a certain design feature affects behavior in practice. This entails that one must choose among the many different ways to operationalize theory into practice.

Although the use of social psychology theory in the open government community domain comes with challenges, it is also fruitful. The theories can serve as an imperfect catalogue or guide to social design. Using theories help avoid previous mistakes, be more explicit about hypothesis, assumptions, and goals, and ensure that theories have an impact in practice. By leveraging the fundamental knowledge about human behavior, it becomes more attainable to design sustainable open government communities. This knowledge also facilitates the design of communities that take human behavior into consideration. After all, it is easier to change community design than people. An important finding in this dissertation is the possibility of changing participation to open government communities by integrating social psychology theory into design. However, this is only possible to a certain degree.

The subjects from the K10 open government community do not react uniformly to the interventions. This entails that some statistical hypotheses are rejected while others are verified. Thus, the insights from this dissertation do not offer any quick fixes to the participation problem. It is a complex issue and the applied social psychology theories are only able to account for a part of it.

Experiment 1 also shows that social psychology inspired design can have unintended consequences. Translating social psychology theory into practice is, thus, not necessarily a recipe for success. Using this theory body to guide design is only advisable insofar it is applied after proper experimentation and data analysis.

11.4. Methodological Insights

In the following sections three aspects of the methodology are described: the experimental designs, the experimental proxies, and the lean experimentation framework.

11.4.1. Experimental Designs

All four experiments of this dissertation are true experiments and rely on three different experimental designs. Two of the experiments use split-plot design, one experiment is conducted through a within-subject design, and one experiment is conducted through a between-subject design.

The different designs have different comparative advantages and disadvantages. The research question and case often reduce the number of possible designs. All three experimental designs succeed in yielding interesting insights, while maintaining a decent internal validity. The one that caused the biggest problem is the split-plot design used in experiment 2. Issues emerged as a significant change in one of the control groups was detected. The split-plot design is, on the other hand, successfully applied in experiment 1.

As described above, all four experiments yield at least some significant results. On a high level, this suggests that the research design is a good mediator between the theories and the subject matter, although the results vary in terms of significance and effect sizes.

11.4.2. Experimental Proxies

In addition to the three research designs used across the experiments, four different proxies are applied in order to conduct the experimental interventions: email, internal community message, survey, and advertisement. Each of these proxies manages to yield significant effects among at least some subject segments. This suggests that the four proxies have the potential to be further applied to open government community research.

Email is the only proxy that is used more than once. It is used independently in experiment 1, and used along with internal community messages in experiment 3. In experiment 1, email as a proxy has a severe drawback as users got upset after receiving emails with social comparison information. The dissatisfaction might be caused by the emails as a medium, by the content of the email, or by the type of users who received them. As described in Chapter 3, most K10 users suffer from severe illnesses that might affect them in this regard.

Experiment 3 shows that K10 users are significantly more likely to open a message sent via the internal community message system than an email. The external validity of this significant finding is unknown. This means that email is a suboptimal mean of content in this case, and that the second proxy, internal community messages, is a better proxy and delivery methodology than email.

Survey probably constitutes the weakest of the four proxies. Although it is fairly easy to ensure random subject assignment among the different conditions, it is difficult to get a representative sample due to self-selection problems. Answering a survey is a much more time consuming task than clicking on an advertisement or opening an email. This potentially skews the sample toward certain active user types. That being said, the survey makes interaction easy, and it did succeed in making some subjects commit to higher level of participation. A commitment of this character increases their post-intervention participation rate. Another quality of surveys is the qualitative depth and insight, they can provide if qualitative responses are given. The added details can potentially make it easier to evaluate theoretical claims.

The fourth proxy, search engine advertisement, is probably the fastest and easiest of the four. It comes with a user-friendly interface and can be started and modified quickly. Google AdWords is, in other words, a great tool for quick iterations. However, advertisement is also the proxy that comes with the most severe limitations. As it is very complicated to systematically expose the same subject to multiple advertisements over time, between-subject designs are most feasible. Furthermore, Google AdWords offers little information about the users. It is therefore difficult to know how the subjects in the various groups differ, and how representative of the population a sample is. Finally, it is time consuming to segment the data to attain individual data points. Experiment 4 has 453 subjects, making the task feasible. But in cases of 4,000 or 40,000 subjects, the segmentation is virtually impossible.

One shortcoming of the research conducted in this dissertation is that only the proxies are examined. To uncover the deign patterns full potential, it is necessary to also implement the changes in the community to see if the findings from the proxy carry over to full implementation. On the other hand, some of the design patterns examined in this dissertation can be permanently implemented through the proxy. For example, participatory information can be sent out regularly through email, as done on experiment 1.

Although each of the proxies hold different advantages and disadvantages, they are generally helpful in open government research. They manage to yield interesting and valid insights into different social psychology based design patterns, which is the purpose of the experimental proxies in this dissertation. The main challenge of the experimental methodology applied is the constraints in terms of temporal and external validity. To overcome these challenges an experimental framework for research and practice is suggested: lean experimentation.

11.4.3. Lean Experimentation

Lean Experiment is a methodological framework for early stage field experiments testing theories or assumptions on real users before developing and implementing a given design. It does not seek to replace existing development processes such as SCRUM, Prince2, etc., but should be seen as primer to these development processes.
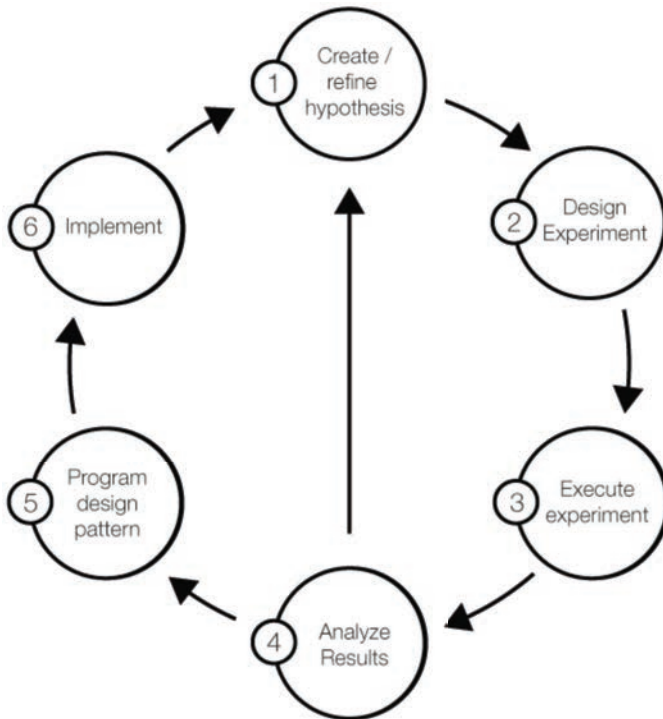
Lean experimentation goes through the following six steps: (1) select theory or assumption; (2) design the experiment (often through proxy); (3) execute the experiment; (4) evaluate results - good ideas can be turned into code while bad ideas can be discarded; (5) turn idea into code; (6) implement the design.

The goal of lean experimentation is to generate insights and knowledge about how a feature or product will work before programming it. Thus, the more insights and knowledge that can be generated in little time and at lower costs, the better the experiment is.

The advantages of lean experimentation is an ability to examine causation, test competitive designs, rely on observed rather than self-reported behavior, influence decisions, economic efficiency, and flexible data output. The disadvantages include fit between task and experimentation, limited to available metrics, confuses users, incrementalism, subject recruitment, lack of qualitative depth, and measures short-term effects.

The framework is based on the findings and process from the dissertation, and has yet to be verified in broader settings.

11.5. Further Research

Based on the substantive and methodological findings in this dissertation, there are four areas in which further research seems particular fruitful: measuring participation quality, creating more complex models, replicating existing findings, and testing and developing the lean experimentation framework.

11.5.1. Measuring Participation Quality

As noted in Chapter 2, the majority of online communities suffer from under-contributions. This problem is certainly present in open government communities, and it is even more salient in small countries where the number of potential contributors is small to begin with. The first step to overcome this participation challenge is to increase the volume in order to attain critical mass. Therefore, this dissertation is primarily concerned with examining participation quantity. It should however be noted that this failure to examine the quality of the participation is a severe limitation of the work. A logical next step is to investigate the quality of such participation, which can be done in at least two ways in future research.

The *first* possibility is to combine the quantitative experimental research with qualitative methodologies such as discourse analysis. This would provide insights into how the discourse changes when social psychology inspired design patterns are implemented. This is especially important in future open government communities in order to assess if more participation also leads to more useful participation. The combination of different methodologies is especially crucial when applying a collaborative democracy perspective.

The *second* way to overcome the limitation of the crude quantitative variable is through improved natural language processing (e.g. Muhlberger & Stromer-Galley 2009) and machine learning (Hsu et al. 2009). As described in Chapter 2, the following metrics have been used to quantitatively assess content quality: text length, characters, question and exclamation mark frequency, capital word frequency, number of references and links in the article, spelling error frequency, readability indices, number of edits, HTML markup, user profile data, user ratings of contributions, historical data, and number of links pointing to the article. But these metrics are not sufficiently developed to be reliably at this point. However, it is expected they will be in the future, thereby giving the opportunity to add qualitative depth to quantitative experimentation.

11.5.2. More Complex Models

In order to reduce the likelihood of errors, the four experiments conducted and reported in this experiment are quite simple. This is problematic because "the level of complexity and indeterminacy in human culture nullifies any illusion of control one might entertain [in experiments]" (Boellstorff et al. 2012: 33). Over two decades ago, Carroll & Kellogg (1989) argued against an overly simplistic approach to HCI: "Articulating the system of theories inherent in HCI designs is not something that is commonly done, but it could be" (Carroll & Kellogg 1989: 14). The simplified research designs used in this dissertation might overlook important interaction effects. Such effects are easier to discover through more

complex models, although it is doubtful if there will ever be a model that includes all measures of social behavior.

On the other hand, several researchers do warn against the pursuit of complex models. It is a significant problem with complex models that they cannot realistically be conducted in practice (Thomke 2001; Sundukovskiy 2010).[275] For example, one model with 10 factors each comprising 3 levels, will result in $3^{10} = 59,049$ variations. And 10 factors is probably only enough to account for merely a fraction of the factors influencing participation in online communities. Thomke (2001) notes that it is impossible to get all information one desired, both on pragmatic and economical grounds. Finally, Kohavi et al. (2009) stress that interaction effects in web experimentation is less frequent than often assumed. This means that the need for comprehensive models that examine interactivity might be smaller than most suspect.

In sum, it is expensive and difficult to gain more complete understandings of complex and interdependent relationships in online communities (Shani & Gunawardana 2011). But this does not mean that more complex models should not be pursued.

11.5.3. Replication

Research replicating existing studies are not valued in the HCI community (Greenberg & Buxton 2008). Consequently, only little of such research is conducted. This is problematic because the validity, scope, and limitations of studies remain largely unknown. The lean experimentation framework makes experiments more efficient and, thus, more attainable. This increases the accessibility of experimentation, and thus hopefully motivates more replicative studies.

---

[275] Similarly, Preece and Schneiderman discourage the use of grand theories: "A unified theory that could be applied across a range of applications and social contexts would be useful, but such a theory is elusive and may not be possible." (Preece and Schneiderman 2009: 15).

11.5.4. Multi Method Research Corpus

As described in Chapter 1 and 5, a plethora of methodologies are used in the IS field, including experiments, statistical analysis, surveys, diaries, case study, interviews, focus groups, ethnography, usability testing, and participatory design. The research in this dissertation relies exclusively on one of them. This approach has caused certain limitations in the research, e.g. lack of thick descriptions of the community and its members.

Boellstorff et al. (2012) argue that quantitative experimental research needs to be part of a broader research methods relying on many methodologies. In order to gain a deeper understanding of open government communities, it will be fruitful to apply qualitative methods such as long ethnographic studies. For example, if one follows the research methodology of Nardi (2010), and spends several years as an active participant in an open government community, one will undoubtedly get a very different perspective, than the one presented in this dissertation. The argument is not that one perspective is inherently superior, but that the diverse methods can interact in a very productive way when aggregated in a research corpus (Boellstorff et al. 2012).

11.5.5. Testing and Developing Lean Experimentation

One of the main contributions of this dissertation is the lean experimentation framework. Lean experimentation builds on Davenport (2009)'s well-tested *Test and Learn wheel*. The contribution of lean experimentation is a reconstruction of the steps in the framework by incorporating an intense emphasis on collecting knowledge early in the development process. This is accomplished by creating efficient experiments that rely on proxies.

Although the framework is implicitly used in the four experiments conducted in this dissertation, more work on the framework is necessary. It needs to be applied to other cases and contexts in order to test the validity of the experiments conducted here and help identify the boundaries of the framework.

11.6. Policy Recommendations

Open government is still an emerging field. As described in Chapter 2 most of the literature is enthusiastic and optimistic in terms of what open government can accomplish. The research conducted in this thesis does not examine if this optimism is grounded in reality. The aim is rather to enhance the likelihood of successful implementation. The following policy recommendations, which are based on the findings from this dissertation, are therefore assuming open government is desirable. In order to create the ideal settings for successful open government implementation, policy makers ought to clarify the following three issues:

Define the aim of open government. Two different open government purposes are examined, compared and discussed throughout this dissertation: collaborative democracy and deliberative democracy. Especially experiment 1 shows that it can be hard to find design patterns that both can increase participation and make it more democratic. Policy makers therefore need to define the primary aim of different open government initiatives: is it to improve output and collaboration, or is it to improve deliberation and democratic processes? It is probably not a question about either/or in general. Different open government projects can have different aims but they need to be clearly defined.

Prioritize among government projects. As demonstrated in three of the experiments in this dissertation, it is possible to increase participation in open government communities. But increased participation in one community comes with opportunity costs. Citizens, companies, and NGOs have a finite amount of time and resources to participate online and offline. So policy makers need to define where participation is important. Is it online or offline? Is it in politics or administration? Is it in health care or patent applications? Once questions like these have been answered, policy makers should aim at optimizing participation to these chosen communities. This will also help mitigate the problem that participation is scattered across different but overlapping communities.

Encourage experimentation. The results of the four experiments conducted in this dissertation are not as clear as one could have hoped. But as described in Chapter 10, many successful companies expect between 50 % - 90 % of their experiments to fail. The overarching lesson of these experiments is that it is important to conduct experiments. This is the cheapest and fastest way to reliably identify design patterns that achieve one's goal. As it is too hard to guess the effect of ideas and design patterns, it is important to encourage experimentation. By leveraging lean experimentation, it is often not necessary to enact in any coding, thereby making it possible for non-developers to conduct experiments. So the real task for policy makers is to encourage and educate relevant civil servants to conduct experiments.

11.7. Chapter Summary

In this dissertation, the following research question is examined: *Can social psychology theory help increase participation in open government communities?* The short answer is that social psychology theory can indeed help guide open government community design using rather simple constructs. Although the results are not as unambiguous as one could hope, theory-based designs make it possible to increase participation and change participation patterns with varying effect-sizes.

One severe caveat of the operationalization of social psychology theory used in this dissertation is that the external validity of the specific findings is unknown. It is therefore not only the social psychology theories *per se* that can help improve the participation in open government communities, but also experimentation which is the underlying methodology of the theories. In online settings, experimentation serves as an efficient mean to examine social psychology theories' effect in any given context.

In order to increase the likelihood of successful open government implementation, different design features ought to be examined through experimentation. This will increase the knowledge about which design elements increases and

decreases participation in different open government contexts. To help facilitate this process, the methodological experimental framework of lean experimentation is proposed.

To develop open government research further, it is necessary to measure participation quality, introduce more complex models, increase replication, build a multi method research corpus and test and develop the lean experimentation framework further.

To enhance the likelihood of successful open government in practice, policy makers ought to define the aim of open government, prioritize different kinds of open government projects, and encourage experimentation.

Bibliography

Abdi, H. Edelman, B, Valentin, D. & Dowling. W. J. (2009). *Experimental Design & Analysis for Psychology,* England: Oxford University Press.

Adair, J.G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, 69(2), pp. 334–345.

Agresti, A. & Finlay, B. (1997). *Statistical Methods for the Social Sciences* 3rd ed., Upper Saddle River, NJ: Prentice Hall.

