Incentive Engineering for Collaborative Online Work: The Case of Crowdsourcing

Zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für

Wirtschaftswissenschaften

am Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Tim Jochen Straub

Tag der mündlichen Prüfung: 15.03.2017

Referent: Prof. Dr. Christof Weinhardt

Korreferent: Prof. Dr. Orestis Terzidis

Karlsruhe, 2017

Contents

List of Abbreviations

A	Auction Treatment	102
CDA	Continuous Double Auction	102
DV	Dependant Variable	56
FE	Feedback on a Equal Competitor	66
FM	Feedback on a Mediocre Competitor	59
Fr	Framing Setting	101
FS	Feedback on a Strong Competitor	59
FW	Feedback on a Weak Competitor	59
FZI	FZI Research Center for Information Technology	103
GLS	Generalized Least Squares	89
HITs	Human Intelligence Tasks	25
IND	Individual Treatment	83
IS	Information Systems	18
KD²Lab	Karlsruhe Design & Decision Lab	103
KIT	Karlsruher Institute of Technology	83
Mo	Money Setting	101
MTurk	Amazon Mechanical Turk	4
NA	No Auction Treatment	102
NF	No Feedback	59
NoFr	No Framing Setting	101
NoMo	No Money Setting	101
OLS	Ordinary Least Squares	56
ORSEE	Online Recruitment System for Economic Experiments	83
PR	Piece Rate	6
ROT	Rank-Order-Tournament	6
SD	Standard Deviation	55
TDR	Tradable Development Right	97
TEAM	Team Treatment	83
USBA	Unified Sealed Bid Auction	102
USD	United States Dollar	20
WWW	World Wide Web	17

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Introduction

> “The world is becoming too fast, too complex and too networked for any company to have all the answers inside.”
>
> YOCHAI BENKLER, 2006

1.1 Motivation

ONLINE collaboration and in particular crowdsourcing services are a new mechanism to share information and goods, participate in political debates, and even work on small to complex tasks in a digital space. Information Technology (IT) enables people to collaborate worldwide via internet access from their own computer, wherever they want. Hence, it is now easier than ever to participate, collaborate, exchange, and work as entry barriers are lowered given the easy and often anonymous access via the internet. Such new services pave the pathway for our future interactions and a potential pathway of how we are going to work in the future.

Many services and platforms, which were developed in the past century, proved how disruptive online collaboration, particularly crowdsourcing, is. Nowadays we share information in private and professional social networks and blogs, communicate via online

messengers, take part in political decisions via online polls or participatory budgeting programs, fund companies on crowdfunding platforms, develop new products in open innovation contests, and even work online on crowdsourcing platforms. On such crowdsourcing platforms, which are sometimes called online labor markets or crowd work platforms, such as Amazon Mechanical Turk (MTurk)[1], crowd workers from all over the world work on small and mostly simple tasks for a small monetary compensation. Tasks are issued from a set of requesters, usually companies, which most of the time break down tasks into small subsets. Hence, one essential property of such tasks is that the overall goal can be divided into subtasks. These so called micro-tasks can then be completed by one or several crowd workers. Since each of these micro-tasks is usually executed quickly the monetary compensation is comparatively low. After the completion of all micro-tasks, the requester aggregates all finished instances of the overall task (see Chapter 2 for more details).

These recently developed online collaboration and crowdsourcing platforms offer a huge potential for companies, practitioners, and researchers. With current internet technology it is now theoretically possible to reach a huge working force without high costs. Crowdsourcing uses this available human capital – the crowd – and offers work for a small price or sometimes even for free to this large pool of people willing to work (Ipeirotis, 2010a; Leimeister, 2010; Paolacci et al., 2010). Recently, an increasing number of companies foster this potential by crowdsourcing work which was previously done in-house. Based on the demand, a variety of platforms evolved not only ranging from simple tasks, such as transcription of receipts, but to complex tasks, such as building logos on platforms like 99designs[2] (Hammon and Hippner, 2012; Kittur et al., 2012, 2013). Besides using "standard" crowdsourcing platforms, which are organized by different online vendors (e.g., MTurk[1]) and offer access to various crowds willing to work on these specific tasks, companies sometimes design their own custom platform. These custom platforms face the challenge of creating and incentivizing their own crowd (see Chapter 2 for more details on crowdsourcing and its related concepts).

The design of such an online platform plays a pivotal role in how well these crowdsourcing applications perform. First, such platforms need to attract many participants to build a crowd and foster the available human capital worldwide. The crowd is the most important factor, as the whole concept of crowdsourcing is built around the cornerstone

[1] `https://www.mturk.com/mturk/welcome` last accessed: September 6, 2016.
[2] `http://en.99designs.de/` last accessed: September 6, 2016.

of having access to such a working force (see Ipeirotis, 2010a; Leimeister, 2010; Paolacci et al., 2010). In some cases a crowd needs to consist of experts (e.g., designers), while in other cases a wide range of user profiles is preferred. All of them need to be attracted by adequate incentives.

Second, the platform must facilitate and attract fast execution to please the companies' desire to receive results quickly. Finally, the platform must provide a certain quality standard of the results, such that requestors' expectations are met. All these challenges target the crowd workers' behavior: (i) crowd workers decide on whether to participate in a crowd, (ii) execute the tasks quickly, and (iii) work carefully to secure the result quality. A crowd worker's behavior is known to be affected by the design of a platform or market (Weinhardt et al., 2003; Weinhardt and Gimpel, 2007; Gimpel et al., 2008). Additionally, most of the design challenges of crowdsourcing platforms are at least partially affected by incentives, since incentives increase the likelihood that someone is motivated to work on a task (Ryan and Deci, 2000; Rheinberg, 2008; Schattke, 2011). Hence, finding the correct incentive for a crowdsourcing application is one of the biggest challenges in this field (Shaw et al., 2011; Wang et al., 2013). Crowdsourcing platforms use a wide variety of incentive and feedback schemes. Crowd workers work for free on platforms such as Wikipedia[3] and Google ReCAPTCHA[4]. On platforms such as MTurk[1] and clickworker[5] crowd workers are paid for each finished (micro-) task. In contrast, platforms such as 99designs[2] and threadless[6] use tournament prizes with different implementations of giving feedback within the tournament to attract crowd workers.

The work at hand stresses the importance of finding the correct incentives for a given crowdsourcing scenario by developing a deeper understanding of crowd workers' behavior under different incentive and feedback schemes.

[3] `https://www.wikipedia.org/` last accessed: September 6, 2016.
[4] `https://www.google.com/recaptcha/intro/index.html` last accessed: September 6, 2016.
[5] `https://www.clickworker.com/` last accessed: September 6, 2016.
[6] `https://www.threadless.com/` last accessed: September 6, 2016.

1.2 Research Outline

This thesis highlights the importance of incentive engineering for crowdsourcing from a platform designer's view. It focuses on the individual crowd worker's behavior. More precisely, it focuses on the effect of different incentives on a crowd worker's motivation. Incentivizing crowd workers is essential, since a crowdsourcing platform's most important asset is the crowd. A platform designer needs to incentivize participation or a certain behavior depending on the goals the crowd should achieve. In this thesis, therefore, different incentive schemes are evaluated to shed light on the effect of incentives on crowd workers' behavior.

Incentives are set by platform designers to motivate participants. Incentives can lead to motivation if a person in his current situation is attracted by these incentives (Rheinberg, 2008; Schattke, 2011). A motivated person is then more likely to change his behavior, e.g., to work harder or with more precision (see Chapter 3 for more details). Motivation can be either of intrinsic or extrinsic nature. Intrinsic motivation targets a participants' inherent satisfaction and not some separable outcome, as extrinsic motivation does, which is further elaborated in Chapter 3 (Ryan and Deci, 2000). Straightforward incentives a platform designer can adjust are of extrinsic nature, e.g., money and prizes (Ryan and Deci, 2000). The two most common monetary incentive schemes used in crowdsourcing are "piece rates (PR)" and "rank-order-tournaments (ROTs)" (see Chapter 2 for more details). PRs pay per finished work and are mainly used in crowdsourcing small tasks, so-called micro-tasks, on platforms like MTurk[1], clickworker[5], and CrowdFlower[7]. ROTs are mainly used for tasks where it is not feasible to pay for every solution, e.g., creative tasks such as designing logos, t-shirts, and innovative products, on platforms like 99designs[2], threadless[6], and InnoCentive[8]. It is well known that different payment schemes affect a participant's behavior (Lazear and Sherwin, 1981; Bull et al., 1987; Ehrenberg and Bognanno, 1990; Becker and Huselid, 1992; Prendergast, 1999; van Dijk et al., 2001). However, for online and crowdsourcing settings in particular, it is unclear which incentive scheme motivates crowd workers better. Hence, a field experiment on the most common micro-task crowdsourcing platform – MTurk – comparing PR and ROT incentive schemes is implemented (see Chapter 4). This approach delivers a comprehensive insight in crowd workers' behavior on a

[7] `http://www.crowdflower.com/` last accessed: September 6, 2016.
[8] `https://www.innocentive.com/` last accessed: September 6, 2016.

crowdsourcing platform. The amount of effort a crowd worker puts into a task depending on the incentive scheme, i.e., how often he finishes a task given a certain time frame, is evaluated. In particular, the following research question is addressed:

> **Research Question 1:** *Which incentive scheme – piece rates or rank-order-tournaments – incentivizes crowd workers' effort best?*

After comparing the two incentive schemes, the question arises if ROTs can be modified in order that crowd workers exert more effort. Most ROTs standardly show a ranking to give feedback about a crowd worker's current position. However, this information may encourage or discourage a crowd worker as it discloses chances to win. Crowd workers who are far off of a winning position might be discouraged, while crowd workers close to winning positions might be encouraged to give their best and close the gap to payout positions (see Mazarakis et al., 2011). By customizing this feedback and only showing certain competitors of different strengths, a crowd worker's motivation might be positively affected (see Mazarakis and Krämer, 2010; Mazarakis and Van Dinther, 2011) (see Chapter 4). Therefore, different feedback schemes in a ROT and their effect on crowd workers' effort are evaluated on MTurk regarding the following research question:

> **Research Question 2:** *Does feedback in rank-order-tournaments affect crowd workers' effort in crowdsourcing settings?*

After examining a crowd worker's effort in ROTs with and without feedback, risk-taking in ROTs is analyzed. Besides an adjustment of feedback, ROTs offer many more design options. It might be that in ROTs risk-taking is influenced by the tournament's design and the feedback mechanism (see van Dijk et al., 2001; Eriksson et al., 2009b). Participants on lower ranks might take higher risks to close the gap to payout ranks or risk attitudes of participants competing in teams might differ from participants competing individually. However, studies for crowd tournaments are missing, although risk-taking might be beneficial for certain crowdsourcing applications. Especially for creative crowdsourcing tasks, who seek extraordinary solutions, designs exerting risk-taking behavior might result in a variety of more diverse and non-standard solutions. High risk propensity might affect crowd workers to avoid conformist solutions. For other tasks, where non-creative, but standard solutions are seeked, risk aversion might stimulate higher quality. Thereby it is important for platform designers to shed light on how their ROT setup influences crowd workers'

risk-taking. Hence, different ROT setups and their influence on risk-taking in a longitudinal crowdsourcing setup are analyzed. In particular an individual ranking, where each crowd worker competes on his own, and a team ranking, where teams compete against other teams, are studied (see Chapter 5). Hence, the following research question with its three sub-questions are analyzed:

> **Research Question 3:** *How is risk-taking behavior affected in long lasting rank-order-tournaments?*
>
> > **Research Question 3.1:** *Does the tournament mode (teams or individuals) affect risk-taking behavior?*
> >
> > **Research Question 3.2:** *Does the tournament progress affect risk-taking behavior?*
> >
> > **Research Question 3.3:** *Does the ranking position affect risk-taking behavior?*

Besides monetary incentives and different setups within these payment schemes, crowdsourcing often is not based on offering a monetary compensation as incentive for crowd workers (Hammon and Hippner, 2012). Many prominent crowdsourcing examples like Wikipedia[3], Google ReCAPTCHA[4], Stack Overflow[9], and OpenIDEO[10] show that a crowd can be motivated to participate without a payment. For a platform designer it is therefore beneficial to understand the motives behind such non-monetary incentives. It is well known from classical motivational theory that besides extrinsic motivation, intrinsic motivation is a strong driver for participation and effort (Ryan and Deci, 2000). From a platform designer's perspective it is hard to target intrinsic motivation with incentives, since it is driven by the personal interest of the crowd worker. However, it might be beneficial to remind a crowd worker about his initial intrinsic motivation or to give insights about the altruistic goals of the platform by a (customized) framing. Hence, the influence of framing on crowd workers' behavior is analyzed in the context of a German government financed project – the "Planspiel Flächenhandel[11]" (see Chapter 6). Moreover, it is of interest for

[9] `http://stackoverflow.com/` last accessed: September 6, 2016.