Ali-Hasan, N.F. (2008). Exploring Social Media Scenarios for the Television, In *Proceedings of International Conference for Weblogs and Social Media*.

Allison, B. (2010). My Data Can't Tell You That, In D. Lathrop & L. Ruma (eds.), *Open Government: Collaboration, Transparency, and Participation in Practice*, Sebastopol, CA: O'Reilly Media, Inc., pp. 257-265.

American Psychological Association (1992). Ethical principles of psychologists and code of conduct, *American Psychologist*, Vol. 47(12), pp. 1597-1611.

Andersen, K. & Henriksen, H. (2006). E-government maturity models: Extension of the Layne and Lee model. *Government Information Quarterly*, 23(2), pp. 236–248.

Andersen, K. (2011). Kommunal borgerservice anno 2011. Frekvens og omkostninger for anvendelse af fysiske og digitale kanaler. In *CAICT Working Papers*. Copenhagen, Denmark: Copenhagen Business School.

Anderson, C. (2006). *The Long Tail*, New York: Hyperion.

Anderson, E.T. & Simester, D. (2011). A Step-By-Step Guide To Smart Business Experiments. *Harvard Business Review*, 89(3), pp. 98–105.

Anonymous blogger (2012). K10-kodeks: Skid på alt og alle - og vær ikke flov! *180grader*. Retrieved 25/8/2012 from http://www.180grader.dk/Krimi/k10-kodeks-skid-p-alt-og-alle-og-vaer-ikke-flov.

Anttiroiko, A.V. (2010). Innovation in Democratic E-Governance. *International Journal of Electronic Government Research*, 6(2), pp. 18–36.

Arazy, O., Kumar, N. & Shapira, B. (2010). A Theory-Driven Design Framework for Social Recommender Systems. *Journal of the Association for Information Systems*, 11(9), pp. 455–490.

Armstrong, C. (2010). Emergent Democracy. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly Media.

Aronson, E. & Carlsmith, J.M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *The Handbook of Social Psychology*, 2, pp. 1–79. Reading, MA, Addison-Wesley

Asch, S.E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70(9), pp. 1–70.

Aschoff, F. R., Schaer, V. & Schwabe, G. (2011). Where should I send my post?: the concept of discourse quality in online forums and its dependency on member-ship size. In *Proceedings of the 5th International Conference on Communities and Technologies*, pp. 69–78.

Ashforth, B.E. & Mael, F. (1989). Social identity theory and the organization. *Academy of management review*, 14 (1), pp. 20–39.

Ba, S., Whinston, A. B., & Zhang, H. (1999). Building trust in the electronic market through an economic incentive mechanism. In *Proceedings of the 20th interna-tional conference on Information Systems,* pp. 208-213.

Ba, S., Whinston, A. B. & Zhang, H. (2003). Building trust in online auction mar-kets through an economic incentive mechanism. *Decision Support Systems*, 35(3), pp. 273–286.

Bainbridge, W.S. (2004). *Berkshire encyclopedia of human-computer interaction*. Great Barrington, Massachusetts: Berkshire Publishing Group

Baldini, M. & Federici, M. (2011). Ethnic discrimination in the Italian rental hous-ing market. *Journal of Housing Economics*, 20(1), pp. 1–14.

Bandura, A. (1962). Social learning through imitation. In Marshall R. Jones (Ed). *Nebraska Symposium on Motivation*, Oxford, England: University Nebraska Press, pp. 211-274.

Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological review*, 84(2), pp. 191–215.

Bandura, A. (1977b). *Social learning theory*, Englewood Cliffs, NJ: Prentice Hall

Bandura, A. (1990). Perceived self-efficacy in the exercise of control over AIDS infection. *Evaluation and program planning*, 13(1), pp. 9–17.

Banks, M. (2006). Towards a continuum of scholarship: The eventual collapse of the distinction between grey and non-grey literature. *Publishing Research Quar-terly*, 22(1), pp. 4–11.

Bansler, J. (1988). Systems Development in Scandinavia: Three Theoretical Schools. *Information Technology & People* 4 (2) pp. 117–133.

Baregheh, A., Rowley, J. & Sambrook, S. (2009). Towards a multidisciplinary defi-nition of innovation. *Management decision*, 47(8), pp. 1323–1339.

Bargh, J. A., Gollwitzer, P. M., & Oettingen, G. (2010). Motivation. In S. Fiske, D. T. Gilbert, & G. Lindzay (Eds.), *Handbook of social psychology* 5th ed., pp. 268-316. New York: Wiley.

Barrett, S. (2009). *Anthropology: A Student's Guide to Theory and Method* 2ⁿᵈ Edition ed., Toronto, Canada: University of Toronto Press.

Baskerville, R.L. & Myers, M.D. (2002). Information systems as a reference discipline. *MIS Quarterly*, pp. 1–14.

Bass, G.D. & Moulton, S. (2010). Bringing The Web 2.0 Revolution to Government. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Baumeister, R.F. & Leary, M.R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin*, 117(3), pp. 497-529.

Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., & Kraut, R. (2004). Using Social Psychology to Motivate Contributions to Online Communities. In *Proceedings of the ACM conference on Computer supported cooperative work* (pp. 212-221).

Beenen, G., Ling, K., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P. & Kraut, R. (2005), Using Social Psychology to Motivate Contributions to Online Communities. *Journal of Computer-Mediated Communication*, 10(4).

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, CT: Yale University Press.

Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage*, 49(3), pp. 2687-2696.

Bertot, J.C., Jaeger, P.T. & Hansen, D. (2011). The impact of polices on government social media usage: Issues, challenges, and recommendations. *Government Information Quarterly,* 29(1), pp. 30-40.

Billig, M. & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, 3(1), pp. 27–52.

Birkinshaw, P. (2010). *Freedom of Information: the Law, the Practice and the Ideal.* Cambridge, England: Cambridge University Press.

Bjerknes, G., Ehn, P. & Kyng, M. (1987). *Computers and democracy - a Scandinavian challenge*, Aldershot, England: Gower.

Bjerknes, G. Bratteteig, T. (1995) User Participation and Democracy: A Discussion of Scandinavian Research on System Development. *Scandinavian Journal of Information Systems,* 7, pp. 73–73.

Blanz, M., Mummendey, A., Mielke, R., Klink, A., (1998). Responding to negative social identity: A taxonomy of identity management strategies. *European Journal of Social Psychology*, 28(5), pp. 697–729.

Blumenstock, J.E. (2008). Size matters: Word Count as a Measure of Quality on Wikipedia. In *Proceedings of the 17th international conference on World Wide Web*, pp. 1095–1096.

Bødker, S. & Grønbæk, K. (1992). Design in action: From prototyping by demonstration to cooperative prototyping. In *Design at work*. Hillsdale, NJ: Erlbaum Associates.

Boellstorff, T., Nardi B., Taylor, T.L., Pearce C. (2012). *Ethnography and Virtual Worlds: A Handbook of Method*, Princeton N.J, Princeton University Press.

Bordens, K. & Abbott, B.B. (2010). *Research Design and Methods: A Process Approach* 8th ed., San Francisco, CA: McGraw-Hill.

Bosch, M., Carnero, M.A. & Farre, L. (2010). Information and discrimination in the rental housing market: Evidence from a field experiment. *Regional Science and Urban Economics*, 40(1), pp. 11–19.

Bovaird, T. (2007). Beyond engagement and participation: User and community coproduction of public services. *Public Administration Review*, 67(5), pp. 846–860.

Brabham, D.C. (2008). Crowdsourcing as a Model for Problem Solving. Convergence. *The International Journal of Research into New Media Technologies* 14 (1): 75.

Brajer, V. & Gill, A. (2010). Yakity-Yak: Who Talks Back? An Email Experiment. *Social Science Quarterly*, 91(4), pp. 1007–1024.

Brewer, M.B. & Kramer, R.M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of personality and social psychology*, 50(3), p. 543.

Brewer, M.B. (1993). Social identity, distinctiveness, and in-group homogeneity. *Social cognition*, 11(1), pp. 150–164.

Brincker, B. & Gundelach, P. (2010). A la Carte Community: Identity and values in the open source software project TYPO3. *Scandinavian Journal of Information Systems*, 22(1), pp. 27–44.

Broota, K.D. (1989). *Experimental Design in Behavioural Research*, New Delhi, India: New Age International.

Brown, M.M. & Brudney, J.L. (2001). Achieving advanced electronic government services: An examination of obstacles and implications from an international perspective. In *Proceedings from National Public Management Research Conference*.

Brown, R. (2000). Social identity theory: past achievements, current problems and future challenges. *European Journal of Social Psychology*, 30(6), pp. 745–778.

Brudney, J. L. & England, R.E. (1983). Toward a definition of the coproduction concept. *Public Administration Review*, pp. 59–65.

Brudney, Jeffrey L. & Duncombe, W.D. (1992). An Economic Evaluation of Paid, Volunteer, and Mixed Staffing Options for Public Services. *Public Administration Review*, 52(5), pp. 474–481.

Burke, M., Marlow, C. & Lento, T. (2009). Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the SIGCHI international conference on Human factors in computing systems*, pp. 945–954.

Burton, M. (2010). A Peace Corps for Programmers. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Butler, B.S. (1999). The dynamics of cyberspace: Examining and modeling online social structure. *Unpublished Dissertation* Pittsburg, PA: Carnegie Mellon University*.*

Butler, D.M. & Broockman, D.E. (2011). Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators. *American Journal of Political Science*, 55(3), pp. 463–477.

Cámara, J. & Kobsa, A. (2009). Facilitating Controlled Tests of Website Design Changes Using Aspect-Oriented Software Development and Software Product Lines. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems*, pp. 116–135.

Cammaerts, B. (2008). Critiques on the Participatory Potentials of Web 2.0. *Communication, Culture & Critique*, 1(4), pp. 358–377.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). *Experimental and Quasi-experimental Designs for Research,* Boston, MA: Houghton Mifflin.

Carmagnola, F., Vernero, F., & Grillo, P. (2009). Sonars: A social networks-based algorithm for social recommender systems. In *User Modeling, Adaptation, and Personalization*, Springer, Berlin Heidelberg, pp. 223-234.

Carrington, P.J., Scott, J. & Wasserman, S. (2005). *Models and methods in social network analysis*, Cambridge, England: Cambridge University Press.

Carroll, J. M., & Kellogg, W. A. (1989). Artifact as theory-nexus: hermeneutics meets theory-based design. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 7-14).

Carroll, J. M. (1997). Human-computer interaction: psychology as a science of design. *Annual review of psychology*, 48(1), pp. 61-83.

Carroll, J. M. (2003). *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, San Francisco, CA: Morgan Kaufmann.

Casaló, L.V., Flavián, C. & Guinalíu, M. (2010). Determinants of the intention to participate in firm-hosted online travel communities and effects on consumer behavioral intentions. *Tourism Management*, 31(6), pp. 898–911.

Castells, M. (2011). *The rise of the network society: The information age: Economy, society, and culture*, Malden: Wiley-Blackwell.

Cha, M. Kwak, H., Rodriguez, P., Ahn, Y. Y., & Moon, S. (2007). I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pp. 1–14.

Chan, C. (2011). Using Online Advertising to Increase the Impact of a Library Facebook Page. *Library Management*, 32(4/5), pp. 361–370.

Chan, K.W. & Li, S.Y. (2008). Understanding consumer-to-consumer interactions in virtual communities: The salience of reciprocity. *Journal of Business Research*, 63(9), pp. 1033–1040.

Cheliotis, G. & Yew, J. (2009). An analysis of the social structure of remix culture. In *Proceedings of the fourth international conference on Communities and technologies*, pp. 165–174.

Chen, Y., Harper, Y. C. F. M., Konstan, J., & Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on movielens. *The American Economic Review*, 100(4), pp. 1358–1398.

Chen, Y., Ho, T. & Kim, Y. (2010). Knowledge market design: A field experiment at Google Answers. *Journal of Public Economic Theory*, 12(4), pp. 641–664.

Chou, S.W. (2010). Why do members contribute knowledge to online communities? *Online Information Review*, 34(6), pp. 829–854.

Chou, S. W., Min, H. T., Chang, Y. C., & Lin, C. T. (2010). Understanding continuance intention of knowledge creation using extended expectation–confirmation theory: an empirical study of Taiwan and China online communities. *Behaviour & Information Technology*, 29(6), pp. 557–570.

Christensen, H.S. (2011). Political activities on the Internet: Slacktivism or political participation by other means? *First Monday*, 16(2), p. 5.

Chryssochoou, X. (2010). Conversation with Claude Flament Executive Committee 1969-1975 and President 1972-1975. *EBSP*, 22(2).

Chua, C.K., Leong, K.F. & Lim, C.S. (2010). *Rapid Prototyping: Principles and Applications*, Singapore: World Scientific Publishing Company.

Chun, S. A., Shulman, S., Sandoval, R., & Hovy, E. (2010). Government 2.0: Making

connections between citizens, data and government. *Information Polity*, 15(1), pp. 1-9.

Cialdini, R.B. (2007). *Influence: The Psychology of Persuasion* Revised. New York, NY: Harper.

Ciffolilli, A. (2007). Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia, *First Monday* 12.

Citron, D. K. (2009). Fulfilling Government 2.0's Promise with Robust Privacy Protections. *George Washington Law Review*, 78, pp. 822.

Clemmensen, T. (2006). Whatever happened to the psychology of human-computer interaction?: A biography of the life of a psychological framework within a HCI journal. *Information Technology & People* (19)(2), pp. 121–151.

Coglianese, C. (2009). The Transparency President? The Obama Administration and Open Government. *Governance*, 22(4), pp. 529–544.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Comscore (2009). Firefox vs Internet Explorer. *Comscore* Retrieved 15/6/2012 from http://blog.comscore.com/2007/04/firefox_vs_internet_explorer_1.html

Conn, V.S., Valentine, J.C., Cooper, H.M., Rantz, M.J. (2003). Grey literature in meta-analyses. *Nursing Research*, 52(4), p. 256.

Cooper, M. (2012). Republican Platform Takes Turn to Right. *The New York Times* Retrieved 29/9/2012 from http://www.nytimes.com/2012/08/29/us/politics/republican-platform-takes-turn-to-right.html

Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. (2003). Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 585–592

Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. (2006). Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1037–1046.

Crook, T., Frasca, B., Kohavi, R., & Longbotham, R. (2009). Seven pitfalls to avoid when running controlled experiments on the web. In *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1105–1114.

Croll, A & Yoskovitz, B. (2013). *Lean Analytics*. Sebastopol, CA: O'Reilly

Cullen, R. & Morse, S. (2011). Who's Contributing: Do Personality Traits Influence the Level and Type of Participation in Online Communities. In *IEEE*, pp. 1–11.

Curioso, A., Bradford, R. & Galbraith, P. (2010). *Expert PHP and MySQL*, Indianapolis, IN: John Wiley & Sons.

Dabbish, L., Kraut, R. E., Farzan, R., & Postmes, T. (2012). Fresh faces in the crowd: turnover, identity, and commitment in online groups. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 245–248.

Danielsson, K., Naghsh A.M., Gumm D., Warr A. (2008). Distributed participatory design. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 3953–3956.

Datatilsynet (2000). *Persondataloven*. Retrieved 30/1/2013 from http://www.datatilsynet.dk/fileadmin/user_upload/dokumenter/Persondatalov spjece/Persondatalovspjece.htm.

Datatilsynet (2001). *The Act on Processing of Personal Data*. Retrieved 30/1/2013 from http://www.datatilsynet.dk/english/the-act-on-processing-of-personal-data/read-the-act-on-processing-of-personal-data/compiled-version-of-the-act-on-processing-of-personal-data/

Davenport, T.H. (2009). How to design smart business experiments. *Harvard Business Review*, 87(2), pp. 68–76.

Davis, M.E. & Phillips, J.A. (2009). *Learning PHP & MySQL: Step-by-Step Guide to Creating Database-Driven Web Sites*, 2nd Edition, Sebastopol, CA: O'Reilly Media, Inc.

Dawes, M.A. (2010). True Experimental Designs. In N. J. Salkind (ed.), *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage, pp. 1543–1546.