[10] `http://www.openideo.com/` last accessed: September 6, 2016.

[11] The "Planspiel Flächenhandel" evaluates a cap and trade system for tradeable development rights. For more details see Subsection 6.2.2 of Chapter 6, and the project's homepage `http://www.flaechenhandel.de/` last accessed: September 6, 2016.

platform designers how such a framing in non-monetary settings performs compared to standardly incentivized settings. It might be that a (customized) framing in non-monetary settings induces similar behavior as in monetarily incentivized settings. This would provide a strong tool for platform designers. In particular the following research questions are analyzed:

> **Research Question 4:** *Is framing capable of influencing behavior in non-monetary settings towards a certain goal?*
>
> **Research Question 5:** *Can a customized framing for a non-monetary setting lead to similar behavior as in monetarily incentivized settings?*

1.3 Structure of the Thesis

The work at hand is structured in four parts as shown in Figure 1.1. Part I motivates this thesis (Section 1.1 of Chapter 1), presents the research outline and the research questions (Section 1.2 of Chapter 1), gives an overview of this thesis' structure (Section 1.3 of Chapter 1), and presents the research development process (Section 1.4 of Chapter 1).

Part II introduces and defines collaborative online work and in particular crowdsourcing as a use case scenario of this work (Chapter 2). Furthermore it presents foundations of motivational theory (Chapter 3).

Part III consists of the three main chapters of this thesis and presents insights about collaborative online work, most importantly crowdsourcing platforms. Chapter 4 evaluates how different payment schemes and giving feedback about the competitor's strength affects a crowd worker's behavior (thereby addressing Research Questions 1 and 2). Further evaluating a crowd worker's behavior, Chapter 5 analyzes how risk-taking behavior is influenced by different tournament designs (thereby addressing Research Question 3 and its sub-questions 3.1, 3.2, and 3.3). Chapter 6 evaluates if crowds in collaborative online work scenarios without monetary incentives can still be incentivized with a framing, even when they work for free (thereby addressing Research Questions 4 and 5)

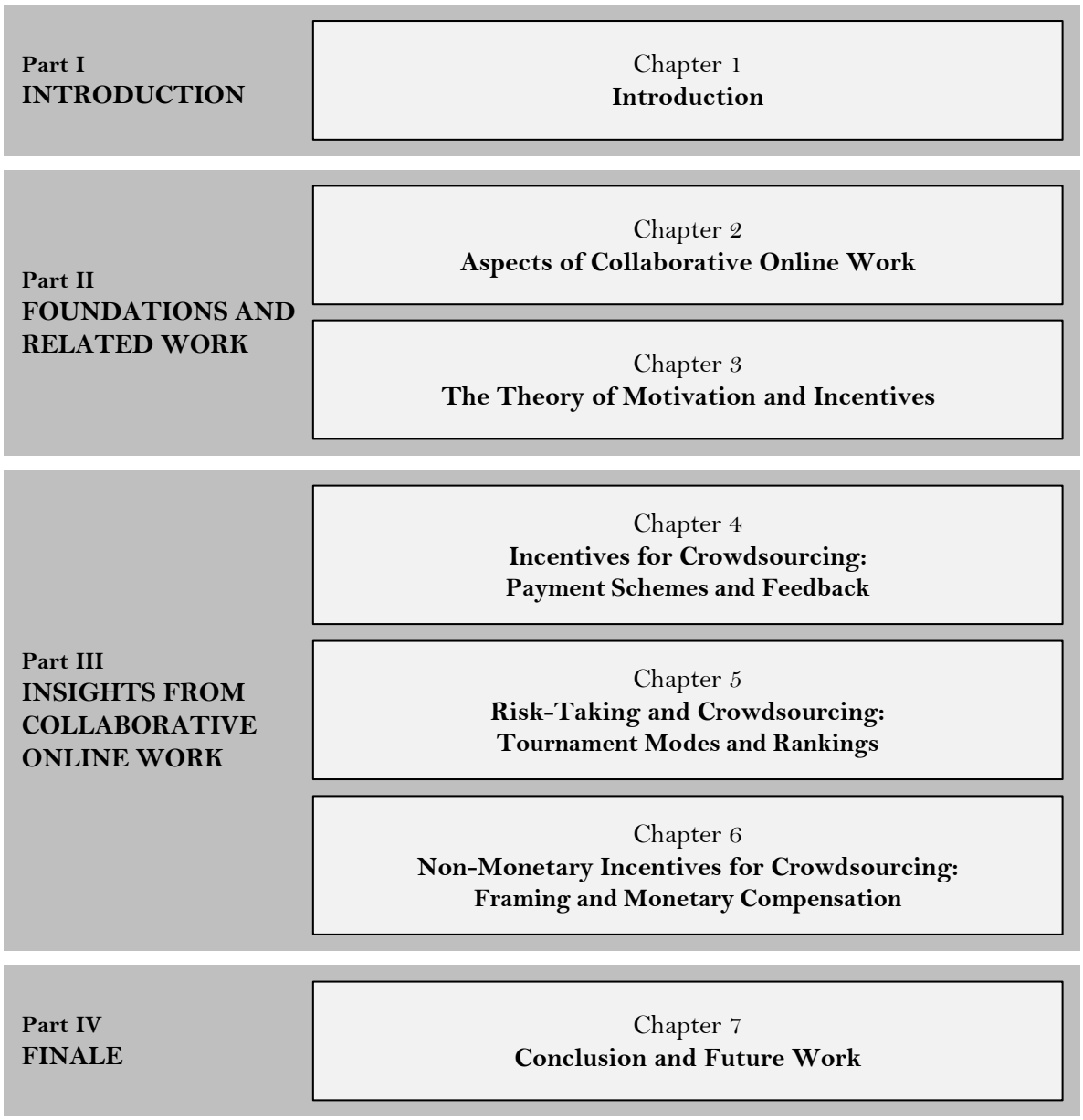


FIGURE 1.1: *Structure of this thesis.*

Part IV concludes this thesis by summarizing the key contributions made in the context of collaborative online work and crowdsourcing (Section 7.1 of Chapter 7). Furthermore, it outlines promising pathways for future research (Section 7.2 of Chapter 7).

Part V contains the appendices.

1.4 Research Development

Parts of the work at hand have been accepted for publication and have been presented at European and international conferences. In addition, parts have been published as articles in double-blind reviewed journals. This section serves to give an overview of this development.

Chapter 4 evaluates the effect of different payment schemes and feedback on a crowd worker's motivation. It was developed in the process of three publications. The basic setup consisted of two extreme feedback schemes as proof of an effect of feedback on a crowd worker's motivation. It was published in the proceedings of the international conference Collective Intelligence 2014 as joint work with Henner Gimpel and Florian Teschner (Straub, Gimpel, and Teschner, 2014). It was ranked six among the top ten recent papers on labor markets as listed on the Social Science Research Network after publication in April 9^{th} 2014. After establishing the initial proof, extensions with more sophisticated feedback mechanisms and theory development were made, which allow a more detailed evaluation of the effects. Hence, an additional extended experiment was conducted targeting the extreme feedbacks and a new mediocre feedback. This revised and extended version of the paper was published at the 22^{nd} European Conference on Information Systems (ECIS 2014) in Tel Aviv, Israel as joint work with Henner Gimpel, Florian Teschner, and Christof Weinhardt (Straub, Gimpel, Teschner, and Weinhardt, 2014). To further strengthen the findings and to integrate the feedback from the two conferences, two new experiments were conducted. Thereby, another feedback mechanism was evaluated and a comparison between the two payment schemes without feedback was evaluated. These three experiments were then published as extensive picture about Research Questions 1 and 2 published in the special issue Computer Supported Cooperative Work & Social Computing of the journal Business & Information Systems Engineering as joint work with Henner

Gimpel, Florian Teschner, and Christof Weinhardt (Straub et al., 2015).

Chapter 5 discusses the effects of different ROT setups on a crowd worker's risk-taking behavior. It complements Chapter 4 by looking at a different aspect – risk-taking – of the user's behavior in crowd tournaments and targets Research Question 3 and its sub-questions 3.1, 3.2, and 3.3. The evaluation was implemented as an online experiment accompanying the FIFA World Cup of 2014. It was presented at the 49^{th} Hawaii International Conference on System Sciences (HICSS 49) as joint work with Timm Teubner and Christof Weinhardt (Straub, Teubner, and Weinhardt, 2016) and was nominated for the best paper award.

Chapter 6 was developed between 2014 and 2016, during the course of the project "Planspiel Flächenhandel" funded by the German Federal Environment Agency. In this project the overall goal of testbedding a cap and trade system for land consumption was implemented and experimentally evaluated with local authorities. Results of these experiments were contributed to the project, including a publication in the book "IÖR Schriften Nr. 69: Flächennutzungsmonitoring VIII" and several result reports, which are accessible on the project's homepage[12]. The book publication was created as joint work with Ralph Henger, Kilian Bizer, Lutke Blecken, Katrin Fahrenkrug, Uwe Ferber, Jens-Martin Gutsche, Tobias Kranz, Michael Melzer, Lukas Meub, Till Proeger, Stefan Siedentop, Tom Schmidt, Achim Tack, and Christof Weinhardt (Henger et al., 2017). Parts of the design and conceptualization of the project context are described in Subsection 6.2.2 and Section 6.3. They were developed as joint work with Tobias Kranz, Timm Teubner, Christof Weinhardt, Ralph Henger, and Michael Schier. In the work at hand, a different angle is taken, accompanying the project's context. Light is shed on the motivational factor of non-monetary incentives, targeting Research Questions 4 and 5. Hence, findings further strengthen the overall picture of incentive engineering in crowdsourcing.

In addition, complementary research in progress in Chapter 7 (Towards a Guideline for Conducting Crowd Experiments) was presented at and published in the proceedings of the Second Karlsruhe Service Summit Research Workshop as joint work with Florian Hawlitschek and Christof Weinhardt (Straub, Hawlitschek, and Weinhardt, 2016). It summarizes lessons learned for experimental economics from the three chapters in Part III.

[12] `http://www.flaechenhandel.de/` last accessed: September 6, 2016.

Particularly conducting online experiments with crowds sometimes needs different approaches as classical experimental economics suggests. Hence, it proposes a way for researchers to collaboratively overcome typical challenges one faces when conducting such online experiments with crowds. However, it is not a main part of this work but ongoing complementary research presented in Section 7.2. Thereby it gives an outlook and a possible direction for future work.

Part II

Foundations and Related Work

Chapter 2

Aspects of Collaborative Online Work

> “ [...T]his is not an information age. It’s an age of networked intelligence, it’s an age of vast promise, an age of collaboration,[...].”
>
> DON TAPSCOTT, 2012

COLLABORATIVE online work is more present than ever. Starting with the Web 2.0 development in the early 2000’s, the World Wide Web (WWW) has changed (Knorr, 2003; O’Reilly, 2006). Before, a set of web developers produced content while internet users consumed. With the Web 2.0 development the user changed from a consumer to a prosumer, who develops and uses content at the same time.

Since then many concepts originated how people collaborate and share information, knowledge, creativity, money, and even work (see Malone et al., 2010; Leimeister, 2010; Hammon and Hippner, 2012; Prpić, 2016). Blogging is the concept of web-users creating their own articles in the web (e.g., blogger[1], WordPress[2], etc.) (see Nardi et al., 2004). Tagging refers, for example, to labeling images to index them for search engines (e.g., Flickr[3], YouTube[4], etc.) (see Golder and Huberman, 2006; Malone et al., 2010; Lee

[1] `https://www.blogger.com/` last accessed: September 6, 2016.
[2] `https://wordpress.org/` last accessed: September 6, 2016.
[3] `https://www.flickr.com/` last accessed: September 6, 2016.
[4] `https://www.youtube.com/` last accessed: September 6, 2016.

et al., 2012). Wikis are a tool for collaboratively sharing and managing content (e.g., Wikipedia[5], SharePoint[6], etc.) (see Oreg and Nov, 2008; Malone et al., 2010; Mazarakis and Van Dinther, 2011; Lee et al., 2012). With social media sites, people can build their own networks, share content publicly or privately, and keep in touch with friends and share personal thoughts (e.g., Facebook[7], twitter[8], etc.) (see Röll, 2010; Wilson et al., 2012; Boecking et al., 2015). Through crowdfunding, internet users can invest money in ideas and products (e.g., Kickstarter[9], Indiegogo[10], etc.) (see Hammon and Hippner, 2012; Feldmann and Gimpel, 2016). Open innovation contests outsource new product developments to users (e.g., Tchibo ideas[11], OpenIDEO[12], etc.) (see Chesbrough, 2006; Terwiesch and Xu, 2008). Using prediction markets, a combination of informed and uninformed traders predict future events by trading stocks and therefore shed light on the likelihood of future events (e.g., Economic Indicator eXchange[13], Iowa Electronic Markets[14], etc.) (see Spann and Skiera, 2003; Teschner et al., 2011; Kranz et al., 2014; Kloker et al., 2016). Last, (crowd) online labor markets refers to outsourcing work to internet users (e.g., MTurk[15], clickworker[16], etc.) (see Ipeirotis, 2010a; Mason and Suri, 2012).

Different research domains take certain angles to tackle this ever-changing development. For example, the service research domain focuses on different conceptual roles customers can take when they are integrated in the processes of companies and the benefits for the company (see Straub et al., 2013; Kunze von Bischhoffshausen et al., 2015). The Information Systems (IS) domain focuses on the impact on electronic markets, how to create crowds, how to incentivize crowds (see for example Chapters 4, 5, and 6), how to secure quality, and how this potential workforce can be used in different setups (Ipeirotis, 2010a; Ipeirotis et al., 2010; Leimeister, 2010; Paolacci et al., 2010; Shaw et al., 2011; Hammon and Hippner, 2012; Wang et al., 2013).

[5] https://www.wikipedia.org/ last accessed: September 6, 2016.
[6] https://products.office.com/en-us/sharepoint/ last accessed: September 6, 2016.
[7] https://www.facebook.com/ last accessed: September 6, 2016.
[8] https://twitter.com/ last accessed: September 6, 2016.
[9] https://www.kickstarter.com/ last accessed: September 6, 2016.
[10] https://www.indiegogo.com/ last accessed: September 6, 2016.
[11] https://www.tchibo-ideas.de/ last accessed: September 6, 2016.
[12] https://openideo.com/ last accessed: September 6, 2016.
[13] http://www.eix-market.de/ last accessed: September 6, 2016.
[14] http://tippie.uiowa.edu/iem/ last accessed: September 6, 2016.
[15] https://www.mturk.com/mturk/welcome last accessed: September 6, 2016.
[16] https://www.clickworker.com/ last accessed: September 6, 2016.

Recently the term crowdsourcing became prominent and is used for many of these Web 2.0 approaches (Howe, 2006a,b). Specifically it is used for concepts who try to harness the human capital available worldwide as working force, i.e. the crowd. The work at hand evaluates how different incentives affect a user's motivation on crowdsourcing platforms. The most important factor of online collaboration and online work in particular is human capital (i.e., the crowd). All of the mentioned concepts only work if a platform can motivate people to participate. Crowdsourcing is used as it targets this factor most clearly. In crowdsourcing there are no users who just consume, like it is the case with blogs or tagging, where users might passively read or profit from better search engines (i.e., consumers), which are enhanced with tagged pictures by other active users (i.e., prosumers) – crowd workers are creating value (Hammon and Hippner, 2012). All crowd workers are users who produce and without them such concepts would not work. Hence, motivation of users is the most important asset in crowdsourcing.

The remainder of this chapter is structured as follows. Section 2.1 introduces and defines the term and concept "crowdsourcing". Subsection 2.1.1 classifies different crowdsourcing applications along different incentive schemes. Subsection 2.1.2 explains several crowdsourcing applications which are later in focus of Chapters 4, 5, and 6. Subsection 2.1.3 delineates the concept from different related collaboration concepts. Section 2.2 then embeds the work at hand in existing works and research streams of collaborative online work, crowdsourcing, and incentives.

2.1 Crowdsourcing

Crowdsourcing and online labor markets (e.g., MTurk[15]) have emerged as new pools of human labor that allow organizations to flexibly scale their workforce and hire experts, typically for a comparatively low price or no price at all (Leimeister, 2010; Paolacci et al., 2010; Hammon and Hippner, 2012). The neologism was created by Howe (2006a,b) as "[...t]he act of a company or institution taking a function once performed by employees and outsourcing it to an undefined network of people in the form of an open call." This definition specifies the key aspects of crowdsourcing in a clear manner and is in line with other definitions of the term crowdsourcing (cf. Estellés and González, 2012; Hammon and Hippner, 2012).

Crowdsourcing is the combination of the two terms "crowd" and "outsourcing" (Rouse, 2010). The concept is similar to classical "outsourcing" with several key differences. First, in classical outsourcing one company outsources tasks once performed in-house to another company (Lacity and Hirschheim, 1993). Therefore it is clear to whom the company outsources its tasks (Rouse, 2010). In crowdsourcing a company or any other individual outsources the task to "[...a]n undefined network of people[...]" – the crowd (Howe, 2006a,b; Rouse, 2010). Therefore they do not know to whom they are outsourcing to. Second, crowdsourcing is an "open call". Hence, it is not clear when the task will be finished or at which level of quality. Normally in outsourcing you have information about the company to whom you are outsourcing your tasks (Rouse, 2010). Hence, you can gather some details about this company (e.g., reputation), which sheds light on these uncertainties. In crowdsourcing this is normally not the case. Third, while in standard outsourcing contracts between the two respective firms exist, they are missing in crowdsourcing (Marjanovic et al., 2012). This makes it almost impossible to embed expectations about the result.

However, besides these key differences to classical outsourcing, crowdsourcing offers great potential for companies, practitioners, and researchers. First, crowdsourcing offers a large potential pool of workers – the crowd (Paolacci et al., 2010; Mason and Suri, 2012). In theory these are all people with internet access. However, they need to be motivated to work (see Chapter 3 on more details about motivation). Some emerged markets are collecting large pools of online workers which are willing to work for a small monetary compensation, like for example MTurk[15], clickworker[16], or 99designs[17], which have up to 850,000 workers[18](Ipeirotis, 2010a). Second, the crowd is almost 24/7 available for work, since they are working from around the globe. Hence, tasks have the potential to be finished fast, when they are attractive for crowd workers, since a subset of the crowd can always work on tasks (Ipeirotis, 2010a). Third, paid online labor is cheap (Leimeister, 2010; Paolacci et al., 2010). The hourly reservation wage is as low as United States dollar (USD) 1.38 while the effective wage is as low as USD 4.80 (Horton and Chilton, 2010; Ipeirotis, 2010a).

[17] `http://en.99designs.de/` last accessed: September 6, 2016.

[18] Note that numbers are mainly based on information from the companies' websites. Amazon Mechanical Turk (MTurk) 500,000 (`https://requester.mturk.com/tour` last accessed: September 6, 2016), clickworker 700,000 (`https://www.clickworker.com/en/about-us/` last accessed: September 6, 2016), and 99designs 850,000 (`http://99designs.de/about/press-releases`)

Crowdsourcing offers opportunities for crowd workers as well. First, crowd workers can work from wherever and whenever they want, as long as they have a computer with an internet connection (Terranova, 2004; Howe, 2006a,b; Brabham, 2008). This offers a flexible work life. Second, crowd workers can make some extra cash or base their living on crowd work. Crowd workers can earn more than USD 1,000 a week working on MTurk (Ipeirotis, 2010b). 12 % percent of U.S. workers and 27 % of Indian workers even use MTurk as their primary source of income (Ipeirotis, 2010b). Third, crowdsourcing offers the potential to use arbitrage effects. Since crowd workers are working from all over the world and requesters (i.e., companies) offer tasks from all over the world as well, the offered wages are often attractive. For example, an effective wage of USD 4.80 per hour might be higher than the average wages in some regions (Horton and Chilton, 2010; Ipeirotis, 2010a). This effect can be seen on MTurk where 34 % of workers are from India. The average yearly wage of an Indian person is roughly at USD 1,500[19]. Fourth, crowd workers can choose which tasks they want to work on (Mason and Suri, 2012). Hence, they only have to work on tasks which they prefer, which is a clear distinction to standard work contracts.

In the past few years, companies have initiated many crowdsourcing activities in order to benefit from the knowledge and potential of the crowd (e.g., 99designs[17], threadless[20], MTurk[15], clickworker[16], CrowdFlower[21], InnoCentive[22], Wikipedia[5], Google ReCAPTCHA[23]). These platforms show the great potential of rightly tuned crowdsourcing applications, which incentives, goals, and collaboration structures attract large crowds. Moreover, researchers started to harness the potential of the crowd as well. Experimental researchers increasingly started using MTurk as low cost participant pool (Chilton et al., 2010; Hall and Caton, 2014). Furthermore, previous work has examined crowdsourcing regarding its requirements for research (Schlagwein and Farhad, 2014), validity and costs (Chilton et al., 2010), and crowd worker demographics (Paolacci et al., 2010; Berinsky et al., 2012). See Mason and Suri (2012), Horton et al. (2011), Kaufmann et al. (2011), Pilz and Gewald (2013), and Teschner and Gimpel (2013) for recent examples.

[19] Note that this number is based on statistics from the world bank in 2013 (`http://data.worldbank.org/indicator/NY.GDP.PCAP.CD` last accessed: September 6, 2016).

[20] `https://www.threadless.com/` last accessed: September 6, 2016.

[21] `http://www.crowdflower.com/` last accessed: September 6, 2016.

[22] `https://www.innocentive.com/` last accessed: September 6, 2016.

[23] `https://www.google.com/recaptcha/intro/index.html` last accessed: September 6, 2016.