Delanty, G. (2009). *Community,* 2nd edition, London, England: Routledge.

Dellarocas, C. (2001). Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. In *Proceedings of the ACM conference on Electronic Commerce*, pp. 171–179.

DeLone, W.H. & McLean, E.R. (1992). Information systems success: The quest for the dependent variable. *Information systems research*, 3(1), pp. 60–95.

Delone, W.H. & McLean, E.R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of management information systems* (4), pp. 9–30.

Digitaliser.dk (2010). Debatmodul på Borger.dk - Digitaliser.dk. *digitaliser.dk*. Retrieved 30/1/2013 from http://digitaliser.dk/resource/556748.

Dourish, P. (2006). Implications for design. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 541–550.

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., & Klemmer, S. R. (2010). Parallel prototyping leads to better design results, more divergence, and

increased self-efficacy. *ACM Transactions on Computer-Human Interaction*, 17(4), p. 18.

Downs, A. (1957). An economic theory of political action in a democracy. *The Journal of Political Economy*, 65(2), pp. 135–150.

Drapeau, M. (2010). Two Way Street: Government with the People. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Drenner, S., Sen, S. & Terveen, L. (2008). Crafting the initial user experience to achieve community goals. In *Proceedings of the ACM conference on Recommender systems*, pp. 187–194.

Dunning, D. ed. (2010). *Social Motivation* 1st ed., Philadelphia: Psychology Press.

Dutton, W.H. (1992). Political science research on teledemocracy. *Social Science Computer Review*, 10(4), pp. 505–522.

Easley, D. & Kleinberg, J. (2012). *Networks, Crowds, and Markets: Reasoning about a highly connected world*. Wiley Online Library.

Eastwick, P.W. & Gardner, W.L. (2009). Is it a game? Evidence for social influence in the virtual world. *Social Influence*, 4(1), p. 18.

Eaves, D. (2010). After the Collapse: Open Government and the Future of Civil Service. In *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Ehn, P. (1993). Scandinavian design: On participation and skill. *Participatory design: Principles and practices*, Hillsdale, NJ: Lawrence Erlbaum, pp. 41–77.

Ellis, P.D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*, Cambridge, England: Cambridge University Press.

Ellison, G. (2002). The Slowdown of the Economics Publishing Process. *Journal of Political Economy*, 110(5), pp. 947–993.

Ellison, G. (2011). Is Peer Review In Decline? *Economic Inquiry*, 49(3), pp. 635–657.

Erikson, E. H. (1994). Identity: Youth and crisis (Vol. 7). New York, NY: Norton.

Eriksson, M. (2010). Using Social Science in Design. In P. Daras et al. (eds). *User Centric Media*. In *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Berlin: Springer, pp. 361–365.

Farina, C. R., Newhart, M. J., Cardie, C., & Cosley, D. (2010). Rulemaking 2.0. *University of Miami Law Review*, 65, pp. 395.

Farina, C. R., Miller, P., Newhart, M. J., & Cardie, C. (2011). Rulemaking in 140 characters or less: social networking and public participation in rulemaking. *Pace Law Review*, 31, pp. 382.

Farnham, S.D., Brown, P.T. & Schwartz, J.L.K. (2009). Leveraging social software for social networking and community development at events. In *Proceedings of the fourth international conference on Communities and technologies,* pp. 235–244.

Farzan, R., Dabbish, L. A., Kraut, R. E., & Postmes, T. (2011). Increasing commitment to online communities by designing for social presence. In *Proceedings of the ACM conference on Computer supported cooperative work*, pp. 321–330.

Farzan, R., Kraut, R., Pal, A., & Konstan, J. (2012). Socializing volunteers in an online community: a field experiment. In *Proceedings of the ACM conference on Computer Supported Cooperative Work*, pp. 325–334.

Fayard, A. L. & DeSanctis, G. (2008). Kiosks, Clubs and Neighborhoods: The Language Games of Online Forums. *Journal of the Association for Information Systems*, 9(10), pp. 677–705.

Fedorowicz, J. & Dias, M.A. (2010). A decade of design in digital government research. *Government Information Quarterly*, 27(1), pp. 1–8.

FedRAMP, (2013). FedRAMP. *Gsa.gov*. Retrieved 30/1/2013 from http://www.gsa.gov/portal/category/102371.

Fernback, J. (1999). There is a there there: Notes toward a definition of cybercommunity. In S. Jones (ed) *Doing Internet research: Critical issues and methods for examining the Net*. Thousand Oaks, CA: SAGE, pp. 203–220.

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, 7(2), pp. 117-140.

Festinger, L. (1957). *A theory of cognitive dissonance*. Palo Alto, CA: Stanford University Press.

Field, A. (2009). *Discovering statistics using SPSS*. Thousand Oaks, CA: Sage Publications Limited.

Fisher, R.J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, pp. 303–315.

Fitzgerald, B. & Howcroft, D. (1998). Towards dissolution of the IS research debate: from polarization to polarity. *Journal of Information Technology*, 13, pp. 313–326.

Flyvbjerg, B. (2001). *Making Social Science Matter: Why Social Inquiry Fails and How it Can Succeed Again*. Cambridge, England: Cambridge University Press.

Fogg, B.J. & Tseng, H. (1999). The elements of computer credibility. In *Proceed-*

ings of the SIGCHI conference on Human factors in computing systems, pp. 80–87.

Fogg, B.J. (2003). *Persuasive technology : using computers to change what we think and do*, San Francisco, CA: Morgan Kaufmann Publishers.

Foley, J. (2012). 10 Developments Show Government Cloud Maturing *Information Week*. Retrieved 30/1/2013 from http://www.informationweek.com/government/cloud-saas/10-developments-show-government-cloud-ma/240002578

Friis Sørensen, I. (2007). Københavns Kommune 2.0. *Kommunikationsforum* Retrieved 30/1/2013 from http://www.kommunikationsforum.dk/artikler/koebenhavns-kommune-to-nul.

Fung & Weil (2010). Open Government and Open Society. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Gaines, B.J., Kuklinski, J.H. & Quirk, P.J. (2007). The Logic of the Survey Experiment Reexamined. *Political Analysis*, 15(1), pp. 1–20.

Gazan, R. (2011). Redesign as an act of violence: disrupted interaction patterns and the fragmenting of a social Q&A community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2847–2856.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438, pp. 900–901.

Gilovich, T., Keltner, D., & Nisbett, R. E. (2010). *Social psychology* . New York, NY: W. W.

Gläser, J. (2006). *Wissenschaftliche Produktionsgemeinschaften: Die soziale Ordnung der Forschung*: Campus Verlag.

Google (2011). URL Builder - Analytics Help. *Google.* Retrieved 30/1/2013 from http://support.google.com/analytics/bin/answer.py?hl=en&answer=1033867&rd=1.

Google (2012a). Using broad match. *Google.* Retrieved 30/1/2013 from http://support.google.com/adwords/answer/2497828?hl=en.

Google (2012b). Using phrase match. *Google.* Retrieved 30/1/2013 from http://support.google.com/adwords/answer/2497584?hl=en.

Google (2013a). High Bounce Rate. *Google.* Retrieved 30/1/2013 from http://support.google.com/analytics/bin/answer.py?hl=en&answer=1009409.

Google (2013b). Pageviews. *Google.* Retrieved 15/6/2013 from http://support.google.com/analytics/bin/answer.py?hl=en&answer=1009409.

Google (2013c.) Time On Site. *Google.* Retrieved 30/1/2013 from

http://support.google.com/googleanalytics/bin/answer.py?hl=en&answer=991 18.

Gottfredson, D.O.N. & Gottfredson, S.D. (1982). Criminal Justice Behavior. *Criminal Justice and Behavior*, 9(3), p .259.

Gottfredson, S.D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, 33(10), p. 920.

Gouldner, A.W. (1960). The norm of reciprocity: A preliminary statement. *American sociological review*, 25(2), pp. 161–178.

Gravetter, F.J. & Wallnau, L.B. (2012). *Statistics for the Behavioral Sciences,* 9th ed., Belmont, CA: Wadsworth Publishing.

Greenberg, S. & Buxton, B. (2008). Usability evaluation considered harmful (some of the time). In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pp. 111–120.

Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, 30(3), pp. 611–642.

Grimmelikhuijsen, S. (2012). Linking Transparency, Knowledge and Citizen Trust in Government: An Experiment. *International Review of Administrative Sciences*, 78(1), pp. 50–73.

Guerini, M., Strapparava, C. & Stock, O. (2010). Evaluation metrics for persuasive NLP with Google Adwords. *Proceedings of LREC'10*, pp. 3459–3463.

Gusfield, J.R. (1978). *Community: A Critical Response*, New York, NY: Harper & Row.

Habermas, J. (1994). Three Normative Models of Democracy. *Constellations*, 1(1), pp. 1–10.

Hansell, S. (2009). Ideas Online, Yes, but Some Not So Presidential. *The New York Times*. Retrieved 20/1/2013 from http://www.nytimes.com/2009/06/23/technology/internet/23records.html

Haney, C., Banks, C., & Zimbardo, P. (1972). *Interpersonal dynamics in a simulated prison*. USA: Defense Technical Information Center.

Harper, F. M., Li, S. X., Chen, Y., & Konstan, J. (2007). Social comparisons to motivate contributions to an online community. *Persuasive Technology*, pp. 148–159.

Harper, F. M., Frankowski, D., Drenner, S., Ren, Y., Kiesler, S., Terveen, L., & Riedl, J. (2007). Talk amongst yourselves: inviting users to participate in online conversations. In *Proceedings of the conference on Intelligent user interfaces*, pp. 62–71.

Harper, F. M., Raban, D., Rafaeli, S., & Konstan, J. A. (2008). Predictors of answer quality in online Q&A sites. In *Proceeding of the SIGCHI conference on Human factors in computing systems*, pp. 865–874.

Harré, R. & Krausz, M. (1996). *Varieties of relativism*, Cambridge, England: Cambridge University Press.

Harrison, T. M., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., & Pardo, T. (2011). Open government and e-government: democratic challenges from a public value perspective. In *Proceedings of the International Conference on Digital Government Research.*

Heeks, R. & Bailur, S. (2007). Analyzing e-government research: Perspectives, philosophies, theories, methods, and practice. *Government Information Quarterly*, 24(2), pp. 243–265.

Heine, S.J. & Norenzayan, A. (2006). Toward a psychological science for a cultural species. *Perspectives on Psychological Science*, 1(3), pp. 251–269.

Heine, S.J. (2010). Cultural Psychology. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (eds). *Handbook of social psychology*. New York, NY: Wiley.

Held, D. (1997). *Models of Democracy* 2nd ed., Palo Alto, CA: Stanford University Press.

Hewett, B., Card, C., Gasen, M., & Perlman, S. (2009). *ACM SIGCHI Curricula for Human-Computer Interaction* 3rd ed., New York, NY: ACM.

Høgenhaven, T. & Andreasen, L.L. (2011). *Når internettet har magten*, Copenhagen, Denmark: DJØFs Forlag.

Hogg, M.A. & Williams, K. D. (2000). From I to we: Social identity and the collective self. *Group Dynamics: Theory, Research, and Practice*, 4(1), pp. 81–97.

Hogg, M.A. (2006). Social Identity Theory. In *Contemporary social psychological theories*. Palo Alto, CA: Stanford University Press.

Hollis, M. (1994). *The philosophy of social science : an introduction*, Cambridge, England: Cambridge University Press.

Holmlid, S. (2009). Participative, co-operative, emancipatory: From participatory design to service design. In *Nordic Conference on Service Design and Service Innovation.*

Holson, L.M. (2009). Putting a Bolder Face on Google. *The New York Times*. Retrieved 30/1/2013 from http://www.nytimes.com/2009/03/01/business/01marissa.html.

Hood, C. (1995). The "New Public Management" in the 1980s: variations on a theme. *Accounting, organizations and society*, 20(2), pp. 93–109.

Hood, C. (2007). What happens when transparency meets blame-avoidance? *Public Management Review*, 9(2), pp. 191–210.

Hornsey, M.J. (2008). Social Identity Theory and Self-categorization Theory: A Historical Review. *Social and Personality Psychology Compass*, 2(1), pp. 204–222.

Hsu, C. F., Khabiri, E. & Caverlee, J. (2009). Ranking Comments on the Social Web. In *International Conference on Computational Science and Engineering*, pp. 90–97.

Hui, G. & Hayllar, M.R. (2010). Creating Public Value in E-Government: A Public-Private-Citizen Collaboration Framework in Web 2.0. *Australian Journal of Public Administration*, 69, pp. 120–131.

Hurst, M. (2011). Ignore the customer experience, lose a billion dollars (Walmart case study). *Good Experience*. Retrieved 12/6/2012 from http://goodexperience.com/blog/2011/04/ignore-the-customer-e.php

Jackson, N. & Carter, P. (1991). In Defence of Paradigm Incommensurability. *Organization Studies*, 12(1), pp. 109–127.

Jaeger, Paul T. & Bertot, John Carlo (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, 27(4), pp. 371–376.

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28(4), pp. 446–456.

Jeon, G.Y., Kim, Y. M. & Chen, Yan (2010). Re-examining price as a predictor of answer quality in an online q&a site. In *Proceedings of the international conference on Human factors in computing systems*, pp. 325–328.

Jeppesen, L.B. & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science*, 17(1), p. 45.

Joyce, E. & Kraut, R. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11(3), pp. 7 23–747.

"Juul" (2011). Kender du K10 - det ekstremistiske website? *180grader*. Retrieved 20/9/2012 from http://www.180grader.dk/Krimi/kender-du-k10-det-ekstremistiske-website.

Kadlec, T. (2012). *Implementing Responsive Design: Building sites for an anywhere, everywhere web*, Berkeley: New Riders.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American economic review*, 93(5), pp. 1449–1475.

Karau, S.J. & Williams, K. D (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of personality and social psychology*, 65(4), p. 681.

Katz, M. (2011). Why context is important when gathering design feedback: an e-commerce case study. In *Proceedings of the (2011) annual conference extended abstracts on Human factors in computing systems*, pp. 467–482.

Kaushik, A. (2009). *Web analytics 2.0: the art of online accountability & science of customer centricity*, Hertfordshire, UK: Sybex.

Keen, A. (2007). *The cult of the amateur: how today's internet is killing our culture*, New York, NY: Doubleday.

Keller, H.E. & Lee, S. (2003). Ethical Issues Surrounding Human Participants Research Using the Internet. *Ethics & Behavior*, 13(3), pp. 211–219.

Kensing, F. & Blomberg, J. (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work*, 7(3), pp. 167–185.

Keuzenkamp, H.A. & McAleer, M. (1995). Simplicity, Scientific Inference and Econometric Modelling. *The Economic Journal*, 105(428), p. 1.

Kim, B., Barua, Anitesh & Whinston, Andrew B. (2002). Virtual field experiments for a digital economy: a new research methodology for exploring an information economy. *Decision Support Systems*, 32(3), pp. 215–231.

Kimball, M.B. (2011). Mandated state-level open government training programs. *Government Information Quarterly*, 28(4), pp. 474–483.

King, G., Keohane, R.O. & Verba, S. (2001). *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton NJ: Princeton University Press.

Kingsley-Hughes, A. & Kingsley-Hughes, K. (2006). *Building forums with vBulletin creating and maintaining online discussion forums*, Birmingham, UK: Packt.

Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. M. (2009). Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1), pp. 140–181.

Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. M. (2009). Online experimentation at Microsoft. In *Third Workshop on Data Mining Case Studies and Practice Prize*.

Kornfeld, W.A. & Hewitt, C. (1981). *The Scientific Community Metaphor*. Retreived 7/4/2012 from http://dspace.mit.edu/handle/1721.1/5693.

Kozinets, R.V. (2002). The field behind the screen: using netnography for marketing research in online communities. *Journal of Marketing Research*, pp. 61–72.

Kraut, R. (2003). Applying Social Psychological Theory to the Problems of Group Work. In John M. Carroll (ed) *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*. San Francisco, CA: Morgan Kaufmann, pp. 325-356.

Kraut, R. & Resnick, P. (2012). *Building Successful Online Communities: Evidence-*

Based Social Design, Cambridge, MA: MIT Press.