2.1.1 Classification of Crowdsourcing

Crowdsourcing can be classified by means of several dimensions (Geiger et al., 2012). Crowdsourcing could be classified at hand of the type of task (Corney et al., 2009; Rouse, 2010; Doan et al., 2011; Brabham, 2012; Geiger et al., 2012; Prpić et al., 2015; Prpić, 2016), who is working on the tasks (Corney et al., 2009; Malone et al., 2010; Rouse, 2010; Zwass, 2010; Doan et al., 2011; Geiger et al., 2012), which control system or quality mechanism is used (Allahbakhsh et al., 2013; Saxton et al., 2013), or which incentive structure is used (Leimeister et al., 2009; Horton and Chilton, 2010; Malone et al., 2010; Rogstadius et al., 2011; Hammon and Hippner, 2012; Chandler and Kapelner, 2013; Kittur et al., 2013). The focus of this thesis is on incentives for crowdsourcing. Hence, we concentrate on classifying crowdsourcing along this dimension.

A prominent classification by Prpić et al. (2015) identifies four main crowdsourcing alternatives along two dimensions as depicted in Figure 2.1. The first dimension separates crowdsourcing alternatives along its content. Crowdsourcing content can either be subjective or objective. Subjective content has no right or wrong answer but rather is based on an individuals opinion, while objective content leads to an unbiased result, so to speak a fact (Prpić et al., 2015; Prpić, 2016). The second dimension differentiates crowdsourcing based on how the contributions are processed by the requester once crowd workers finished the task(s). Aggregated contributions collectively generate the desired value, while other contributions must be filtered based on their quality to secure their value for the requester (Prpić et al., 2015; Prpić, 2016). Based on these two dimensions, four main crowdsourcing alternatives can be identified. First, in crowd-voting crowd workers collectively decide between alternatives by each voting for their subjective best option. Second, micro-task crowdsourcing divides a big task into small subsets of repeatable micro-tasks which can then be worked on individually (see Subsection 2.1.2 of this chapter for an example of such a platform). Afterwards the individual contributions must be aggregated from the requester, in order that the overall task is solved. Third, in idea crowdsourcing the creativity of the crowd is desired. Since the requester desires a unique and diverse solution the contributions must be filtered afterwards, in order that the contributions meet subjective expectations. Fourth, in contrast to idea crowdsourcing, solution crowdsourcing has a well described problem which can be objectively solved. The contribution either solves the problem, then the content is correct, or not (Prpić et al., 2015; Prpić, 2016).

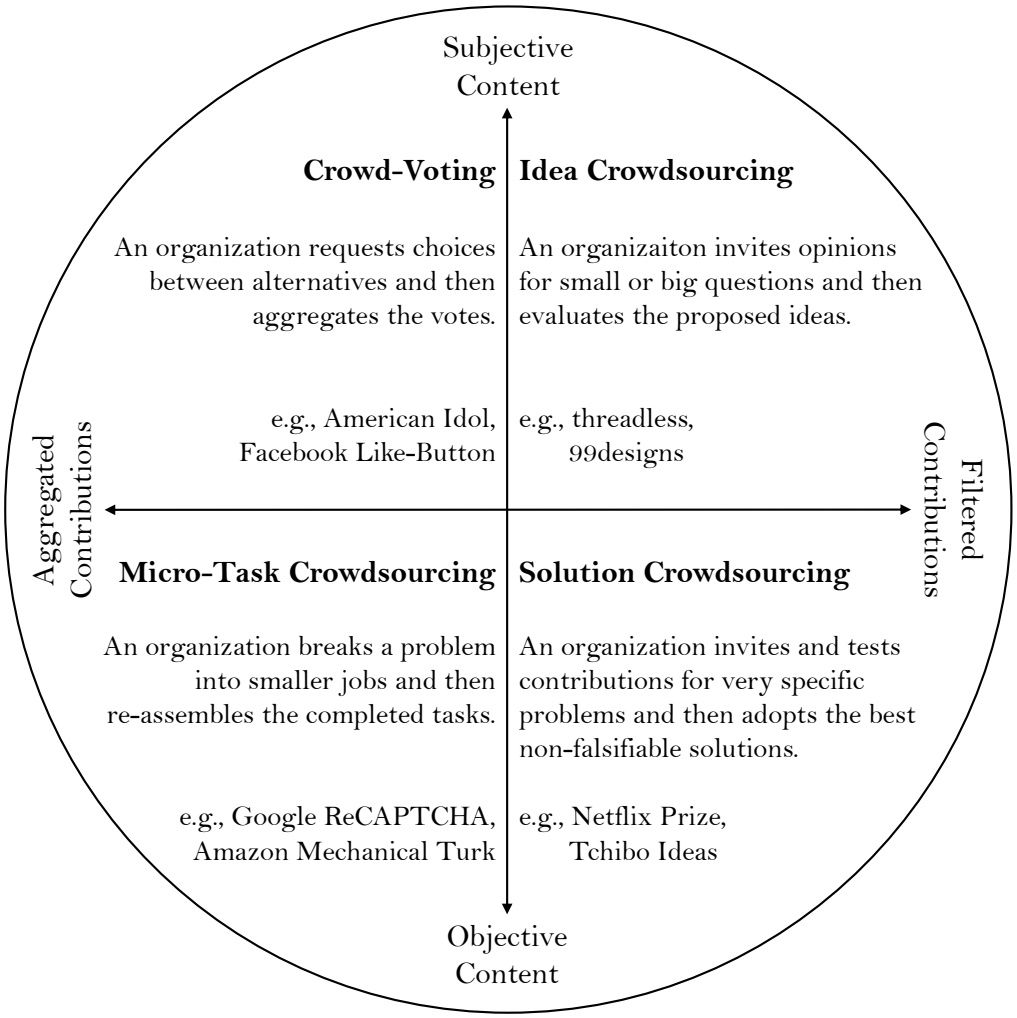


FIGURE 2.1: *Classification of crowdsourcing into four main alternatives by Prpić et al. (2015), extended with further examples.*

This classification gives an exhaustive overview of different crowdsourcing alternatives. However, there is no clear distinction of different applicable and used incentives for crowdsourcing. For example, two prominent examples for micro-task crowdsourcing are Google ReCAPTCHA[23] and MTurk[15]. Google ReCAPTCHA[23] does not offer a monetary compensation but rather enforces participation in order to get access to a certain website, while MTurk[15] offers a monetary compensation to its crowd workers. Hence, we need to classify crowdsourcing along the dimension of incentives in order to have a fuller picture of crowdsourcing.

Following Hammon and Hippner (2012) crowdsourcing platforms can be divided in paid- and non-paid crowdsourcing applications. Paid crowdsourcing is mostly initiated by profit oriented companies submitting their tasks on crowdsourcing markets where a large pool of potential workers are registered to work for money (e.g MTurk[15], CrowdFlower[21], clickworker[16], 99designs[17], threadless[20]). Non-paid crowdsourcing mostly consists of tasks following different goals than earning money. It is mostly initiated from non-profit organizations or from non-profit projects within companies (e.g., Wikipedia[5], Google ReCAPTCHA[23]) (see Subsection 2.1.2 for an exemplary platform description and Chapter 6 for details on how such non-monetary incentives influence user behavior).

Paid-crowdsourcing applications can be further divided by their payment scheme (Malone et al., 2010; Kittur et al., 2013). On the one hand, there are crowdsourcing tournaments that are applied for tasks which are mostly complex and often of creative nature (e.g., 99designs[17], threadless[20]) (see Subsection 2.1.2 for an exemplary platform description and Chapters 4 and 5 for details on how this incentive structure influences user behavior). In such cases it is mostly not suitable to pay every worker, but only one who's submission is accepted and used by the company. Hence, the payment scheme follows a ROT where only one or several contestants get paid in the end. The principle of such ROTs can be briefly summarized as follows: Typically, a task and a set of rewards are announced to attract potential crowd workers who then compete against each other in completing the task. Over the course of the tournament, an optional ranking may inform the crowd workers about their current performance evaluation and (or just) their payoff-relevant position in the ranking. The most successful contributors in the eyes of the issuer or an external jury win a (graduated) set of rewards (Araujo, 2013). On the other hand, there are PR payment schemes. PRs are mostly used for tasks which can be divided into small, simple, and repetitive chunks, so called micro-tasks (e.g., MTurk[15], clickworker[16]) (see Subsection 2.1.2 for an exemplary platform description and Chapter 4 for details on how this incentive structure influences user behavior). In the case of a PR payment scheme a crowd worker usually works on one ore several of these micro-tasks and gets paid per finished task. Payment is usually lower than in ROTs, since each task is normally small and can be executed fast (Hammon and Hippner, 2012; Mason and Suri, 2012). This type of work can, for example, be data entry, image tagging, or verification of addresses. Following this we can classify crowdsourcing based on the incentive scheme as depicted in Figure 2.2.

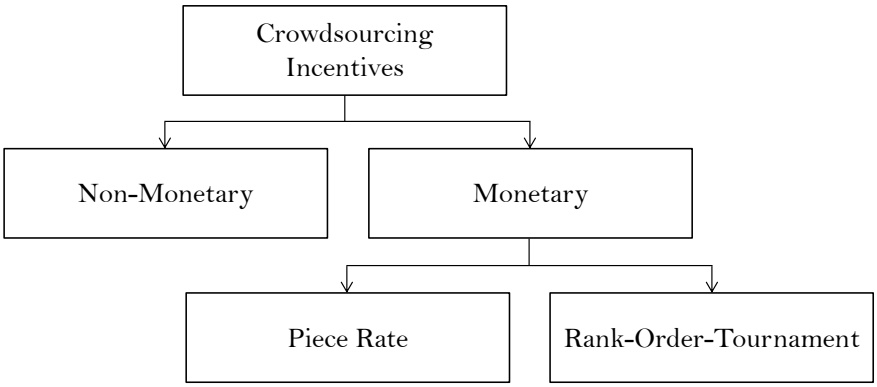


FIGURE 2.2: *Classification of crowdsourcing based on the incentive scheme. Own presentation based on Malone et al. (2010), Hammon and Hippner (2012), and Kittur et al. (2013).*

2.1.2 Exemplary Crowdsourcing Platforms

To better understand crowdsourcing and to get a feeling about its scope this section briefly discusses several crowdsourcing platforms. These platforms represent the three mainly used incentive types as depicted in Figure 2.2 (see Subsection 2.1.1) and are each prominent and successful examples of these incentive types. Moreover, the following platforms and their crowdsourcing application type are in focus of the Chapters 4, 5, and 6.