Kriplean, T., Beschastnikh, I. & McDonald, D.W. (2008). Articulations of wiki-work: uncovering valued work in wikipedia through barnstars. In *Proceedings of the ACM conference on Computer supported cooperative work*, pp. 47–56.

Krumm, J., Davies, N. & Narayanaswami, C. (2008). User-Generated Content. *IEEE Pervasive Computing*, 7(4), pp. 10–11.

Kuhn, T.S. (1996). *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.

Kuper, S. & Szymanski, S. (2009). *Soccernomics*, Philadelphia, PA: Nation Books.

Kyng, Morten. (2010). Bridging the Gap between Politics and Techniques. *Scandinavian Journal of Information Systems*, 22 (1), pp. 49–68

de Laat, M. (2002). Network and content analysis in an online community discourse. In *Proceedings of the Conference on Computer Support for Collaborative Learning,* pp. 625–626.

Lacanilao, F. (1997). Continuing problems with gray literature. *Environmental Biology of Fishes*, 49(1), pp. 01–05.

Laitin, D.D. (2003). The Perestroikan Challenge to Social Science. *Politics & Society*, 31(1), pp. 163–184.

Lakhani, K.R. & Panetta, J.A. (2007). The principles of distributed innovation. *Innovations: technology, governance, globalization*, 2(3), pp. 97–112.

Lakhani, K. R., Austin, R. D., & Yi, Y. (2010). *Data.gov.* Harvard Business School.

Lampe, C., Wash, R., Velasquez, A., & Ozkaya, E. (2010). Motivations to participate in online communities. In *Proceedings of the international conference on Human factors in computing systems*, pp. 1927–1936.

Lampe, C., LaRose, R., Steinfield, C., & DeMaagd, K. (2011). Inherent barriers to the use of social media for public policy informatics. *The Innovation Journal: The Public Sector Innovation Journal*, 16, pp. 1–17.

Lampel, J. & Bhalla, A. (2007). The role of status seeking in online communities: Giving the gift of experience. *Journal of Computer-Mediated Communication*, 12(2).

Lathrop, D. & Ruma, L. (2010). *Open Government: Collaboration, Transparency, and Participation in Practice*, Sebastopol, CA: O'Reilly.

Layne, K. & Lee, J. (2001). Developing fully functional E-government: A four stage model. *Government information quarterly*, 18(2), pp. 122–136.

Lazar, J., Feng, J.H. & Hochheiser, H. (2010). *Research Methods in Human-*

Computer Interaction 1st ed., New York, NY: Wiley.

Lee, B., Barua, A. & Whinston, A.B. (1997). Discovery and representation of causal relationships in MIS research: A methodological framework. *MIS quarterly*, pp. 109–136.

Lee, F., Edmondson, A. C., Thomke, S., & Worline, M. (2004). The mixed effects of inconsistency on experimentation in organizations. *Organization Science*, pp. 310–326.

Levitt, S.D. & List, J.A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *The Journal of Economic Perspectives*, 21(2), pp. 153–174.

Lewis, R. & Reiley, D. (2010). *Does retail advertising work: Measuring the effects of advertising on sales via a controlled experiment on Yahoo*, Yahoo! Working paper.

Li, J. & Chignell, M. (2010). Birds of a feather: How personality influences blog writing and reading. *International Journal of Human Computer Studies*, 68(9), pp. 589–602.

Limm, J. (2007). Jelsoft's future: acquisition news, *vBulletin*. Retrieved 6/4/2013 from http://www.vbulletin.com/forum/forum/vbulletin-announcements/vbulletin-announcements_aa/241138-jelsoft-s-future-acquisition-news.

Lin, C.P. (2010). Learning Virtual Community Loyalty Behavior From a Perspective of Social Cognitive Theory. *International Journal of Human–Computer Interaction*, 26(4), pp. 345–360.

Lin, H.F. (2008). Determinants of successful virtual communities: Contributions from system characteristics and social factors. *Information & Management*, 45(8), pp. 522–527.

Linders, D. (2011). We-Government: an anatomy of citizen coproduction in the information age. In *Proceedings of the International Digital Government Research Conference*, pp. 167–176.

Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., & Zimmerman, J. (2011). I'm the mayor of my house: examining why people use foursquare - a social-driven location sharing application. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 2409–2418.

Locke, E.A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 3(2), pp. 157–189.

Locke, E.A. & Latham, G.P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), p. 705.

Ludford, P. J., Cosley, D., Frankowski, D., & Terveen, L. (2004). Think different:

increasing online community participation using uniqueness and group dissimilarity. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 631–638.

Luger, E, Moran, S, Rodden, T. (2013). Consent for All: Revealing the Hidden Complexity of Terms and Conditions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*.

Luhtanen, R. & Crocker, J. (1992). A Collective Self-Esteem Scale: Self-Evaluation of One's Social Identity. *Personality and Social Psychology Bulletin*, 18(3), pp. 302–318.

Ma, M. & Agarwal, R. (2007). Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research*, 18(1), p. 42.

Mailchimp (2010). Email Marketing Benchmarks by Industry. *MailChimp*. Retrieved 28/4/2012 from http://www.mailchimp.com/resources/research/email-marketing-benchmarks-by-industry

Malamoud, C. (2010). By The People. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly Media.

Markus, M. L., Bjørn-Andersen, N. 1987. Power over Users: Its Exercise by System Professionals. *Communications of the ACM* 30 (6), pp. 498–504

Mathiassen, L. (1998). *Reflective Systems Development*. Doctoral dissertation, Aalborg, Denmark: Aalborg University.

Mattus, M. (2007). Finding Credible Information A Challenge to Students Writing Academic Essays. *Human IT* (9.2), pp. 1–28.

McAuley, L., Pham, B., Tugwell, P., & Moher, D. (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *The Lancet*, 356(9237), pp. 1228–1231.

McDermott, P. (2010). Building open government. *Government Information Quarterly*, 27(4), pp. 401–413.

McGrath, J. (1995). Methodology Matters: Doing research in the behavioral and social sciences. In *Readings in Human-Computer Interaction: Toward the Year (2000), Second Edition*. San Francisco, CA: Morgan Kaufmann.

Meijer, A. & Thaens, M. (2010). Alignment 2.0: Strategic use of new internet technologies in government. *Government Information Quarterly*, 27(2), pp. 113–121.

Meister, M. (2010). København spildte 5 millioner på hjemmeside-flop til klimatopmøde. *Version2.* Retrieved 30/1/2013 from

http://www.version2.dk/artikel/koebenhavn-spildte-5-millioner-paa-hjemmeside-flop-til-klimatopmoede-17379.

Meister, M. (2011). Dokumentation: Det københavnske Facebook-flop i kroner og øre. *Version2*. Retrieved 30/1/2013 from http://www.version2.dk/artikel/dokumentation-det-koebenhavnske-facebook-flop-i-kroner-og-oere-17672.

Mergel, I. (2010). Gov 2.0 revisited: Social media strategies in the public sector. *PA Times*, 33(3), p. 7.

Mermelstein, R., Cohen, S., Lichtenstein, E., Baer, J. S., & Kamarck, T. (1986). Social support and smoking cessation and maintenance. *Journal of Consulting and Clinical Psychology*, 54(4), pp. 447–453.

Milgram, S. (1974). *Obedience to Authority: An Experimental View*, New York, NY: Harper and Row.

Miller, E. (2010). Disrupting Washington's Golden Rule. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Missingham, R. (2011). E-parliament: Opening the door. *Government Information Quarterly*, 28(3), pp. 426–434.

Molka-Danielsen, J. (2011). Exploring the Role of Virtual Worlds in the Evolution of a Co-creation Design Culture. In *Nordic Contributions in IS Research*. Springer, pp. 3–15.

Montgomery, D.C. (2004). *Design and Analysis of Experiments* 6th ed., New York, NY: Wiley.

Moon, J.Y. & Sproull, L.S. (2008). The Role of Feedback in Managing the Internet-Based Volunteer Work Force. *Information Systems Research* (19)(4), pp. 494–515.

Moore, M.H. (1997). *Creating Public Value: Strategic Management in Government*, Cambridge, MA: Harvard University Press.

Muhlberger, P. & Stromer-Galley, J. (2009). Automated and hand-coded measurement of deliberative quality in online policy discussions. In *Proceedings of the Annual International Conference on Digital Government Research*, pp. 35–41.

Muller, M. (2012). Lurking as personal trait or situational disposition: lurking and contributing in enterprise social media. In *Proceedings of the ACM conference on Computer Supported Cooperative Work*, pp. 253–256.

Murphy, C.A., Coover, D. & Owen, S.V. (1989). Development and Validation of the Computer Self-Efficacy Scale. *Educational and Psychological Measurement*, 49(4), pp. 893–899.

Nam, T. (2011). Suggesting frameworks of citizen-sourcing via Government 2.0. *Government Information Quarterly*, 29(1), pp. 12–20.

Nambisan, S. & Baron, R.A. (2009). Different Roles, Different Strokes: Organizing Virtual Customer Environments to Promote Two Types of Customer Contributions. *Organization Science* 21(2), pp. 554-572.

Nardi, B. (2010). *My life as a night elf priest: An anthropological account of World of Warcraft*, Ann Arbor, MI: University of Michigan Press.

Netherland, W. & McCroskey, C. (2010). Case Study: Tweet Congress. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Newell, A. & Card, S.K. (1985). The prospects for psychological science in human-computer interaction. *Human-computer interaction*, 1(3), pp. 209–242.

Nicklisch, A. & Salz, T. (2008). Reciprocity and Status in a Virtual Field Experiment. *SSRN eLibrary*.

Nonnecke, B. & Preece, J. (2000). Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 73–80.

Nonnecke, B., Andrews, D. & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1), pp. 7–20.

Nov, O., Naaman, M. & Ye, C. (2008). What drives content tagging: the case of photos on Flickr. In *Proceeding of the SIGCHI conference on Human factors in computing systems*, pp. 1097–1100.

Noveck, S. (2010). The Single Point of Failure. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Nygaard, K, & Bergo, O. T. (1975). The Trade Unions - New Users of Research. *Personnel Review* 4 (2), pp. 5–10.

O'Reilly III, C.A., Main, B.G. & Crystal, G.S. (1988). CEO compensation as tournament and social comparison: A tale of two theories. *Administrative Science Quarterly*, pp. 257–274.

O'Reilly, T. (2010). Government As a Platform. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Odum, H.W. (1952). Folk sociology as a subject field for the historical study of total human society and the empirical study of group behavior. *Social Forces*, 31, p. 193.

OECD (2008). *The Future Of E-Government: Agenda 2020*. Paris, France: OECD

Oktay, H., Taylor, B.J. & Jensen, D.D. (2010). Causal discovery in social media using quasi-experimental designs. In *Proceedings of the First Workshop on Social Media Analytics*, pp. 1–9..

Olivera, F., Goodman, P.S. & Swee-Lin Tan, S. (2008). Contribution behaviors in distributed environments. *MIS Quarterly*, 32(1), pp. 23–42.

Olson, G.M. & Olson, J.S. (1991). User-centered design of collaboration technology. *Journal of Organizational Computing and Electronic Commerce*, 1(1), pp. 61–83.

Olson, G.M. & Olson, J.S. (2003). Human-Computer Interaction: Psychological Aspects of the Human Use of Computing. *Annual Review of Psychology*, 54(1), pp. 491–516.

Olsson, T. (2009). Understanding collective content: purposes, characteristics and collaborative practices. In *Proceedings of the fourth international conference on Communities and technologies*, pp. 21–30.

Open Government Partnership (2013). Country Commitments, *Opengovpartnership.org.* Retrieved 30/1/2013 from http://*opengovpartnership.org*.

Osimo, D. (2008). Web 2.0 in Government: Why and How? *European Commission Joint Research Centre Institute for Prospective Technological Studies*. JRC, European Commission.

Parks, W. (1957) Open Government Principle: Applying the right to know under the Constitution. *George Washington Law Review*, 26, p. 1.

Parycek, P. & Sachs, M. (2010). Open Government–Information Flow in Web 2. 0. *European Journal of ePractice*, 9(3), pp. 1–12.

Peters, B.G. & Pierre, J. (1998). Governance Without Government? Rethinking Public Administration. *Journal of Public Administration Research and Theory*, 8(2), pp. 223–243.

PEW (2012). Who's Online: Internet User Demographics. *PEW*. Retrieved 5/9/2012 from http://pewinternet.org/Trend-Data-(Adults)/Whos-Online.aspx

Pollitt, C. (2012). *New Perspectives on Public Services: Place and Technology*, Oxford, England: Oxford University Press.

Popper, K. (2002). *The logic of scientific discovery*, London: Routledge.

Porter, S.R. & Whitcomb, M.E. (2004). Understanding the effect of prizes on response rates. *New Directions for Institutional Research* (121), pp. 51–62.

Posner, R.A., (1998). Rational choice, behavioral economics, and the law. *Stanford Law Review*, pp. 1551–1575.

Preece, J., Maloney-Krichmar, D. & Abras, C. (2003). History of emergence of online communities. In B. Wellman (ed) *Encyclopedia of Community.* Thousand Oaks, CA: Sage.

Provan, K.G. & Milward, H.B. (2001). Do networks really work? A framework for evaluating public-sector organizational networks. *Public Administration Review*, 61(4), pp. 414–423.

Purchase, H.C. (2012). *Experimental Human-Computer Interaction: A Practical Guide with Visual Examples*, Cambridge, England: Cambridge University Press.

Quora (2013). How has Facebook's UI changed over time? *Quora*. Retrieved 30/1/2013 from http://www.quora.com/permalink/B8aWbYqXZ.

Rafaeli, S., Ravid, G. & Soroka, V. (2004). De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital. *Proceedings of the Annual Hawaii International Conference.* IEEE, pp. 10ff.

Rafaeli, S., Raban, D.R. & Ravid, G. (2005). Social and economic incentives in Google Answers. In *ACM Workshop Sustaining Community: The role and design of incentive mechanisms in online systems.*

Rashid, A. M., Ling, K., Tassone, R. D., Resnick, P., Kraut, R., & Riedl, J. (2006). Motivating participation by displaying the value of contribution. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 955–958.

Reich, S. (2010). Adolescents' Sense Of Community On Myspace And Facebook: A Mixed-Methods Approach. *Journal Of Community Psychology*, 38(6), pp. 688–705.

Reips, U. D. (2002a). Internet-Based Psychological Experimenting. *Social Science Computer Review*, 20(3), pp. 241 –249.

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

Reips, U. D. (2007). The methodology of Internet-based experiments. *The Oxford Handbook of Internet Psychology*, New York, NY: Oxford University Press

Reis, H.T. & Gosling, S.D. (2010). Social psychological methods outside the laboratory. *Handbook of social psychology*. New York: John Wiley & Sons.

Ren, Y., Kraut, R. & Kiesler, S (2007). Applying Common Identity and Bond Theory to Design of Online Communities. *Organization Studies*, 28(3), pp. 377 –408.

Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*, Cambridge, MA: MIT press.

Rhodes, R. A. W. (1996). The New Governance: Governing without Government1. *Political Studies*, 44(4), pp. 652–667.

Richins, M.L. (1995). Social comparison, advertising, and consumer discontent. *American Behavioral Scientist*, 38(4), pp. 593–607.

Richtel, M. & Sack, K. (2012). Facebook Urges Members to Add Organ Donor Status. *The New York Times*. Retrieved 22/1/2013 from http://www.nytimes.com/2012/05/01/technology/facebook-urges-members-to-add-organ-donor-status.html.

Ries, E. (2011). *The Lean Startup.* New York, NY: Random House.

Robinson, D., Yu, H., Zeller, W., & Felten, E. (2009). Government data and the invisible hand. *Yale Journal of Law & Technology*, 11, pp. 160.

Robinson, D., Yu, H. & Felten, E. (2010). Enabling Innovation for Civic Engagement. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Rodden, K., Hutchinson, H. & Fu, X. (2010). Measuring the user experience on a large scale: user-centered metrics for web applications. In *Proceedings of the SIGCHI international conference on Human factors in computing systems*, pp. 2395–2398.

Ross, L., Lepper, M. & Ward, A. (2010). History Of Social Psychology: Insights, Challenges, And Contributions To Theory And Application. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (eds). *Handbook of social psychology*. New York, NY: Wiley.