Amazon Mechanical Turk

MTurk[15] presents itself as "marketplace for work". It is one of the biggest, most noticed, and examined crowdsourcing platforms (Ipeirotis, 2010a; Paolacci et al., 2010). Mainly it is a micro-task platform, where requesters (mostly profit oriented companies) ask crowd workers to execute simple and mostly repetitive tasks which suit to fulfill an overarching goal – so called Human Intelligence Tasks (HITs) (Mason and Suri, 2012; Chandler et al., 2014). The crowd workers are paid per finished task (PR) and payment is rather low due to the simplicity of tasks (Paolacci et al., 2010). Hence, MTurk is an example of a platform using a monetary PR incentive type as depicted in Figure 2.2 (see Subsection 2.1.1). Based on these dimensions MTurk is comparable to platforms like clickworker[16] or CrowdFlower[21]. Such platforms are sometimes called "online labor markets" or "crowd work" platforms to differentiate them from non-paid crowdsourcing applications (see Horton, 2010; Kittur et al., 2013). The process of such crowdsourcing platforms is rather straightforward, as depicted in Figure 2.3. A requester publishes his task on the platform. The requester then

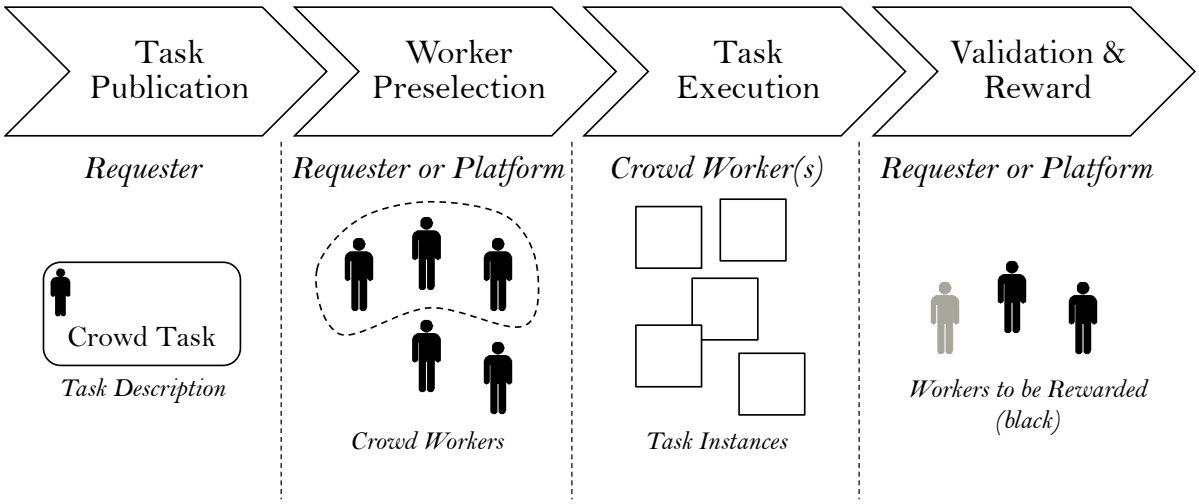


FIGURE 2.3: *The high-level steps of crowdsourcing and the respective actors of Tranquillini et al. (2015).*

may set requirements for workers, so called qualifications (Mason and Suri, 2012). Doing so, only workers who fulfill these requirements are allowed to work on such a task. On MTurk such crowd worker requirements include (i) a rate of how many of the previous tasks were processed correctly, (ii) how many tasks he already finished, and (iii) in which country a worker resides (Mason and Suri, 2012). Workers who fulfill these requirements then start working on the tasks. Once they are finished, the requester checks the results and if they are valid, pays the crowd workers, if not, rejects the task (Tranquillini et al., 2015). MTurk is the focus of the implemented analyses of Chapter 4.

99designs

99designs[17] is a crowdsourcing application where requesters (mostly profit oriented companies) can outsource the design of logos to crowd workers, i.e., a logo for a company or a website (Schlagwein and Farhad, 2014). The platform uses a tournament setup to decide who gets paid (Tranquillini et al., 2015). Hence, 99designs is an example of a platform using a monetary ROT incentive type as depicted in Figure 2.2 (see Subsection 2.1.1). A requester only pays for one design, the design picked in the end (Huang et al., 2012). Therefore he creates a crowdsourcing tournament and gives the crowd workers (designers) some specifications about the desired design (Cavallo and Jain, 2012). Several crowd workers then start designing logos and compete for the prize. During the tournament the requester can rank the designs and give personalized feedback to the crowd workers.

After several iterations, where the crowd workers can submit revised versions of their designs, the requester picks a winning submission (Cavallo and Jain, 2012; Huang et al., 2012; Tranquillini et al., 2015). Such contests are especially suitable for tasks where the requester does not have a clear idea about the outcome, as it is the case for creative work (Cavallo and Jain, 2012; Tranquillini et al., 2015). The crowd worker whose design is picked by the requester gets the prize, usually up to a couple hundred USD. The overall process is similar to the process depicted in Figure 2.3. However the key difference is, that the task is not divided into small micro-tasks, but is left as it is (Tranquillini et al., 2015). 99designs represents a platform with a ROT payment scheme for creative tasks. By this it is similar to crowdsourcing platforms like threadless[20] or InnoCentive[22]. Crowdsourcing platforms using a ROT payment scheme are the main focus of Chapter 5.

Wikipedia

Wikipedia[5] was created in 2001 with the goal to create the biggest encyclopedia of the world. The platform uses crowd workers to collaboratively write, proofread, and update the published articles (Rafaeli and Ariel, 2008; Leimeister, 2010). The platform does not use money to incentivize participation (Subramani and Peddibhotla, 2003; Rafaeli and Ariel, 2008). Hence, Wikipedia is an example of a platform using a non-monetary incentive type as depicted in Figure 2.2 (see Subsection 2.1.1). Access to the articles is free of charge as well. Overall more than 70,000 active crowd workers follow the idea of sharing their knowledge for free and created approximately 38,000,000 articles in 290 different languages[24]. Thereby, Wikipedia follows an altruistic goal with an epic meaning, which suits as incentive for participation (Subramani and Peddibhotla, 2003; Rafaeli and Ariel, 2008). Such non-monetary incentives are the focus of Chapter 6.

2.1.3 Related Concepts

Crowdsourcing is an umbrella term used for a multitude of different concepts (e.g., online labor, co-creation, crowdfunding, blogging, social networking, knowledge management, etc.). To clearly understand the scope and generalizability of this thesis we briefly discuss the differences from related concepts to crowdsourcing.

[24]Numbers are based on the project's homepage accessible at `https://en.wikipedia.org/wiki/Wikipedia` last accessed: September 6, 2016.

Open Innovation

Open innovation is a similar concept to crowdsourcing regarding the fact that some tasks once performed in-house are outsourced to a certain set of people. However, open innovation targets the innovation process of a company (Chesbrough, 2006; Leimeister, 2010). In many ways this concept focuses more on integrating customers or a special subset of them, such as lead users (cf. Straub et al., 2013; Kunze von Bischhoffshausen et al., 2015). Hence, it is often not an "open call" to the crowd, but rather a selective sample of (well) known customers.

Crowdfunding

Crowdfunding is the concept of using the crowd as seed capital investors for new startups and product ideas (Hammon and Hippner, 2012; Niemeyer et al., 2016). Compared to standard fundraising, where a small set of angel investors invest a large sum of money, the funds are spread within the crowd. Following this, many crowd workers only invest a small amount each, but the collective investments sum up to a large amount. While using the crowd for its purposes, here the financial assets, it does not focus on the work or labor of the crowd. Hence, incentives and motivation have a different nature. Therefore, crowdfunding is not in focus of this work.

Social Networks

Social networks are sometimes framed as crowdsourcing (Doan et al., 2011). However, social network are a tool to keep connected to your peers (Boecking et al., 2015). While still the crowd produces content for a subset of the crowd – a person's network – the focus of such platforms is to keep in touch with each other. Based on Doan et al. (2011) this is called exploiting the network. Keeping in touch with one's own network is not a crowdsourcing process. Based on the definition presented in Section 2.1 no "outsourcing" from processes within the company is applied (Howe, 2006a,b). Hence, social networks are not in focus of this work.

2.2 Existing Work on Collaborative Online Work and Incentives

Incentives for collaborative online work and in particular incentives for crowdsourcing applications have recently been subject to research (e.g., Horton and Chilton, 2010; Mason and Watts, 2010; Brabham, 2008, 2010; Chandler and Kapelner, 2013). As introduced in Figure 2.2, incentives for crowdsourcing can be categorized as non-monetary and monetary incentives. Monetary incentives can be further divided into PRs and ROTs (see Subsection 2.1.1 for more details).

To this end, most researchers concentrate on understanding monetary and non-monetary incentives in a crowdsourcing scenario independently of each other. For example, early works on PR payment schemes focus on the link between the amount of paid compensation and performance (e.g., Horton and Chilton, 2010; Mason and Watts, 2010) (see Chapters 4 and 5 for a detailed overview of related work on monetary incentives). Most studies on non-monetary incentives in crowdsourcing focus on incentives and motives for participation (see Chapter 6 for a detailed overview of related work on non-monetary incentives). Hence, implications on their effect on outcome variables such as performance (i.e., effort; see Chapter 4 for a detailed overview on related work on effort) and risk-taking (see Chapter 5 for a detailed overview on related work on risk-taking) can not be deduced as they are mainly questionnaire based (e.g., Brabham, 2008, 2010).

Comparisons between monetary incentives, in particular PRs and ROTs, has been analyzed in the context of classical labor (e.g., van Dijk et al., 2001; Eriksson et al., 2009a, 2009b) (see Chapter 4 for a detailed overview of related work on the different effects of PRs and ROTs). However, classical labor is different from collaborative online work and crowdsourcing (see Section 2.1 for more details). Motives for participation and hence the effect of incentives might differ vastly in a crowdsourcing scenario (see Chapter 3).

Following this, the work at hand thereby addresses and unifies several research streams and links them to several outcome variables, as depicted in Table 2.1. All of them are important to fully understand the research domain of collaborative online work and crowdsourcing. A detailed overview of related work will be presented in the respective Chapters 4, 5, and 6.

Literature	Incentive Type			Outcome		Context	
	Non-Monetary	Piece Rate	Rank-Order-Tournament	Performance	Risk-Taking	Crowdsourcing	Classical Labor
Horton and Chilton (2010)	○	●	○	●	○	●	○
Mason and Watts (2010)	○	●	○	●	○	●	○
Ehrenberg and Bognanno (1990)	○	◑	●	●	○	○	●
Bracha and Fershtman (2013)	○	◑	●	●	◑	○	●
van Dijk et al. (2001)	○	●	●	●	○	○	●
Eriksson et al., (2009a)	○	●	●	●	◑	○	●
Eriksson et al., (2009b)	○	●	●	●	○	○	●
Taylor (2003)	○	○	●	○	●	○	●
Bothner et al. (2007)	○	○	●	○	●	○	●
Nieken and Sliwka (2010)	○	○	●	○	●	○	●
Qiu (2003)	○	○	●	○	●	○	●
Grund and Gürtler (2005)	○	○	●	○	●	○	●
Brabham (2008)	●	○	○	○	○	●	○
Brabham (2010)	●	○	◑	○	○	●	○
Kaufmann et al. (2011)	◑	◑	○	○	○	●	○
Rogstadius et al. (2011)	◑	●	○	◑	○	●	○
Chandler and Kapelner (2013)	◑	●	○	●	○	●	○
This Work	●	●	●	●	●	●	◑

TABLE 2.1: *Overview of relevant domains and main related literature of the work at hand.*

Chapter 3

The Theory of Motivation and Incentives

> “There is no medicine like hope, no incentive so great, and no tonic so powerful as expectation of something better tomorrow.”
>
> ORISON SWETT MARDEN, 1908

DESIGNERS of online collaboration platforms and in particular crowdsourcing platforms face several challenges when they create platforms, because they need to generate specific outcomes. Two of the main challenges of crowdsourcing platforms are to secure qualitative online work and to incentivize participation (Shaw et al., 2011; Wang et al., 2013). This work focuses on the latter – incentives for crowdsourcing applications.

It is key to understand the related concepts around incentives and the interaction between these concepts for designing platforms where the crowd – an undefined set of people in the web (Howe, 2006a,b) – needs to participate and behave in a specific manner to achieve a desired outcome. People work, share their knowledge, or participate when they are motivated to do so. Incentives have the potential to spur motivation, as they depict a certain goal (e.g., winning a prize, earning money, etc.) (Rheinberg, 2008; Schattke, 2011; Ryan and Deci, 2000). If the person desires this goal, then his willingness to take action is addressed. Hence, this person is motivated by this incentive (Ryan and Deci, 2000). We

therefore have to understand the link between incentives and motivation in more detail to design the best incentives for crowdsourcing platforms. Crowdsourcing applications, for example, use different monetary and non monetary incentive schemes (see Chapter 2) such as PRs (e.g., MTurk[1], clickworker[2]), ROTs (e.g., 99designs[3], threadless[4]), and no monetary incentives (e.g., Wikipedia[5], Google ReCAPTCHA[6]) as discussed in Subsection 2.1.1. Clearly these incentives effect a crowd workers' motivation differently. This chapter outlines the motivation theories behind these incentives to position this work's studies in the motivation theory literature (see Chapters 4, 5, and 6).