Sadri, G. & Robertson, I.T. (1993). Self-efficacy and Work-related Behaviour: A Review and Meta-analysis. *Applied Psychology*, 42(2), pp. 139–152.

Salamon, L.M. (1995). *Partners in public service: government-nonprofit relations in the modern welfare state*, Baltimore, MD: Johns Hopkins University Press.

Sanford, C. & Rose, J. (2007). Characterizing eParticipation. *International Journal of Information Management*, 27(6), pp. 406–421.

Sani, F. & Todman, J. (2006). *Experimental design and statistics for psychology: a first course*, Malden, MA: Wiley-Blackwell.

Sarkar, C., Wohn, D.Y. & Lampe, Cliff (2012). Predicting length of membership in online community "everything2" using feedback. In *Proceedings of the ACM conference on Computer Supported Cooperative Work Companion*, pp. 207–210.

Schelling, T.C. (1978). *Micromotives and Macrobehavior*, New York, NY: Norton.

Schram, S. & Caterino, B. (2006). *Making Political Science Matter: Debating Knowledge, Research, And Method*, New York, NY: NYU Press.

Schuler, D. (2010). Online Deliberation and Civic Intelligence. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Searle, J.R. (1995). *The construction of social reality*, New York: Free Press.

Shah, C. & Pomerantz, J. (2010). Evaluating and predicting answer quality in community QA. In *Proceedings of the international ACM SIGIR conference on Research and development in information retrieval*, pp. 411–418.

Shani, G. & Gunawardana, A. (2011). Evaluating recommendation systems. *Recommender Systems Handbook*, Springer, pp. 257–297.

Shen, C. & Chiou, J. (2009). The effect of community identification on attitude and intention toward a blogging community. *Internet Research* (19)(4), pp. 393–407.

Shen, K. & Khalifa, M. (2009). Design for Social Presence in Online Communities: A Multidimensional Approach. *AIS Transactions on Human-Computer Interaction*, 1(2), pp. 33–54.

Shen, K., Yu, A.Y. & Khalifa, M. (2010). Knowledge contribution in virtual communities: accounting for multiple dimensions of social presence through social identity. *Behaviour and Information Technology*, 29(4), pp. 337–348.

Shiflett, C. (2009). *Essential PHP Security*, Sebastopol, CA: O'Reilly.

Shin, H.K. & Kim, K.K. (2010). Examining identity and organizational citizenship behaviour in computer-mediated communication. *Journal of Information Science*, 36(1), pp. 114 –126.

Short, J. W. (2001). E. and Christie, B. (1976) *The social psychology of telecommunications*. London, England: Wiley.

Sifry, M. (2010). "You can be the eys and ears": Barack Obama and the Wisdom of Crowds. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly, pp. 115–122.

Skelcher, C., Mathur, N. & Smith, M. (2005). The Public Governance of Collaborative Spaces: Discourse, Design and Democracy. *Public Administration*, 83(3), pp. 573–596.

Smith, P.B. (2005). Is there an indigenous European social psychology? *International Journal of Psychology*, 40(4), pp. 254–262.

Sniderman, P.M. & Piazza, T. (1993). *The scar of race*, Cambridge, England: Cambridge University Press.

Spinuzzi, C. (2002). A Scandinavian challenge, a US response: methodological assumptions in Scandinavian and US prototyping approaches. In *Proceedings of the international conference on Computer documentation*, pp. 208–215.

Spinuzzi, C. (2005). The methodology of participatory design. *Technical Communication*, 52(2), pp. 163–174.

Starbuck, W.H. (2003). Turning Lemons into Lemonade Where Is the Value in

Peer Reviews? *Journal of Management Inquiry*, 12(4), pp. 344–351.

Starbuck, W.H. (2005). How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science*, pp. 180–200.

Statistics Denmark (2011a). 97 pct. af unge kvinder bruger sociale medier. *Danmarks Statistik*. Retrieved 30/1/2013 from http://www.dst.dk/pukora/epub/Nyt/(2011)/NR403.pdf.

Statistics Denmark (2011b). Publikation: Offentligt forsørgede - 16-64 årige. *Danmarks Statistik*. Retrieved 4/2/2012 from http://www.dst.dk/da/statistik/publikationer/vispub.aspx?cid=16672.

Statistics Denmark (2012). Tabeller i Statistikbanken: Folke- og førtidspension. *Danmarks Statistik*. Retrieved 30/1/2013 from http://www.dst.dk/da/statistik/nyt/relateret.aspx?psi=1303.

Stoddard, J. L., Augustson, E. M., & Moser, R. P. (2008). Effect of adding a virtual community (bulletin board) to smokefree. gov: randomized controlled trial. *Journal of medical Internet research*, 10(5).

Strahan, E. J., Wilson, A. E., Cressman, K. E., & Buote, V. M. (2006). Comparing to perfection: How cultural norms for appearance affect social comparisons and self-image. *Body Image*, 3(3), pp. 211–227.

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2005). Assessing information quality of a community-based encyclopedia. In *Proceedings of the International Conference on Information Quality*, pp. 442–454.

Sundukovskiy, S.L. (2010). *The impact of experimentation on product development in companies involved in interactive marketing*. Doctoral Dissertation,Capella University.

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor Books.

Swallow, S.R. & Kuiper, N.A. (1988). Social comparison and negative self-evaluations: An application to depression. *Clinical Psychology Review*, 8(1), pp. 55–76.

Swann, W.B. (1983). Self-verification: Bringing social reality into harmony with the self. *Psychological perspectives on the self*, Hillsdale, NJ: Erlbaum,

Swann, W.B. & Bosson, J.K. (2010). Self and identity. *Handbook of psychology*. Oxford, England: Oxford University Press.

Swartz, A. (2010). When Is Transparency Useful? In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Tajfel, H. et al. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2), pp. 149–178.

Tajfel, H. & Turner, J.C. (1979). An integrative theory of intergroup conflict. *The social psychology of intergroup relations*, 33, p. 47.

Tang, D. et al. (2010). Overlapping experiment infrastructure: more, better, faster experimentation. In *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 17–26.

Tauberer, J. (2010). Case Study: Govtrack.us. In D. Lathrop & L. Ruma (eds). *Open Government: Collaboration, Transparency, and Participation in Practice*. Sebastopol, CA: O'Reilly.

Taylor, S.E., Buunk, B.P. & Aspinwall, L.G. (1990). Social Comparison, Stress, and Coping. *Personality and Social Psychology Bulletin*, 16(1), pp. 74–89.

Tedjamulia, S.J.J. Dean, D.L., Olsen, D.R., Albrecht, C. C. (2005). Motivating Content Contributions to Online Communities: Toward a More Comprehensive Theory. In *Proceedings of the Hawaii International Conference on System Sciences,* p. 193b.

Teerling, M.L. & Pieterson, W. (2011). How to improve e-government use: An empirical examination of multichannel marketing instruments. *Information Polity: The International Journal of Government & Democracy in the Information Age*, 16(2), pp. 171–187.

The European Parliament and The Council (1995). DIRECTIVE 95/46/EC. Retrieved 30/1/2013 from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

Thoits, P.A. (1995). Social Psychology: The Interplay between Sociology and Psychology. *Social Forces*, 73(4), p. 1231.

Thomke, Stefan, (1998). Managing Experimentation in the Design of New Products. *Management Science*, 44(6), pp. 743–762.

Thomke, Stefan (2001). Enlightened experimentation: The new imperative for innovation. *Harvard Business Review*, 79(2), pp. 66–75.

Thomsen, M. (2011). Nyt flop: København spildte 7 mio. på kommunal Facebook-kopi. *Version2*. Retrieved 30/1/2013 from http://www.version2.dk/artikel/nyt-flop-koebenhavn-spildte-7-mio-paa-kommunal-facebook-kopi-17522.

Tohidi, M. (2006). Getting the right design and the design right. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1243–1252.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), pp. 1124–1131.

Tyler, T. (2011). *Why people cooperate : the role of social motivations*, Princeton NJ: Princeton University Press.

Ullman, L. (2011). *PHP and MySQL for Dynamic Web Sites: Visual QuickPro Guide* 4th ed., San Francisco, CA: Peachpit Press.

United Nations (2008). *UN E-Government Survey (2008)*, New York: Department of Economic and Social Affairs Division for Public Administration and Development Management. Retrieved 23/3/2011 from http://unpan1.un.org/intradoc/groups/public/documents/un/unpan028607.pdf.

United Nations (2010). United Nations E-Government Survey (2010) Leveraging e-government at a time of financial and economic crisis.

Utz, S. (2003). Social identification and interpersonal attraction in MUDs. *Swiss Journal of Psychology*, 62(2), pp. 91–101.

Vaidyanathan, G. & Mautone, S. (2009). Security in dynamic web content management systems applications. *Communication of the ACM*, 52(12), pp. 121–125.

vBulletin (2010). vB4Mance Part 4: The Basics - Understanding vBulletin4 PHP Script and Hosting. *vBulletin*. Retrieved 6/4/2013 from http://www.vbulletin.com/forum/blogs/ibxanders/3935821.

Violi, N., Shneiderman, B., Hanson, A., & Rey, P. J. (2011). Motivation for Participation in Online Neighborhood Watch Communities: An Empirical Study Involving Invitation Letters. In *IEEE Third International Conference on Social Computing*, pp. 760–765.

Vonhof, S. (2010). Deficiencies of Undergraduate Forestry Curricula in Their Social Sciences and Humanities Requirements. *Journal of Forestry*, 108(8), pp. 413–418.

Wagner, H.R. (1963). Types of Sociological Theory: Toward a System of Classification. *American Sociological Review*, 28(5), pp. 735–742.

Walls, J.G., Widermeyer, G.R. & El Sawy, O.A. (2004). Assessing information system design theory in perspective: how useful was our 1992 initial rendition? *Journal of Information Technology Theory and Application*, 6(2), p. 6.

Walsham, G. (1993). *Interpreting Information Systems in Organizations* 1st ed., New York, NY: John Wiley & Sons

Wang, Y. & Fesenmaier, D.R. (2004). Towards understanding members' general participation in and active contribution to an online travel community. *Tourism Management*, 25(6), pp. 709–722.

Weatherall, J. (2001). *Vejen til førtidspension: programområde 9: effektivitet i velfærdsproduktionen*, Copenhagen, Denmark: Socialforskningsinstituttet.

Webster, J. & Watson, R.T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly*, 26(2), p. 3.

Weimer, M., Gurevych, I. & Mühlhäuser, M. (2007). Automatically assessing the post quality in online discussions on software. In *Proceedings of the Annual Meet-*

ing of the ACL on Interactive Poster and Demonstration Sessions: Association for Computational Linguistics, pp. 125–128.

White House. (2009). Open government directive. *Memorandum for the Heads of Executive Departments and Agencies*.

White House (2011). A Status Report on the Administration's Commitment to Open Government. *The White House*. Retrieved 17/2/2012 from http://www.whitehouse.gov/blog/2011/09/16/status-report-administration-s-commitment-open-government.

White House (2013). Fact Sheet: Open Government Flagship Initiatives, *The White House*. Retrieved 30/1/2013 from http://www.whitehouse.gov/open/documents/flagship-initiatives.

Whittaker, Z. (2012). Sweet irony: EU imposes cookie law, ignores own rules, *ZDNet*. Retrieved 30/1/2013 from http://www.zdnet.com/blog/london/sweet-irony-eu-imposes-cookie-law-ignores-own-rules/4975.

Wikimedia Foundation (2010). Wikimedia Foundation Feb 2010 Letter to the Board - Strategic Planning. *Wikimedia*. Retrieved 30/1/2013 from http://strategy.wikimedia.org/wiki/Wikimedia_Foundation/Feb_(2010)_Letter_to_the_Board.

Wilson, S.M. & Peterson, L.C. (2002). The anthropology of online communities. *Annual review of anthropology*, pp. 449–467.

Wilson, T.D., Aronson, Elliot & Carlsmith, K. (2010). The Art of Laboratory Experimentation. In Susan T. Fiske, Daniel T. Gilbert, & Gardner Lindzey (eds). *Handbook of Social Psychology*. New York, NY: Wiley.

Wohn, D., Velasquez, A., Bjornrud, T., & Lampe, C. (2012). Habit as an explanation of participation in an online peer-production community. In *Proceedings of the ACM SIGCHI annual conference on Human Factors in Computing Systems*, pp. 2905–2914.

Wolf, C. G., Carroll, J. M., Landauer, T. K., John, B. E., & Whiteside, J. (1989). The role of laboratory experiments in HCI: help, hindrance, or ho-hum? In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 265–268.

Wolff, W.M. (1970). A study of criteria for journal manuscripts. *American Psychologist*, 25(7), p. 636.

Wu, K., Zhu, Q., Zhao, Y., & Zheng, H. (2010). Mining the Factors Affecting the Quality of Wikipedia Articles. In *International Conference of Information Science and Management Engineering*, pp. 343–346.

Wyld, D.C. (2009). *Moving to the cloud: An introduction to cloud computing in government*, IBM Center for the Business of Government. Retrieved 30/1/2013

fromhttp://www.ukeig.org.uk/sites/default/files/WyldCloudReport_0.pdf

Yang, H.L. & Lai, C.Y. (2010). Motivations of Wikipedia content contributors. *Computers in Human Behavior*, 26(6), pp. 1377–1383.

Yildiz, M. (2007). E-government research: Reviewing the literature, limitations, and ways forward. *Government Information Quarterly*, 24(3), pp. 646–665.

Yin, R.K. (2002). *Case Study Research: Design and Methods,* 3rd ed., Thousand Oaks, CA: SAGE Publications.

Zadro, L., Williams, K. D & Richardson, R. (2004). How low can you go? Ostracism by a computer is sufficient to lower self-reported levels of belonging, control, self-esteem, and meaningful existence. *Journal of Experimental Social Psychology*, 40(4), pp. 560–567.

Zhang, P., Benbasat, I., Carey, J., Davis, F., Galletta, D., & Strong, D. (2002). Human-computer interaction research in the MIS discipline. *Communications of the Association for Information Systems (9)*, p. 355.

Zitt, M. & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management*, 42(6), pp. 1513–1531.

Appendices

Appendix 1   Summary of Included Papers

Author	Theory	Number Subjects	Research Design	Methodology	Case
Cosley et al. (2003)	Group Conformity	N = 70,000 n = 488	Single Case	Field experiment	MovieLens
Utz (2003)	Social identity theory	N = n = 217	Single Case	Survey Field Experiment	Undisclosed
Ludford et al. (2004)	Collective Effort Model	N = 8,500 n = 245	Single Case	Field experiment Survey	MovieLens
Wang and Fesenmaier (2004)	Self-efficacy	N = 150,000 n = 322	Single Case	Survey	Undisclosed
Beenen et al. (2005)	Collective Effort Model (exp 1) Goal-setting (exp 2)	N = 80,000 n = 830 (exp 1) n = 834 (exp 2)	Single Case	Field experiment	MovieLens
Cosley et al. (2006)	Collective Effort Model	N = n/a n = 2,723	Single Case	Field experiment	MovieLens
Joyce and Kraut (2006)	Reciprocity	N = n = 2,777	Longitudinal study (6 months)	Log analysis	netscape.public.mozilla.ui alt. support.diet alt.support.cancer.breast alt.politics.usa.constitution.gun-rights alt. sports.hockey.nhl.ny-rangers alt.baldspot
Rashid et al. (2006)	Collective Effort Model	N = n/a n = 160	Single Case	Field experiment Survey	MovieLens
Harper et al. (2007)	Collective Effort Model	N = 100,000 n = 2,021	Single Case	Field experiment	MovieLens

444

Study	Construct/Theory	N	Study Type	Method	Platform
Ma and Agarwal (2007)	Perceived identity verification	N = 5,559 n = 666	Comparative case study	Survey	Lexus IS300 Quitnet
Chan & S. Y. Li (2008)	Reciprocity	N = n/a n = 899 (survey)	Single Case	Netnography Survey	Onlylady
Drenner et al. (2008)	Cognitive Dissonance	N = 583	Single Case	Field experiment	MovieLens
Nov, Naaman and Ye (2008)	Social presence	N = n/a n = 237	Single Case	Survey Log analysis	Flickr
Lin (2008)	Need to belong	N = n/a n = 198	Theory Building	Survey	n/a
Burke, Marlow and Lento (2009)	Social learning Feedback	N = 150,000,000 n = 140,292	Single Case	Log analysis Semi-structured interviews	Facebook
Shen and Chiou (2009)	Social identity theory	N = n/a n = 243	Theory Building	Survey	n/a
Shen and Khalifa (2009)	Social presence	N = n/a n = 430	Comparative case study	Survey	Four undisclosed communities
Carmagnola et al. (2009)	Social comparison Group conformity	N = n/a n = 45	Single Case	Evaluation	Facebook
Nambisa and Baron (2009)	Social identity theory	N = n/a n = 152	Comparative case study	Survey Log analysis	IBM Microsoft
Yang and Lai (2010)	Self-efficacy	N = n/a n = 219	Single Case	Survey	Wikipedia
Shin and Kim (2010)	Social identity theory	N = n/a n = 192	Comparative case study	Survey	Naver Daum Yahoo
Chen et al. (2010)	Social comparison	N = 109,366 n = 398	Single Case	Field experiment	MovieLens
Lin (2010)	Social learning	N = n/a	Theory Building	Survey	n/a

445

		n = 303			
Shen, Yu, and Khalifa (2010)	Social presence	N = > 400,000 n = 430	Theory Building	Survey	n/a
Lampe et al. (2010)	Social identity theory	N = 1,850,000 n = 599	Single Case	Survey Log analysis	Everything2
Chou et al. (2010)	Perceived Identity Verification	N = n/a n = 429	Comparative case study	Survey	Baidu Yahoo! knowledge+
Chou (2010)	Perceived Identity Verification	N = n/a n = 241	Comparative case study	Survey	eetimes.com China eetimes.com Taiwan
Casaló, Flavián, and Guinalíu (2010)	Social Identity Theory	N = n/a n = 456	Comparative case study	Survey	http://www.es.lastminute.com/ http://www.minube.com http://www. losviajeros.com http://comunidad.muchoviaje.com/CS/ http://www.turismo20.com/, http://www.foro.geoplaneta.com http://www. travelmarketing.biz/ http://www.escapadarural.com/ http://www.ruralon.com, http://www.viajered.com http://www.viajaris.com, http://www.es.ulises.com

446

Appendix 2   Research Procedure and Selection Strategy in Literature Review

Research into online communities has been conducted since the 1990s (Preece et al. 2003). Early research was mainly empirical descriptions of early online communities, and theory-based research is a rare sight before the turn of the millennium. Therefore, the time period included in this review is 2000-2010.

Whereas many literature reviews only quantify the research into different categories, I use a more qualitative approach in this one. This qualitative approach is deemed appropriate due to the high degree of diversity and fragmentation of the field. As many different theories and methodologies are being used, a short description of each paper is included in the review.

The purpose is to give a complete review of the literature within the scope, which, according to Webster and Watson, "covers relevant literature on the topic and is not confined to one research methodology, one set of journals, or one geographic region" (Webster and Watson 2002: xv).

There are two methods often used to perform literature reviews: journal-centric and concept-centric (Larsson and Hrastinski 2011). As mentioned above, social psychology-based research into online communities is as multidisciplinary as research gets. It is complicated to rely strictly on a journal-centric approach, as research is published in very different fields.

This literature review relies on a two-sponged approach, as recommended by Webster and Watson (2002). It begins with a journal-centric approach to identify work from the top IS journals, and subsequently expands to a concept-centric approach based on searches in Web of Knowledge. The journal-centric part of this review is conducted by scanning table of contents in leading IS journals, and the concept-centric part is conducted by searching Web of Knowledge for relevant keywords.

Phase 1: Scanning journals

Webster and Watson (2002) suggest that the research starts with key journals. The advantage of this approach is that the articles from high quality journals are included in the review. The disadvantage is that the vast majority of research is published outside of these journals. Furthermore, niche fields and areas might be completely left out of such a review.

In the initial scanning process, the following journals' tables of content were scanned for relevant articles in the period: *European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of the Association for Information Systems, Journal of Management Information Systems, Management Information Systems Quarterly,* and *Scandinavian Journal of IS.*

In addition to the IS senior scholars' basket of six journals, the Scandinavian Journal of IS was also included to root the research in a Scandinavian tradition. Table 67 shows the number of articles for each journal/year identified to be potentially relevant for this review.

Journal	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
SJIS	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	1
EJIS	1	0	1	0	1	0	0	0	0	0	1
ISJ	0	1	2	0	0	0	1	1	0	2	1
ISR	1	1	0	1	1	1	0	2	5	0	0
JAIS	0	0	0	0	1	0	0	1	3	1	1
JMIS	0	1	1	0	0	2	0	1	0	4	1
MISQ	0	1	1	0	0	3	1	0	1	0	0

Table 67: Iteration 1: Relevant papers found from these journals by reading title

Scanning the table of content from the seven journals produced 34 potentially relevant papers. The criteria used to judge whether the full text of an article should be read were broad, as all articles addressing online communities and/or social psychology are included.

This preliminary search includes many papers that are excluded from the scope of this review. Three inclusion criteria are used in the selection process. *First*, studies must rely on a social psychology theory. Consequently, most articles on online communities were excluded because they either offered no theory base or relied on another theoretical perspective, such as economics.

Second, the subject matter of the study must be an online community. One weakness about this criterion is the lack of an existing clear definition of online communities. The weakness has been mitigated by including keywords of synonyms (virtual community and internet community) and closely related terms (web 2.0 and user generated content).

Third, the research must be published in a journal or a conference paper. Books and book chapters are excluded from this review.

Journal	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
SJIS	n/a	n/a	n/a	n/a	n/a	0	0	0	0	0	0
EJIS	0	0	0	0	0	0	0	0	0	0	0
ISJ	0	0	0	0	0	0	0	0	0	0	0
ISR	0	0	0	0	0	0	0	1	0	0	0
JAIS	0	0	0	0	0	0	0	0	0	0	0
JMIS	0	0	0	0	0	0	0	0	0	0	0
MISQ	0	0	0	0	0	0	0	0	1	0	0

Table 68: Iteration 2: Relevant papers after reading full text of the 34 papers, identified in the first iteration.

Based on an analysis of the full text of the 34 papers, only two papers fulfill all three criteria for inclusion. Although the IS senior scholars' basket of six journals cannot be neglected in IS research, it is obviously not the focal point for research into online communities. This finding is consistent with Lee, Vogel, and Limayem (2003), who found that "virtual community research is lacking and is being ignored in the most prestigious journals." (Lee, Vogel, and Limayem 2003: 48). The

lack of acceptance in the top journals makes it necessary to do a broader search for literature.

Phase 2: Searching ISI Web of Knowledge

Scanning the seven journals resulted in only two papers fulfilling the criteria for this review. A benefit from the first phase is that multiple keywords are identified. These are included in the second phase of the literature gathering process: searching the ISI Web of Knowledge database. The keywords were searched for in "topic", meaning that keywords are searched for in title, abstract, author keywords, and keyword plus. The searches were conducted in August of 2011. The results from each query are listed in Table 69.

Query	Results	Included
Topic=(internet community) AND Topic=(social psychology)	23	3
Topic=(online community) AND Topic=(social psychology)	21	6
Topic=(virtual community) AND Topic=(social psychology)	30	5
Topic=(web 2.0) AND Topic=(social psychology)	8	1
Topic=(user generated content) AND Topic=(social psychology)	2	0
Topic=(online community) AND Topic=(motivation)	157	7
Topic=(social psychology) AND Topic=(online forum)	13	0
Topic=(online community) AND Topic=(contribution)	239	16
Topic=(online community) AND Topic=(participation)	452	14
Topic=(online community) AND Topic=(social comparison)	32	2
Topic=(online community) AND Topic=(social learning theory)	50	1
Topic=(online community) AND Topic=(cognitive dissonance)	1	0
Topic=(online community) AND Topic=(self-concept theory)	1	1
Topic=(online community) AND Topic=(collective effort model)	4	0
Topic=(online community) AND Topic=(social loafing)	3	0
Topic=(online community) AND Topic=(social identity theory)	22	6
Topic=(CMC) AND Topic=(social psychology)	13	0
Topic=(online community) AND Topic=(hierarchy of needs)	7	0
Topic=(online community) AND Topic=(Social Presence)	66	3
Topic=(online community) AND Topic=(need to belong)	17	0
Topic=(online community) AND Topic=(goal setting)	28	1
Total	**1161**	**65**
- Redundant	**247**	**36**
Unique included	**914**	**29**

Table 69: Queries conducted in the Web of Science database.

The queries in web of knowledge produced a total of 1,161 results. When reading the abstract of the 1,161 results, 92 potentially relevant papers (i.e. addressing online communities and/or social psychology) were selected for full-text read-

ing. The exclusion process is further explained in the next section. A total of 29 papers fulfilled all three criteria for inclusion. One of the papers found in the first phase of the data collection process was also retrieved through the Web of Knowledge Search. This gives a total of 30 papers included in this review (2 papers from phase one and 28 unique papers from phase two).

Publication outlet	Number of papers
Journal	19
Conference	11

Table 70: Publication outlets for the literature included in this review

As Table 3 shows, 19 of the papers are published in journals, while 11 are published at conferences. A tendency shows that papers using experimental designs are published at conferences, while papers relying on surveys are published in journal outlets.

Scope and Limitations

For each article included, a number of variables are coded. *General information*: (a) date of publication; (b) publication form (journal article, conference article). *Theoretical information*: (a) theory applied. *Methodological information*: (a) methodology (field experiment, laboratory experiment, survey, behavioral data); (b) research design (case study, comparative study). *Case information*: (a) community website property.

The selection process has a number of limitations. *First*, the Web of Knowledge is not a complete database of all research. If other databases such as ProQuest or Google Scholar were used, the results are likely to be somewhat different. *Secondly*, only articles written in English are included. This might be problematic as social psychology research points to some potential differences between behavior in collective and individualistic cultures (Heine 2010). This limitation is mitigated by the fact that several included articles are conducted on communities in Asia (e.g. Chou et al. 2010; Shin and Kim 2010). *Thirdly*, a single researcher has

conducted this review. This might lead to subjectivity in the selection and classi-fication.

The vast majority of the papers retrieved through the Web of Knowledge search does not meet these criteria and are excluded from this review. The high exclu-sion rate could be the result of a poor search strategy. Two questions seem sali-ent: (a) was the search too broad? And (b)were the wrong keywords used?

Because the first phase of this literature review produced only two results, I used very broad search queries (for example "online community AND participation" to make sure as many papers as possible are identified. This has caused extra and somewhat unnecessary work, as many abstracts were read in vain.

There is always a risk of searching for the wrong keywords. But by including all known synonyms for online communities (internet community, virtual commu-nity) and related terms (web 2.0, online forum, web user generated content), this risk has been minimized. Furthermore, a search was completed for each social psychological theory relevant to online communities. The high redundancy rate in the search results also show that much of the literature is covered in several queries.

In short, the high exclusion rate of papers is a result of conducting a broad search. The excluded research focuses on many related topics. For example, some papers are concerned with *identity* and *social dynamics* in an online com-munity (e.g. Perotta 2006; Szell and Turner 2010). These papers are not design-oriented, but merely concerned with social psychology dynamics. Another group of papers excluded from this review are papers examining online *e-learning* and learning communities (e.g. Tseng and Kupo 2010). This research is usually not relying on social psychology. Yet another strand of literature is concerned with the *individual consequences of online participation*, e.g. participating in online health forums, weight loss groups etc. (e.g. Hwang et al. 2010).

Social psychology is often applied as an alternative to economic theory. However, *Economic theory* is also used to examine participation to online communities. Research based on this perspective is excluded from this review. Comparing design practices based on social psychology and economic theory would be an interesting research area. Researchers use economic theory to examine *auction websites* such as eBay (e.g. Heshan 2010), and they often focus on trust and transactions.

Another research area is *knowledge contribution in closed organizational settings*. This research is concerned with employee participation rather than volunteer participation (e.g. Kankanhalli, Tan and Wei 2005). Although there are many similarities between participation to volunteer communities and organizational communities, the latter is excluded from this paper. People act for different reasons at work and in their spare time, and that makes workplace-oriented research irrelevant in this dissertation.

Finally, motivation to use websites without online communities is also being studied in much of the research (e.g. Sangwan, Siguaw and Guan 2010). As this research is not concerned with any social aspect of the usage, it is excluded from this review.

Classification

The research on online communities can basically be classified in three ways: (a) by case, (b) by design pattern, or (c) by theory.

The major benefit of classifying *by case* is that the case gets illuminated very well. As this review comprises several papers describing the same case, e.g. MovieLens and Wikipedia, a case-centric review would help explain contributions to these communities. The downside to this approach is that the time span on the web is short. The websites that are interesting today are probably not so interesting a couple of years from now. Moreover, there is a moving-target, meaning that the websites are changing significantly and rapidly over time problem (Fogg

2003). Therefore, a study of Wikipedia anno 2006 might not be directly comparable to a study of Wikipedia in 2010.

A second way to classify the literature is by *design pattern*. Most of the research included in this review contains design patterns, e.g. using higher entrance barriers (Drenner et al. 2008); giving behavioral information to users (Ling et al. 2005); or highlighting uniqueness of the users' contributions (Ludford et al. 2004). This provides an opportunity to do a design pattern-centric classification, as done by Crumlish and Malone (2009) and used on Yahoo! Pattern Library (developer.yahoo.com/ypattern). The advantage of this approach is how applicable it becomes to practitioners. The disadvantage is that the category is a suboptimal fit for most of the research, as design patterns only take up a small part of the work.

Thirdly, the literature can be classified by *theory*. Theories are sometimes – but not always – used to guide research examining and improving online communities through design. The main advantage of synthesizing the literature in a theory-centric way is that theories are more stable than the empirical domain. A theory about human behavior has more longitudinal sustainability than even the most popular online communities. By focusing on theory, knowledge can be systemically accumulated over time (Gregor 2006).

This theory-centric classification is not without problems of its own. As pointed out by Ling et al. (2004) and Arazy et al. (2010), there is a mismatch between the goals of designers and social science theories. A social science theory, e.g, social learning theory, cannot account for all variables at play in an online community. This critique can be raised to a more general critique of social psychology, as it comprises a lot of mid-range theories (Beenen et al. 2004) that are sometimes complimentary and other times mutually exclusive. Many theories can be used on any given community, making it hard to create clear categories due to the unclear lines of theoretical overlap. Another downside of this approach is that much research is excluded from this review, because it contains no clear theoretical foundation.

Appendix 3   Source Credibility and Inclusion Criteria

There is no objective and universally accepted criteria for assessing source credibility (Mattus 2007). Credibility cannot be evaluated based on a single dimensions, but requires simultaneous evaluation of many dimensions. Two key parameters are trustworthiness (truthful, unbiased, and well-intentional) and expertise (knowledgeable, experienced, and competent) (Fogg and Tseng 1999). Furthermore, when evaluating a source, such as an academic text, it is not sufficient to take the text *per se* into consideration. One must also include the context and scientific communities in which an article is written. Thus, credibility assessments are made in a given context in relations to some particular aims and goals (Mattus 2007).

Fogg & Tseng (1999) argue that "credibility is a perceived quality; it doesn't reside in an object, a person, or a piece of information" (Fogg and Tseng 1999: 80). Accepting credibility as a perceived and subjective assessment creates a need to define how source credibility is assessed in different contexts relevant to this dissertation.

Under optimal conditions, there would be one set of inclusion criteria accepted by all. But different research fields have different traditions. In political science, journals are ranked higher than conference proceedings. The opposite is true in the HCI research field. Even more problematic, different research fields have reached different stages of maturity. For example, social psychology has been evolving over a century, while open government has only been around for a couple of years. The inclusion criteria are, consequently, less rigid in open government than in social psychology. As the four different scientific communities, I interact with, each constitutes different contexts, the sources must be evaluated on somewhat different grounds, appropriate to the specific context. In the following sections the general inclusion criteria, as well as inclusion criteria for each of the scientific communities, are defined.