The remainder of this chapter is structured as follows. Section 3.1 defines the terms motive, incentives, motivation, and behavior to clearly distinct between the four concepts. Section 3.2 then summarizes motivation theory, by introducing content theories, process theories, and intrinsic and extrinsic motivation.

3.1 Definitions

To better understand and differentiate between the related concepts motive, incentives, motivation, and behavior, they will be briefly defined and delineated based on the classical approach to motivation from Rheinberg (2008) and Schattke (2011) as depicted in Figure 3.1:

Motive

A motive is a personal attribute (Rheinberg, 2008, 2011). Motives depict a person's overarching objectives and hopes. A motive then is the desire to achieve this outcome. However, motives vary across different persons, as they are individual attributes (Rheinberg, 2008; Schultheiss et al., 2009; Schattke, 2011). Hence, motives are, in contrast to motivation, time-stable and character-dependent variables (Rheinberg, 2011; Mazarakis, 2013).

Incentive

Incentives and motives are linked (Rheinberg, 2008). Incentives depict the achievement of

[1] `https://www.mturk.com/mturk/welcome` last accessed: September 6, 2016.
[2] `https://www.clickworker.com/` last accessed: September 6, 2016.
[3] `http://en.99designs.de/` last accessed: September 6, 2016.
[4] `https://www.threadless.com/` last accessed: September 6, 2016.
[5] `https://www.wikipedia.org/` last accessed: September 6, 2016.
[6] `https://www.google.com/recaptcha/intro/index.html` last accessed: September 6, 2016.

certain goals. If this goal is in line with a person's motives, the interaction of the incentive and his motives increases the chances that the person will take action (Schattke, 2011). However, persons differ in how attractive they find those incentives, as they might have different motives. Even the reaction within a person to an incentive might differ based on the current situation (Rheinberg, 2008; Schattke, 2011). For example, a person who currently needs money to pay his bills finds working for a moderately loan attractive, while several years later the same person might not find it attractive, since he currently does not urgently need money. In the latter situation the person might base his decision to work on different factors, such as if the work is enjoyable or interesting.

Motivation

Motivation results out of the interaction of motives and incentives (Rheinberg, 2008, 2011; Schattke, 2011; Mazarakis, 2013). Individuals are motivated, when they are driven to do something (Ryan and Deci, 2000). Motivation therefore is the result of different factors (i.e., motives and incentives) which stimulate a person to a certain action (see Section 3.2). Hence, motivation results out of any motive(s) which drive an individual towards a certain action and those incentive(s), e.g., money and the person's current situation, which trigger these motive(s). The interaction of motives and incentives lead to motivation, which increases the willingness to act (Rheinberg, 2008; Schattke, 2011). Hence, motivation is, in contrast to motives, a time- and character-dependent variable (Rheinberg, 2011; Mazarakis, 2013).

Behavior

A person who is motivated for a task, may change his behavior, e.g., work harder. However, it is important to understand that motivation does not necessarily lead to a certain behavior or participation. Motivation only increases the likelihood to act (Schattke, 2011). Hence, a platform designer can profit from having insights on how different *incentives* address different *motives* and, hence, *motivation*, which will result in higher chances to behave in a certain way (Rheinberg, 2008; Schultheiss et al., 2009; Schattke, 2011).

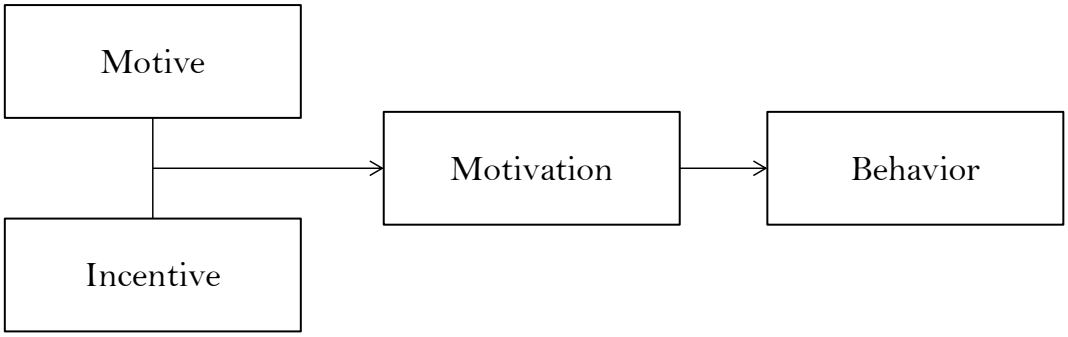


FIGURE 3.1: *The classical approach to motivation adapted from Rheinberg (2008).*

3.2 Motivation Theory

Understanding motivation is crucial for designing the right incentives for a specific platform. However, a clear definition of motivation is still under scientific debate, as many different and refined theories exist (cf. Maslow, 1943; Deci and Ryan, 1985; Csikszentmihalyi, 1990; Amabile et al., 1994; Ryan and Deci, 2000; Rheinberg, 2008; Vallerand et al., 2008; Vansteenkiste et al., 2010; Schattke, 2011). These theories can be divided into content- and process theories (Kanfer, 1990). Content theories describe a person's motives and which variables influence a person's motivation. Hence, content theory focuses on the question "*what* gets a person motivated" (see Maslow, 1943; Herzberg, 1968; Hackman and Oldham, 1976; Herzberg et al., 2011). Process theories, on the other hand, focus on how motivation originates. Hence, process theory focuses on the question "*how* to get or keep a person motivated" (see Vroom, 1964; Lawler, 1971; Locke, 2001; Locke and Latham, 2002; Rheinberg, 2011). Both theories will be briefly introduced in the following Subsections (3.2.1, 3.2.2) to highlight the motivation theory streams which are currently under scientific debate. Afterwards a third research stream, which differentiates motivation into intrinsic and extrinsic motivation, will be introduced in more detail as this categorization serves as main theoretical framework of motivation in the context of this work (Subsection 3.2.3).

3.2.1 Content Theories

One of the first content theories is Maslow's hierarchy of needs (Maslow, 1943). Maslow (1943) divides a person's needs in a hierarchical system where more basic needs are at

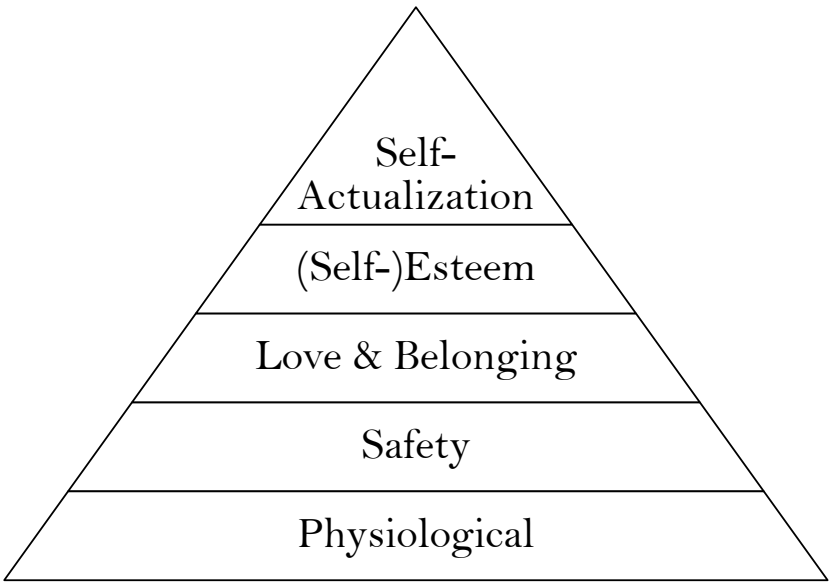


FIGURE 3.2: *Maslow's hierarchy of needs. Own presentation based on Maslow (1943).*

the bottom: (i) physiological needs, (ii) safety needs, (iii) love and belonging needs, (iv) (self-) esteem needs, and (v) self-actualization needs, as depicted in Figure 3.2. This hierarchical view clarifies that a person's needs are not independent from each other. A person only seeks to fulfill higher needs once basic needs are satisfied (Maslow, 1943) (see Figure 3.2).

The two-factor-theory from Herzberg (1968) follows a similar idea. Herzberg (1968) divides motivation into two distinct factors: (i) hygiene factors and (ii) motivators. Hygiene factors are neutral to a person's motivation for a certain work task. However, if they are missing a person gets dissatisfied with his job. In a work environment, such factors can be salary or the relationship with peers (Herzberg, 1968). Motivators, on the other hand, get a person motivated for doing a task. However, they do not necessarily lead to dissatisfaction with the job if they are missing. Such motivators can be the work itself, growth, and recognition (Herzberg, 1968).

Hackman and Oldham (1976) describe in their job-characteristics-model five variables as main reasons for an internal motivation towards a task or job: skill variety, task identity, task significance, autonomy, and feedback. An employee becomes intrinsically motivated (see Subsection 3.2.3 for further details on intrinsic motivation), if a task fulfills these psychological states (Hackman and Oldham, 1976).

Content theories, in summary, help to identify motives behind motivated actions. The process of getting a person motivated to perform a certain task is focus of process theory, which will be focus of the next Subsection (3.2.2).

3.2.2 Process Theories

The expectancy theory by Vroom (1964) relies on the three factors: (i) valence, (ii) instrumentality, and (iii) expectancy. Motivation is defined as a process of a person, which decides between several alternatives of tasks. Each task eventually leads to a certain outcome. A person decides whether to do a task based on valence, instrumentality, and expectancy. Valence describes if a person desires an outcome (positive valence) or wants to avoid the outcome (negative valence). Instrumentality describes if a person's actions lead to a certain (desired) outcome, if the goals of the task are met. Expectancy describes a person's subjective expectation that one's effort will lead to a performance good enough to meet the goals of the task (Vroom, 1964). Motivation is then the product of the person's valence towards the task outcome (e.g., positive valence of potential promotion), the instrumentality if the task will lead to the outcome (e.g., acquisition of a big industry project with a potential project manager position), and how much effort the person expects to fulfill the task (e.g., several project acquisition meetings) (Vroom, 1964).

Locke and Latham (2002) describe in their goal-setting theory motivation as pursuit to achieve goals. Goals serve as a directive function. Hence, if a person wants to achieve a goal the person focuses his effort and attention towards goal relevant activities. Goal irrelevant tasks are seen as a distraction (Locke and Latham, 2002). Furthermore, the harder it is to achieve a desired goal, the more effort a person puts into the tasks to fulfill the goal. Hence, a person is more motivated to do the task (Locke and Latham, 2002). In addition, Locke and Bryan (1969) found that giving feedback about a person's current performance directs more effort into tasks in line with their goals.

Process theories, in summary, help to understand the process of how people get motivated. However, there exists another prominent research stream besides these process- and content theories, which differentiates between intrinsic and extrinsic motivation. In the following Subsection (3.2.3) this distinction will be further explained.

3.2.3 Intrinsic and Extrinsic Motivation

One of the most prominent theories and categorizations of motivation differentiates between the drivers that lead to a certain reaction (Deci and Ryan, 1985; Ryan and Deci, 2000). The so called self-determination theory categorizes two different types of motives as main drivers for motivation – intrinsic and extrinsic motivation. This categorization is used in many other publications (cf. Hars and Ou, 2001; Tedjamulia et al., 2005; Reiss, 2012; Mazarakis, 2013; Kosonen et al., 2014; Xu and Li, 2015) and will be used in the context of the work at hand as well, as main theoretical framework for motivation.

Intrinsic Motivation

When a person is intrinsically motivated the person takes action because the task is enjoyable or the person is interested in the task. Intrinsic motivation mostly lacks of an observable and separable outcome. Hence, purely intrinsic motivated tasks are executed by a person for the task itself and not some separable incentive or outcome (Deci and Ryan, 1985; Ryan and Deci, 2000). For example, a person who starts to solve a logic puzzle, such as the rubik's cube, out of interest, curiosity, and fun. Intrinsic motivation is seen as the more beneficial motivational factor than extrinsic motivation (Ryan and Deci, 2000). However, intrinsic motivation is highly dependent on the individual (see Section 3.1). One person might be intrinsically motivated for a task, while, regarding the same task, this is not true for a different person (Ryan and Deci, 2000). Moreover, social demands and responsibilities in work-life for non-intrinsic tasks undermine intrinsic motivation after childhood (Ryan and Deci, 2000). Harter (1981) observed that intrinsic motivation of school children is decreasing with every school year. Furthermore, tasks can lack the potential of motivating individuals intrinsically, such as boring and repetitive tasks (Ryan and Deci, 2000; Gill and Prowse, 2012; Mazarakis, 2013). Intrinsic motivation is a main part of the study conducted in Chapter 3.

Extrinsic Motivation

Extrinsic motivation refers to taking an action because of a separable outcome, for example a monetary reward (Deci and Ryan, 1985; Ryan and Deci, 2000). Purely extrinsic motivated tasks are executed by a person solely for this separable outcome and not for the task itself (Deci and Ryan, 1985; Ryan and Deci, 2000). Hence, extrinsic motivation contrasts intrinsic motivation. For example, working for a salary or filling out a questionnaire

for the chance to win a prize. Extrinsic motivation can vary in its autonomy (Ryan and Deci, 2000; Mazarakis, 2013). Persons can execute tasks with disinterest and resistance or with willingness, because they have accepted the task. In both cases the task can still be extrinsically motivated, since the person is performing the task for a separable outcome (Ryan and Deci, 2000). For example, a person can work on a project only because he fears sanctions from his boss, while a different person might perform the same work to increase chances of a promotion. Both persons would still be extrinsically motivated, while the former case is an example for external control and the latter involves a feeling of choice (Ryan and Deci, 2000). Extrinsic motivation is a main part of the studies conducted in Chapters 4 and 5.

Most of the times people are motivated due to several reasons. For example, a person might work since the work is enjoyable to some extent (intrinsic motivation) and since earning money to pay bills is a necessity (extrinsic motivation). Hence, a person's motivated state results out of the interplay of intrinsic and extrinsic motivation. Moreover, motivation varies along different dimensions. First, it can vary in strength, with the antipodes being not or highly motivated. Second, motivation can vary in the underlying reasons for the motivation – intrinsic, extrinsic, or the interaction of both. Following this, a person, for example, can be highly motivated to work on 99designs[3], since the person enjoys creating logos and wants to earn money (intrinsic and extrinsic motivation), while being moderately motivated to work on MTurk[1] for a monetary reward (extrinsic motivation) even if the person finds the tasks not enjoyable or interesting (no intrinsic motivation). However, it is important to notice that the interaction of intrinsic and extrinsic motivation is not linear, and is strongly context dependent (Frey and Jegen, 2001; Bénabou and Tirole, 2003; Frey, 2012). For example, extrinsic incentives (e.g., monetary rewards) can undermine intrinsic motivations – the so called crowding-out effect. It has been observed that adding an extrinsic incentive does not necessarily increase or in some cases even reduces effort and participation when participants were highly intrinsically motivated before the extrinsic factor was added (Frey and Jegen, 2001; Bénabou and Tirole, 2003; Frey, 2012). This non-linear relation means platform designers, by just adding incentives, do not necessarily improve the outcome. This and the fact that the reaction to incentives is context- (i.e., the crowdsourcing platform) and person- (i.e., the crowd) specific, makes clear that insights from work-life or related concepts, such as social networks, open innovation, and crowdfunding (see Subsection 2.1.3 of Chapter 2), can not be transferred one to one to

crowdsourcing. This stresses the importance of evaluating different incentives (see Subsection 2.1.1 of Chapter 2) in the field of crowdsourcing and its different platforms (see Subsection 2.1.2 of Chapter 2) as implemented in Chapters 4, 5, and 6.

Part III

Insights from Collaborative Online Work

Chapter 4

Incentives for Crowdsourcing: Payment Schemes and Feedback

> “ Call it what you will, incentives are what get people to work harder.”
>
> NIKITA KHRUSHCHEV

THIS chapter evaluates the effect of different payment schemes (PR and ROT, see Subsection 2.1.1 of Chapter 2) and feedback on one of the outcomes of crowdsourcing platforms – effort of crowd workers. Moreover, it introduces a research model of the coherences of feedback and crowd workers' effort in crowdsourcing and evaluates the effects. Therefore, this chapter focuses on Research Questions 1 and 2 (see Section 1.2 of Chapter 1).

Research Question 1: *Which incentive scheme – piece rates or rank-order-tournaments – incentivizes crowd workers' effort best?*

Research Question 2: *Does feedback in rank-order-tournaments affect crowd workers' effort in crowdsourcing settings?*

This chapter builds on and extends the collective works: Straub, Gimpel, and Teschner (2014), Straub, Gimpel, Teschner, and Weinhardt (2014), and Straub, Gimpel, Teschner, and Weinhardt (2015). It starts with an introduction and motivates the context and importance of this evaluation (Section 4.1). Section 4.2 summarizes specific related literature for this chapter. Section 4.3 develops and outlines the research model. Section 4.4 explains the overall study design. Sections 4.5 through 4.7 (studies 1–3) explain the specific study design (4.5.1, 4.6.1, 4.7.1) and show the results (4.5.2, 4.6.2, 4.7.2) of the respective study. Section 4.8 summarizes the findings of all three studies and shows pathways for future directions.

4.1 Introduction

Recently, crowdsourcing has gained attention by both practitioners and researchers as model to outsource human work on demand to a broad, diverse, and distributed workforce. Monetarily incentivized (paid-) crowds are provided by many commercial vendors, e.g., MTurk[1], upwork[2], clickworker[3], 99designs[4], and InnoCentive[5]. These platforms provide access to a range of different crowd workers who work on a wide range of tasks – from simple repetitive e-mail tagging to creative and more complex tasks such as building logos (Kittur et al., 2012, 2013; Hammon and Hippner, 2012).

In these online labor markets one challenge for organizations is to properly incentivize crowd workers' effort and quality of work (Huang et al., 2010; Shaw et al., 2011; Wang et al., 2013). Work relations are short-lived and commonly one-shot labor relations, i.e., the relation between the requester and worker only exists once for the given task and is finished after its completion. This is in contrast to standard employment contracts in offline labor situation, where contracts between the employer and employee usually last over at least several month. Quality control is mostly done by repetition of work by different crowd workers (Ipeirotis et al., 2010; Kokkodis and Ipeirotis, 2013; Wang et al., 2013). In paid crowdsourcing settings, crowd workers are usually incentivized by a PR (pay per

[1] `https://www.mturk.com/mturk/welcome` last accessed: September 6, 2016.
[2] `https://www.upwork.com/` last accessed: September 6, 2016.
[3] `https://www.clickworker.com/` last accessed: September 6, 2016.
[4] `http://en.99designs.de/` last accessed: September 6, 2016.
[5] `https://www.innocentive.com/` last accessed: September 6, 2016.

finished task) or by a ROT (see Subsection 2.1.1 of Chapter 2). PR payments are most commonly observed for the collection of crowd input on activities that can be divided into small pieces that can be done (mostly) independently of each other (Malone et al., 2010). Paying a price to the best performing crowd worker or a small set of top performers is most commonly seen for tournaments when only one or a few good solutions are needed. Examples include the design of a good algorithm or logo. On crowdsourcing platforms hosting such tournaments, like 99designs, the employer (i.e., the person or organization that creates a crowdsourcing task and posts it on the crowd labor market) typically has to decide whether to provide feedback on a crowd worker's current competitive position. Standardly these platforms give the option to display leaderboards and signal who is the provisional winner or to hide this information from crowd workers (Araujo, 2013). In addition the requester can directly contact the workers by commenting their submissions (Araujo, 2013). ROTs are also commonly used in traditional work places (Microsoft, GE, Yahoo! etc.) and sports (poker, soccer leagues etc.). Given their wide usage and appeal of using competitive elements to incentivize workers, some organizations using crowdsourcing even employ ROTs on platforms like MTurk[1], where PR payments are seen as the standard. Setting up and controlling a ROT is clearly more cumbersome than straightforward PR payments. Handling this complexity might, however, pay off when the crowd workers' performance is higher; given that both incentive schemes provide the same average wage for crowd workers.

Overall, this raises two main questions: (i) Do rank-order-tournaments lead to better performance by crowd workers than piece rate payments? (ii) When conducting a crowd labor tournament, should one provide feedback on a crowd worker's competitive position? Those questions are implemented as Research Question 1 and 2 presented in Section 1.2 of Chapter 1:

> **Research Question 1:** *Which incentive scheme – piece rates or rank-order-tournaments – incentivizes crowd workers' effort best?*
>
> **Research Question 2:** *Does feedback in rank-order-tournaments affect crowd workers' effort in crowdsourcing settings?*

In this chapter both these research questions are investigated in an exploratory way using a series of three real effort studies on MTurk[1] with overall 874 crowd workers participating.

4.2 Background

4.2.1 Payments in Online Labor

Crowdsourcing and online labor markets usually need smaller payments than traditional work relations. Paolacci et al. (2010) report that to get comparable results to traditional offline labor settings, crowdsourcing needs rather small monetary incentives. The hourly reservation wage of crowd workers on MTurk is USD 1.80, while the effective hourly wage is USD 4.80 (Horton and Chilton, 2010; Ipeirotis, 2010a).

Several studies analyzed how changes in the payment structure influence the behavior of crowd workers in online labor markets (Mason and Watts, 2010; Shaw et al., 2011). Shaw et al. (2011) show that linking monetary incentives to the responses of other crowd workers (e.g., penalty for disagreeing to the majority) influences the performance. In their experiment they decreased the payment (i.e., 10 %) for a task if the crowd worker's response disagreed with the majority of other crowd workers' responses. In contrast to just paying a PR, crowd workers' show a higher performance under this penalty condition.

Mason and Watts (2010) show that more money leads to more effort by crowd workers. The experimental setup varied the payment and difficulty of the task. Participants who were exposed to a setup where more money was paid, finished more tasks than participants in lower payment treatments. However, quality is not affected by an increased payment. Participants made the same amount of mistakes in the different difficulty setups, independent of the payment. Moreover, compared to a PR, an overall lower quota pay scheme, which only pays for a set of completed tasks, leads to a greater output.

To sum up, it is an open debate which incentive and information structures are best suited to stimulate crowd workers' performance.

4.2.2 Incentives: Piece Rates and Rank-Order-Tournaments

In ROTs two or more people compete against each other and are ranked according to their performance. Top performer(s) win the tournament. Such settings are used in sports (poker, soccer, etc.), traditional labor relations and work places (Microsoft, GE, Yahoo!, etc.), and crowdsourcing platforms (99designs[4], threadless[6], InnoCentive[5], by some requesters on MTurk[1], etc.).