Evaluating source credibility across multiple domains is a complex task. At least two different factors need to be included: the scientific domain and the source medium. Time/maturity and research tradition are the two factors in the scientific domain that are relevant in this assessment. In the following, an assessment of the different source mediums is presented, followed by an analysis of the scientific domains.

Source Media

When assessing source credibility, one needs to distinguish between academic and non-academic sources. Different criteria ought to be used when evaluating them (Mattus 2007). The source mediums examined here are: journals, conference proceedings, and grey literature.

Academic Journals

A key distinction in assessing credibility is the one between academic and non-academic publications.[276] The gold standard of academic research is peer-reviews (Banks 2006).[277] The benefit of the peer-review institution is, at least in principle, that an article is evaluated by other academics, thereby being of sufficiently high credibility to be used and cited (Starbuck 2003). This benefit of the peer-review system is, however, contested on several grounds: good research appears in low-prestige journals (Starbuck 2005: 180); although most reviewers can agree upon which criteria should be used to evaluate an article, they seldom interpret the criteria in the same way (Wolff 1970; Gottfredson 1978; Gottfredson and Gottfredson 1982; Starbuck 2005); and a growing amount of research is published on the internet, making it increasingly more accepted to publish in outlets that are not peer reviewed (Ellison 2011).

---

[276] An academic article can be identified as it fulfils a basic requirement: "Academic articles are the articles which appear in academic journals ... An academic journal is, almost by definition, one whose contents have been subject to peer-review" (O' Dochartaigh 2001 quoted in Mattus 2007:2).

[277] Peer-review is defined as "a system of decision making by referees, editors, and research program directors in evaluating the quality of scientific research" (Cicchetti 1991, 119).

In sum, peer-reviews are still the most widely accepted methodology of assessing credibility. However, it is important to be aware that it is not a perfect credibility indicator, and that good research is also published other places than peer-reviewed journals

Studying the peer review process in the economic domain, Ellison (2002) finds "a dramatic slowdown of the publication process at top economic journals" (Ellison 2002: 947). In the field of economics, the review process takes 12-18 months longer in 2000, than it did in 1970 (Ellison 2002). The same trend is likely to apply to other research domains. This creates a significant time delay between the rise of a new phenomenon and the description of it in top-ranked journals. This delay is particular problematic in new fields. For example, open government is just rapidly emerging in practice, but the research published in top-ranked journals is scarce at this point. Consequently, the inclusion criteria need to be less strict for emerging research fields than the ones more well established.

Conference Proceedings

In some research domains, conference proceedings are considered to be less credible than journal articles. Some conference proceedings are not peer-reviewed. But even the conference articles that are peer-reviewed are considered inferior to those published in journals due to higher time pressure on the editor and less anonymity about reviewer, as they tend to be conference attendees. However, in scientific communities such as HCI, conference proceedings are preferred over journals. Most HCI researchers prefer to publish their work at popular conferences such as CHI and CSCW. This means the best and most credible research is published here.

Grey Literature

Grey literature is understood as "unpublished reports, dissertations, articles in obscure journals, some online journals, conference abstracts, policy documents, reports to funding agencies, rejected or unsubmitted manuscripts, non-English language articles, and technical reports" (Conn et al. 2003: 256). The obvious

drawback of grey literature is that it is not peer-reviewed. But it does come with some advantages: (1) the reporting is less biased toward significant results (Banks 2006; McAuley et al. 2000; Conn et al. 2003)[278] and (2) it is generally more available than peer-reviewed articles (Banks 2006).

Banks (2006) suspects that the spread of the internet, and the greater availability and faster publication will lead to more high-quality papers being published as grey literature, thereby helping to eliminate the distinction between peer-reviewed literature and grey literature over time. But for now, grey literature still remains questionable.[279] With some exceptions, grey literature is therefore generally excluded from this dissertation. Insofar grey literature is included, it comes from reputable publishers such as The White House, The UN, OECD, Statistics of Denmark etc.

Research Communities

Each of the four research communities that this dissertation interacts with is assessed on two parameters here: maturity and tradition. The inclusion criteria are summed up in Table 71.

Domain Medium	General	Open Government	IS	Social psychology
Peer-reviewed journals	Yes	Yes	Yes	Yes
Non peer-reviewed journals	No	Yes	No	No
Peer-reviewed Conference Proceedings	Yes	Yes	Yes	No
Non peer-reviewed Conference Proceedings	No	Yes	No	No
Grey Literature	No	Yes	No	No*

Table 71: Inclusion criteria for literature in this dissertation. *With the exception of the highly esteemed Handbook of Social Psychology

---

[278] One severe caveat of peer-reviewed literature is that it is biased toward significant results, as scholars tend to report these studies (Banks 2006). Several meta-study analyses from medicine science (McAuley et al. 2000; Conn et al. 2003) show that leaving out grey literature creates biases toward studies with significant results and inflated effect sizes.

[279] For example, in the domain of environmental biology, Lacanilao (1997) detects a heavy reliance on grey literature compromising the overall credibility.

General Inclusion Criteria

Unless research is related to the one of the four research communities addressed below, the following inclusion criteria are applied: only peer-reviewed journals and conference proceedings are included. Grey literature is excluded.

Inclusion Criteria For Open Government

In most research fields, scientific results are gradually accumulated and validated over many years (Kuhn 1996). Open government is in its early stage, and only few journal articles are published. The aim of Chapter 2 on open government is to ground the field and identify important topics, trade-offs, and discussions. Relying exclusively on peer-reviewed papers would make Chapter 2 very short. Not including any literature is hardly desirable, as the overview of the field would be subject to many incredulity errors. When it comes to open government research, all sources are included.

In a review of the eParticipation literature, Sanford and Rose (2007) note that the eParticipation literature is suggestive, normative and optimistic. The same characteristics apply to open government literature. Much of the open government literature can be characterized as suggestive, normative, and evangelizing with scarce data to back up the claims. The aim of some open government literature is to persuade government agencies to engage in more and better open government. The Anthology *Open Government* by Lathrop and Ruma (2010) serves as an illustration of this. Although it does not claim to be an academic work, it has little rigorous empirical evidence. Thus, Chapter 2 is useful to ground the open government field, but does not provide an exhaustive empirical review of open government. This is both frustrating and normal in emerging disciplines. It takes time to get solid evidence. The slow peer review process in esteemed journals contributes to this delay.

Inclusion Criteria For IS

As described in Chapter 1, the IS sub-discipline HCI has been around since the 1970s and is thus well established. As noted above, much high-quality work in

the HCI field is published in conference proceedings. Some of the most important proceedings relevant to this dissertation are CHI, CSCW, and Hawaii International Conference on System Sciences. The inclusion criteria for HCI related work are, therefore, peer-reviewed journals and peer-reviewed conferences. All other sources are excluded.

Inclusion Criteria for Social Psychology

Social psychology is an old and established discipline. In the field, a majority of the best research is published in academic journals such as Journal of Experimental Social Psychology, and Personality and Social Psychology Bulletin. Therefore, only research from journals is included regarding social psychology. One exception is the *Handbook of Social Psychology*, which has been around since 1935 and is a highly esteemed publication.

Potential Fallacies

When evaluating source credibility one can make two types of errors: gullibility errors and incredulity errors (Fogg and Tseng 1999). These errors are parallels to Type I and Type II errors, discussed earlier in this chapter. *Gullibility errors* mean that sources are accepted, although they should not have been due to lack of credibility. *Incredulity errors* entail that a source is not used due to credibility concerns, although it is in fact sufficiently credible (Fogg and Tseng 1999).

Both types of errors are critical, as they reduce the quality and usefulness of research when conducted (Mattus 2007). In the general inclusion criteria, HCI and social psychology, the biggest concern is incredulity errors, as the criteria are set high. In open government and experimentation, the more dominating concern is gullibility errors, as the inclusion criteria are more lax.

Appendix 4   IRB Application and Exempt

Cornell University
Office of
Research Integrity and Assurance

Institutional Review Board for Human Participants

Concurrence of Exemption

To: Amit Sharma

From: Matthew Aldridge, Senior IRB Administrator

Date: May 01, 2012

RE: **Protocol ID#:**   1204002989

            Project(s):   Encouraging Contribution to Online Rulemaking at Regulation.gov

A member of the Office of Research Integrity and Assurance (ORIA) has reviewed the above-referenced project and found it to qualify for **Exemption from IRB Review** according to paragraph #2 of the Department of Health and Human Services Code of Federal Regulations 45 CFR 46.101(b).

This proposal has not been evaluated for scientific merit, except to weigh the risk to the human participants in relation to the potential benefits.

Please be aware of the following:

° Exemption from IRB review does not absolve the investigator from ensuring that the welfare of the research subjects is protected and that methods used and information provided to gain participant consent are appropriate to the activity. It is your responsibility as a researcher to familiarize yourself with and conduct the research in accordance with the ethical standards of the Belmont Report (http://ohsr.od.nih.gov/guidelines/belmont.html).
° You must notify the ORIA office of changes or amendments to the above-referenced protocol **BEFORE** their implementation.
° You are not required to submit progress reports or requests for continuing review/approval to ORIA, unless you modify your study protocol.

c: Daniel Cosley

Appendix 5 MySQL Queries for Randomization in Experiment 1

Add everyone to group 1: UPDATE `k10_user` SET `thomas_group`=1. Add all even numbers to group 2: UPDATE `k10_user` SET `thomas_group`=1 WHERE `userid`%2=0. Add all users which user id that can be divided by 3 to group 3: UPDATE `k10_user` SET `thomas_group`=1 WHERE `userid`%3=0

Appendix 6   Survey In Experiment 2

Afslut denne undersøgelse

K10 Brugerundersøgelse

K10 Brugerundersøgelse

Kære K10 bruger

Vi forsøger hele tiden at forbedre K10. Vi håber du vil hjælpe os ved at udfylde denne korte undersøgelse.

Informationen er KUN til internt brug på K10, og bliver IKKE videregivet til andre.

Som tak for hjælpen trækker vi lod blandt deltagerne om tre gavekort af 250 kr. stykket til saxo.dk.

Hvis du vil deltage i lodtrækningen skal du indtaste dit brugernavn nederst.

Husk at trykke færdig i besvarelsen nederst på siden, når du er færdig.

1. I hvilket år er du født?

2. Hvilket køn er du?

Mand

Kvinde

3. Hvad kan du bedst lide ved K10?

4. Hvad kunne du godt tænke dig var anderledes ved K10?

5. Hvor mange indlæg forventer du i alt at skrive på K10, de næste to uger?

Ingen

1-5

6-10

11-20

21 eller flere

Andet (angiv venligst)

6. Er der et K10 medlem, som gør noget særlig godt for andre brugere?

7. Vi trækker lod om tre gavekort. Det er helt frivililgt, om du ønsker at være med i lodtrækningen. For at kunne vinde gavekortet, skal du skrive dit brugernavn.

Brugernavn:

Færdig

Appendix 7   User Replies to Emails and Messages Sent in Experiment 3

User A   Jeg har ikke nogen spørgsmål, men vil bare sige tak for at få at vide, at folk kan bruge mine indlæg.

User B   Hej

11.7.1.1.1. Ej nu kan jeg jo slet ikke finde ud af hvilket ben jeg skal stå på.

11.7.1.1.2. Føler enlig ikke at jeg gør nogen speciel indsats, men er ikke i tvivl om at jeg gør mig bemærket med mine temmelig laaaaange indlæg.

11.7.1.1.3. Sys der bestemt også er andre der fortjener lidt ros (går ud fra flere har fået *ss*) men som jeg selv svarede kan og vil jeg ikke nævne en bestemt, fordi hver bruger har sin stærke side, hvor de bidrager, k10 ville ikke være til uden langt de fleste brugere det er i min verden temmelig simpelt, så den ene er ikke bedre end den anden.

11.7.1.1.4. Det falder ellers lige oven i at jeg har besluttet at holde en pause med at skrive, især fordi en bestemt bruger irritere mig lidt for meget (tror enlig ik det er med vilje) og såre simpelt for ikke at komme til at åbne min kæft lidt for højt, så hellere holde den lukket

11.7.1.1.5. Ud over at det sagtens kan være mig der er lidt for nærtagende i tiden.

Men immervæk så varmer det, og jeg ville ønske jeg kunne sige TAK for rosen.

User C   11.7.1.1.6. Hej..

11.7.1.1.7. Sådan et skulderklap varmer bare helt vildt.. *S*

11.7.1.1.8. Og netop i en tid hvor man kæmper med både at få an-
erkendt at man ikke kan så meget mere - og selv aner-
kende det.... og samtidig føler fordømmelsen fra omgi-
velserne som sygemeldt kontanthjælpsmodtager... er det
den slags tilkendegivelser af og til der er med til at holde
en oppe...

User D   Hej.
Tusind tak.

User E   11.7.1.1.9. Tusind tak. Det er en dejlig hilsen at få. Jeg er glad for at
kunne give lidt tilbage efter al den hjælp som jeg selv har
fået her på K10.

Appendix 8   Google AdWords Ad Copy in Experiment 4

Group	Ad Copy
Highlight individual benefit B1	**Få råd om Fleksjob** Få råd og hjælp om fleksjob fra andre på K10
Highlight small group benefit B2	**Del dine viden on Fleksjob** Del dine erfaringer om fleksjob med ligesindede på K10
Highlight large group benefit B3	**Del dine viden on Fleksjob** Del dine erfaringer om fleksjob med andre I Danmark på K10
Control group / generic advertisement B4	**Forum om fleksjob** Forum om fleksjob. Meld dig ind på K10

Histogram

Mean = 00:03:55
Std. Dev. = 0.0051 days
N = 460

Histogram

Mean = 4.62
Std. Dev. = 5.649
N = 460

Histogram

Mean = .02
Std. Dev. = .146
N = 460

Histogram

Mean = .02
Std. Dev. = .167
N = 460

Histogram

Mean = .1
Std. Dev. = .294
N = 460

Histogram

Mean = .39
Std. Dev. = .489
N = 460

Frequency (y-axis)

Bounce (x-axis)
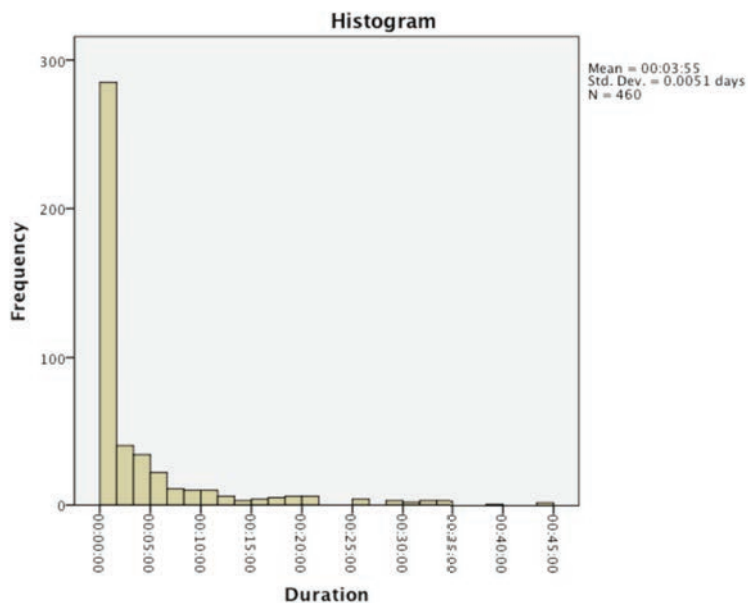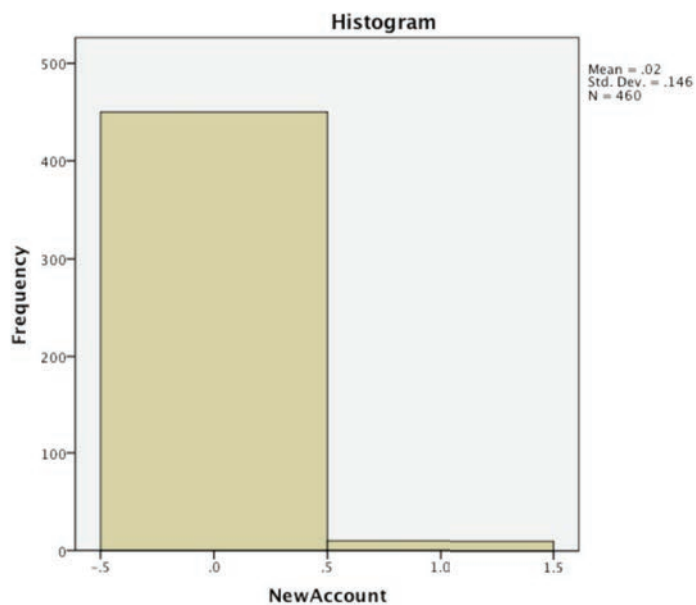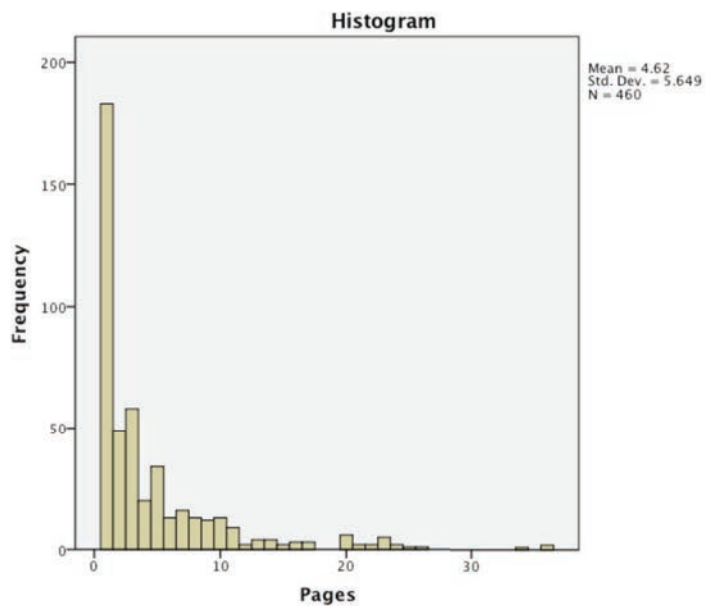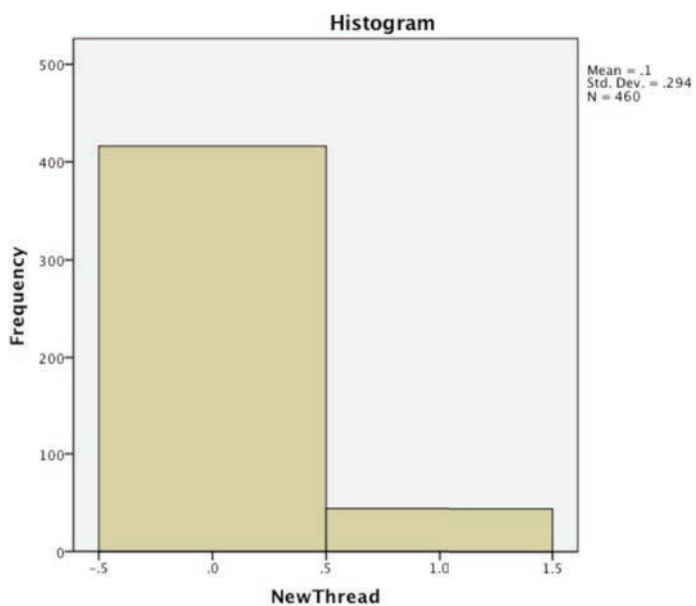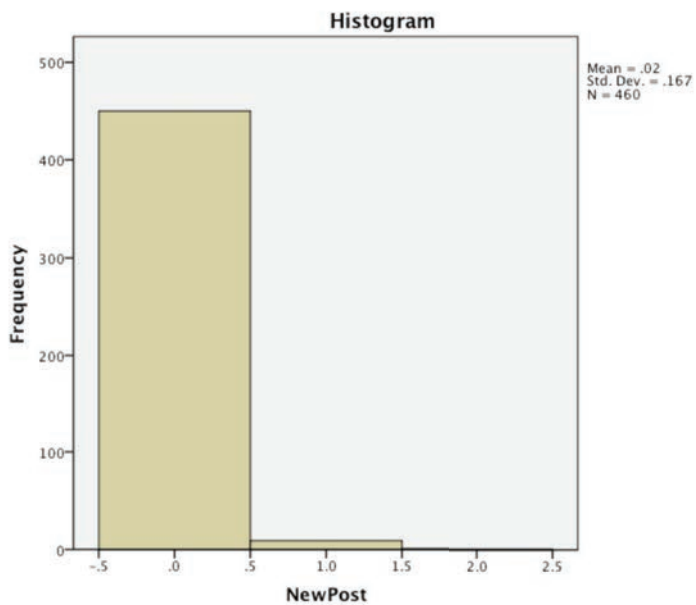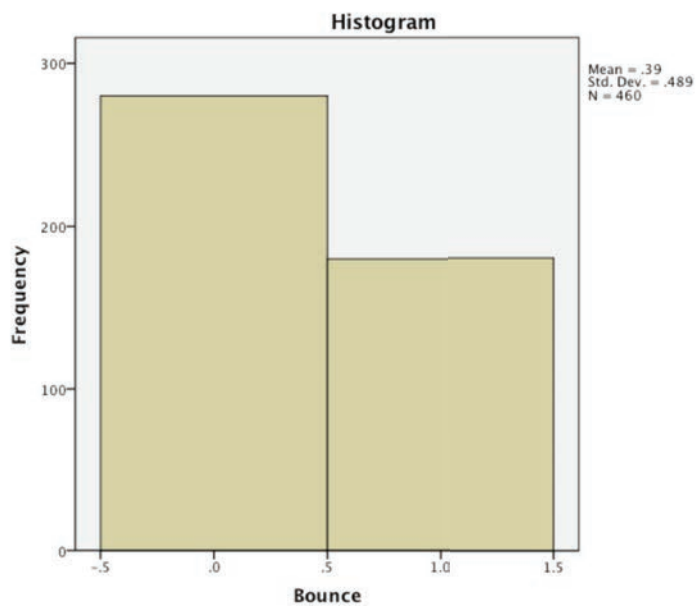
Appendix 10   Email and Message Sent in Experiment 3

Email / message content

Titel: Din deltagelse på K10 er værdsat

Hej [Brugernavn]

I sidste måned havde vi en spørgeskemaundersøgelse på K10. Blandt spørgsmålene var følgende: *Er der et K10 medlem, som gør noget særlig godt for andre brugere?*

Der var flere brugere, der fremhævede netop dig for din indsats på siden.

Så vi vil bare lige fortælle dig, at andre brugere bemærker dig og sætter pris på din indsats på K10.

Hvis du har nogen spørgsmål, kan du bare svare på denne besked.

Mvh K10

Ps. Mange brugere skrev også, at K10 løfter i flok – og at alle brugere hjælper på deres egen måde. Dette er vi helt enige i. Vi ønsker derfor ikke at fremhæve nogle brugere frem for andre, men blot at give dig et stort tak for din indsats på K10. Den betyder meget for andre mennesker.

19. Thomas Lyse Hansen
Six Essays on Pricing and Weather risk in Energy Markets

20. Sabine Madsen
Emerging Methods – An Interpretive Study of ISD Methods in Practice

21. Evis Sinani
The Impact of Foreign Direct Investment on Efficiency, Productivity Growth and Trade: An Empirical Investigation

22. Bent Meier Sørensen
Making Events Work Or, How to Multiply Your Crisis

23. Pernille Schnoor
Brand Ethos
Om troværdige brand- og virksomhedsidentiteter i et retorisk og diskursteoretisk perspektiv

24. Sidsel Fabech
Von welchem Österreich ist hier die Rede?
Diskursive forhandlinger og magtkampe mellem rivaliserende nationale identitetskonstruktioner i østrigske pressediskurser

25. Klavs Odgaard Christensen
Sprogpolitik og identitetsdannelse i flersprogede forbundsstater
Et komparativt studie af Schweiz og Canada

26. Dana B. Minbaeva
Human Resource Practices and Knowledge Transfer in Multinational Corporations

27. Holger Højlund
Markedets politiske fornuft
Et studie af velfærdens organisering i perioden 1990-2003

28. Christine Mølgaard Frandsen
A.s erfaring
Om mellemværendets praktik i en transformation af mennesket og subjektiviteten

29. Sine Nørholm Just
The Constitution of Meaning
– A Meaningful Constitution?
Legitimacy, identity, and public opinion in the debate on the future of Europe

2005
1. Claus J. Varnes
Managing product innovation through rules – The role of formal and structured methods in product development

2. Helle Hedegaard Hein
Mellem konflikt og konsensus
– Dialogudvikling på hospitalsklinikker

3. Axel Rosenø
Customer Value Driven Product Innovation – A Study of Market Learning in New Product Development

4. Søren Buhl Pedersen
Making space
An outline of place branding

5. Camilla Funck Ellehave
Differences that Matter
An analysis of practices of gender and organizing in contemporary workplaces

6. Rigmor Madeleine Lond
Styring af kommunale forvaltninger

7. Mette Aagaard Andreassen
Supply Chain versus Supply Chain
Benchmarking as a Means to Managing Supply Chains

8. Caroline Aggestam-Pontoppidan
From an idea to a standard
The UN and the global governance of accountants' competence

9. Norsk ph.d.

10. Vivienne Heng Ker-ni
An Experimental Field Study on the

12. Claus Bajlum
*Essays on Credit Risk and
Credit Derivatives*

13. Anders Bojesen
*The Performative Power of Competen-
ce – an Inquiry into Subjectivity and
Social Technologies at Work*

14. Satu Reijonen
*Green and Fragile
A Study on Markets and the Natural
Environment*

15. Ilduara Busta
*Corporate Governance in Banking
A European Study*

16. Kristian Anders Hvass
*A Boolean Analysis Predicting Industry
Change: Innovation, Imitation & Busi-
ness Models
The Winning Hybrid: A case study of
isomorphism in the airline industry*

17. Trine Paludan
*De uvidende og de udviklingsparate
Identitet som mulighed og restriktion
blandt fabriksarbejdere på det aftaylo-
riserede fabriksgulv*

18. Kristian Jakobsen
*Foreign market entry in transition eco-
nomies: Entry timing and mode choice*

19. Jakob Elming
*Syntactic reordering in statistical ma-
chine translation*

20. Lars Brømsøe Termansen
*Regional Computable General Equili-
brium Models for Denmark
Three papers laying the foundation for
regional CGE models with agglomera-
tion characteristics*

21. Mia Reinholt
*The Motivational Foundations of
Knowledge Sharing*

22. Frederikke Krogh-Meibom
*The Co-Evolution of Institutions and
Technology
– A Neo-Institutional Understanding of
Change Processes within the Business
Press – the Case Study of Financial
Times*

23. Peter D. Ørberg Jensen
*OFFSHORING OF ADVANCED AND
HIGH-VALUE TECHNICAL SERVICES:
ANTECEDENTS, PROCESS DYNAMICS
AND FIRMLEVEL IMPACTS*

24. Pham Thi Song Hanh
*Functional Upgrading, Relational
Capability and Export Performance of
Vietnamese Wood Furniture Producers*

25. Mads Vangkilde
*Why wait?
An Exploration of first-mover advanta-
ges among Danish e-grocers through a
resource perspective*

26. Hubert Buch-Hansen
*Rethinking the History of European
Level Merger Control
A Critical Political Economy Perspective*

2009
1. Vivian Lindhardsen
*From Independent Ratings to Commu-
nal Ratings: A Study of CWA Raters'
Decision-Making Behaviours*

2. Guðrið Weihe
*Public-Private Partnerships: Meaning
and Practice*

3. Chris Nøkkentved
*Enabling Supply Networks with Colla-
borative Information Infrastructures
An Empirical Investigation of Business
Model Innovation in Supplier Relation-
ship Management*

4. Sara Louise Muhr
*Wound, Interrupted – On the Vulner-
ability of Diversity Management*

integration in the process of product innovation

19. Throstur Olaf Sigurjonsson
Governance Failure and Icelands's Financial Collapse

20. Allan Sall Tang Andersen
Essays on the modeling of risks in interest-rate and inflation markets

21. Heidi Tscherning
Mobile Devices in Social Contexts

22. Birgitte Gorm Hansen
Adapting in the Knowledge Economy Lateral Strategies for Scientists and Those Who Study Them

23. Kristina Vaarst Andersen
Optimal Levels of Embeddedness The Contingent Value of Networked Collaboration

24. Justine Grønbæk Pors
Noisy Management A History of Danish School Governing from 1970-2010

25. Stefan Linder
Micro-foundations of Strategic Entrepreneurship Essays on Autonomous Strategic Action

26. Xin Li
Toward an Integrative Framework of National Competitiveness An application to China

27. Rune Thorbjørn Clausen
Værdifuld arkitektur Et eksplorativt studie af bygningers rolle i virksomheders værdiskabelse

28. Monica Viken
Markedsundersøkelser som bevis i varemerke- og markedsføringsrett

29. Christian Wymann
Tattooing The Economic and Artistic Constitution of a Social Phenomenon

30. Sanne Frandsen
Productive Incoherence A Case Study of Branding and Identity Struggles in a Low-Prestige Organization

31. Mads Stenbo Nielsen
Essays on Correlation Modelling

32. Ivan Häuser
Følelse og sprog Etablering af en ekspressiv kategori, eksemplificeret på russisk

33. Sebastian Schwenen
Security of Supply in Electricity Markets

2012
1. Peter Holm Andreasen
The Dynamics of Procurement Management - A Complexity Approach

2. Martin Haulrich
Data-Driven Bitext Dependency Parsing and Alignment

3. Line Kirkegaard
Konsulenten i den anden nat En undersøgelse af det intense arbejdsliv

4. Tonny Stenheim
Decision usefulness of goodwill under IFRS

5. Morten Lind Larsen
Produktivitet, vækst og velfærd Industrirådet og efterkrigstidens Danmark 1945 - 1958

6. Petter Berg
Cartel Damages and Cost Asymmetries

7. Lynn Kahle
Experiential Discourse in Marketing A methodical inquiry into practice and theory

8. Anne Roelsgaard Obling
Management of Emotions in Accelerated Medical Relationships

TITLER I ATV PH.D.-SERIEN

1992

1. Niels Kornum
Servicesamkørsel – organisation, økonomi og planlægningsmetode

1995

2. Verner Worm
Nordiske virksomheder i Kina
Kulturspecifikke interaktionsrelationer ved nordiske virksomhedsetableringer i Kina

1999

3. Mogens Bjerre
Key Account Management of Complex Strategic Relationships
An Empirical Study of the Fast Moving Consumer Goods Industry

2000

4. Lotte Darsø
Innovation in the Making
Interaction Research with heterogeneous Groups of Knowledge Workers creating new Knowledge and new Leads

2001

5. Peter Hobolt Jensen
Managing Strategic Design Identities
The case of the Lego Developer Network

2002

6. Peter Lohmann
The Deleuzian Other of Organizational Change – Moving Perspectives of the Human

7. Anne Marie Jess Hansen
To lead from a distance: The dynamic interplay between strategy and strategizing – A case study of the strategic management process

2003

8. Lotte Henriksen
Videndeling
– om organisatoriske og ledelsesmæssige udfordringer ved videndeling i praksis

9. Niels Christian Nickelsen
Arrangements of Knowing: Coordinating Procedures Tools and Bodies in Industrial Production – a case study of the collective making of new products

2005

10. Carsten Ørts Hansen
Konstruktion af ledelsesteknologier og effektivitet

TITLER I DBA PH.D.-SERIEN

2007

1. Peter Kastrup-Misir
Endeavoring to Understand Market Orientation – and the concomitant co-mutation of the researched, the re searcher, the research itself and the truth

2009

1. Torkild Leo Thellefsen
Fundamental Signs and Significance effects
A Semeiotic outline of Fundamental Signs, Significance-effects, Knowledge Profiling and their use in Knowledge Organization and Branding

2. Daniel Ronzani
When Bits Learn to Walk Don't Make Them Trip. Technological Innovation and the Role of Regulation by Law in Information Systems Research: the Case of Radio Frequency Identification (RFID)

2010

1. Alexander Carnera
Magten over livet og livet som magt
Studier i den biopolitiske ambivalens