In settings, other than crowdsourcing, economics suggest that ROTs incentivize workers better than piece rates (PRs) (Bracha and Fershtman, 2013; Lazear and Sherwin, 1981; Ehrenberg and Bognanno, 1990; Bull et al., 1987; van Dijk et al., 2001). Ehrenberg and Bognanno (1990) find that professional golf players are positively incentivized by tournaments. Lazear and Sherwin (1981) give evidence that ROTs, used in work places, incentivize risk-averse workers equally well as a PR. Bracha and Fershtman (2013) distinct effort into labor effort and cognitive effort. Participants who work under a ROT exert more labor effort but at the same time less cognitive effort than when working under a PR. Reasons are that competition incentivizes workers but they are less able to do cognitive tasks under pressure. In addition, van Dijk et al. (2001) find that effort levels and variance in ROTs are higher compared to PRs. In addition, low ability workers work harder. Similarly, Bull et al. (1987) find a higher variance of effort in ROTs compared to PRs.

These results suggest the strength of the competitors as explanation for effort variance: Some participants might lose interest when falling behind (Mazarakis and Krämer, 2010; Mazarakis and Van Dinther, 2011; Mazarakis et al., 2011; Mazarakis, 2013). Others who are in front might relax. Some who are close to each other might actually be competing harder. Eriksson et al., (2009a) present experimental evidence that if participants can choose between ROTs and PRs, variance decreases and effort levels increase in ROTs. They further find that risk-averse participants tend to choose a PR.

Eriksson et al., (2009b) experimentally study the influence on participants' effort by giving feedback on their current position with PR payments and ROTs. Three different feedback rules on relative performance are observed – (i) no feedback, (ii) feedback given half way through the experiment, and a (iii) continuously updated feedback. On average feedback does not change effort, but participants who are behind make more mistakes

[6] `https://www.threadless.com/` last accessed: September 6, 2016.

under continuous feedback and almost never drop out of the ROT. The reason could be a social norm to never give up (Eriksson et al., 2009b). Arguably this relation might, however, be stronger in a laboratory setting, which is used in these studies, than an anonymous digital crowdsourcing settings. Hence, drop outs might be more pronounced in online labor markets.

Evidence for this is presented by Fershtman and Gneezy (2011): Quitting is often avoided by participants because it is socially stigmatized. Still, higher rewards lead participants to exert more effort and quit more often at the same time. Finally, Pull et al. (2013) show that in dyadic tournaments where the ability of participants is heterogeneous, effort levels should decrease, because both know that one will win anyway. When participants' abilities are homogeneous, effort levels should be high. In consequence it is expected that continuous feedback will lead to the same effect. In detail, if participants get feedback and performed better than expected, they decrease their effort but expect to be better in the future (Kuhnen and Tymula, 2012). While workers who performed worse than their expectations will increase their effort but reduce their expectations. This implies that showing feedback has the potential to improve and lower performance of participants depending on the current position in the tournament.

4.2.3 Motivation in Paid Micro-Task Crowdsourcing

Intrinsic motivation refers to doing something because one wants to do it out of pure interest or fun, while extrinsic motivation refers to the motivation of doing something out of external reasons like getting a reward (Ryan and Deci, 2000; Eccles and Wigfield, 2002) (see Subsection 3.2.3 of Chapter 3). Both affect the performance of crowd workers in crowdsourcing settings (Brabham, 2008, 2010; Kaufmann et al., 2011; Pilz and Gewald, 2013). Paid crowdsourcing settings (e.g., MTurk, 99designs[4], threadless[6], InnoCentive[5], clickworker[3], CrowdFlower[7], etc.) mainly targets extrinsic factors, by paying crowd workers – an extrinsic motivation. Thus, this study focuses on extrinsic motivation by using monetary compensation as incentive for performing well. It is not possible to completely exclude intrinsic motivation. However, on MTurk the main motivation of crowd workers

[7] `http://www.crowdflower.com/` last accessed: September 6, 2016.

is to earn money (Kaufmann et al., 2011). Hence, crowd workers are initially mainly externally motivated. Furthermore, by using a real effort task, which from design has no overall meaning for society or the person itself, is boring, and does not improve major skills of a person, it is as much as possible excluded that altruism, entertainment, and personal development, which are often referred to as intrinsic motivation, play a role (Gill and Prowse, 2012). Additionally, Deci et al. (1981) argue that competition decreases intrinsic motivation. More accurately, competitive situations where participants feel a pressure to win undermine intrinsic motivation (Reeve and Deci, 1996).

4.3 Research Model

The first aim is comparing performance of crowd workers, i.e., crowd workers' effort, in rank-order-tournaments (ROTs) and piece rate payments (PR). Following van Dijk et al. (2001), Bracha and Fershtman (2013), and Ehrenberg and Bognanno (1990), it is hypothesized that when both mechanisms yield the same expected payout ROTs should be associated with higher performance. PR payments offer little room for designing the incentive scheme; the key parameter is the PR itself which is set to be equal to the average ROT payout. A ROT, on the contrary, opens up more design options. Hence, the second aim is to evaluate these and aim for the best ROT design. Therefore the related work reviewed above is summarized in the research model depicted in Figure 4.1.

Following the sequential distinction of service quality in (i) structure, (ii) process, and (iii) outcome (Donabedian, 1980, 2003), a crowd worker's performance is considered as outcome and is hypothesized to be related to the work process and structures. Structural constructs are classified as (i) individual, (ii) crowd, or (iii) system level (Figure 4.1). Potentially, this structure will prove useful for more extensive conceptualization on the interrelation of crowdsourcing incentives and quality. Evaluating this belief is future work; here the generic structure is used as frame for a specific moderated mediation model.

The model shows the hypothesized correlates of crowd workers' performance in ROTs, which are the following: Performance is the achievement of a worker on a given task in a given time frame, in this case a crowd worker's effort. It is operationalized as number of successfully completed instances of a task in a certain time frame. Performance is directly

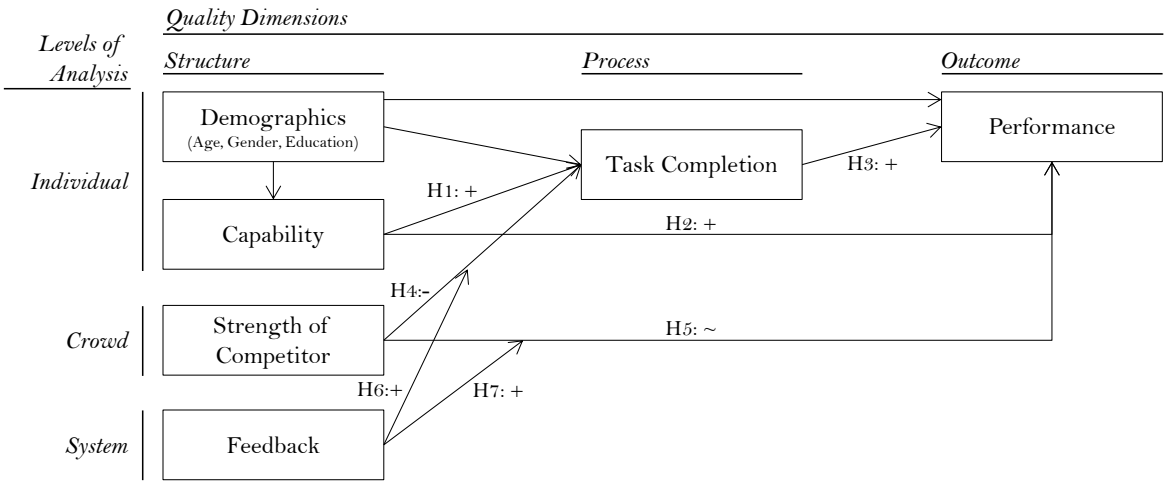


FIGURE 4.1: *Hypothesized research model on the correlates of crowd workers' performance in rank-order-tournaments.*

related to the crowd worker's capability, i.e., his ability to perform the specific task. Hence, capability is measured as the number of finished tasks in a pre-round. Capable crowd workers are expected to perform better. Strength of competitors is the performance of the respective competitors. Feedback indicates whether participants are informed about their current position in the tournament. In this case whether a leaderboard is shown. Based on the work by Eriksson et al., (2009b), performance might be related to the competitors' strength in cases when feedback on the performance and current standing in a ROT is provided. Therefore the correlation might be moderated by feedback. Given evidence from studies on ROTs, the direction of the moderated effect of the competitors' strength on performance is, however, not ex-ante clear (Eriksson et al., 2009b; Fershtman and Gneezy, 2011; Pull et al., 2013). Whether a crowd worker finishes the task or not is indicated by task completion. Following Fershtman and Gneezy (2011) and Eriksson et al., (2009b) it is assumed that a strong association with performance exists. Task completion is hypothesized to mediate the correlation of capability and strength of competitors on performance. Crowd workers able to do a task will finish it more often. Therefore a positive correlation between capability and task completion is assumed. Strength of competitors is assumed to be correlated with task completion: Similar to Fershtman and Gneezy (2011), it is believed that falling behind leads to quitting, hence, the stronger the competitor, the more likely crowd workers quit the task. Feedback moderates the association of strength of competitors with both task completion and performance. Only when feedback is given, the

competitors' strength can be seen and hence show a relation. Facing strong competitors is expected to lead to a stronger relation between competitors' strength and task completion than facing weaker competitors (Pull et al., 2013). For strong competitors, the hypotheses is that the relation to performance is positive while it is expected to be negative for weak competitors. In other words: When a crowd worker sees that he is falling behind but does not quit the task, the feedback is expected to increase performance. When the crowd worker is ahead, he might relax and therefore performance decreases. When the crowd worker is facing an equally good competitor and always has to excel to win, a performance increase and almost no dropout rates are expected, since there is always a fair chance to win. Finally, a crowd worker's age, gender, and education is expected to be correlated with capability, task completion, and performance – at least these demographics might serve as surrogate measurement for less observable individual characteristics. Directions of this correlation are not hypothesized, since this is not in focus of the work at hand.

4.4 General Study Design and Procedures

In this chapter the relations between performance (i.e., crowd workers' effort), strength of competitors, and feedback as implemented in the research model (Figure 4.1) are explored. Hence, results of three studies are presented: Study 1 compares PR payments with the simplest dyadic ROT providing no performance feedback. Study 2 investigates the performance in dyadic ROTs depending on the strength of the competitor and whether feedback is provided or not. Study 3 further tweaks the design of the dyadic ROT by featuring a group matching where individual crowd workers are matched with supposedly equally well performing competitors to spur their performance.

All three studies in this chapter have similar designs and can be referred to as "exploratory studies using experimental techniques" (see Section 7.1 of Chapter 7 for further details). They implement a real effort task to measure performance (i.e., crowd workers' effort), similar to the slider task by Gill and Prowse (2012). A real effort task is designed to measure effort while excluding accompanying motivational factors (Gill and Prowse, 2011, 2012). In the case of the studies of this chapter, crowd workers have a fixed time adjusting as many sliders as possible ranging from 0 to 100 to a value of 50. Correct positioned sliders reset, with a changed shape (width and horizontal orientation), until

either the time for the task elapses or the crowd worker quits. The number of sliders a crowd worker correctly sets prior to the end of the task is the measure of performance (i.e., effort). The rather simple, needless work is on purpose and typical for real effort experiments. The intention is to measure crowd workers' reaction with a task that depends as little as possible on pre-existing knowledge, learning by doing effects, randomness, or guessing (Gill and Prowse, 2012). In addition it partially excludes intrinsic motivational factors like entertainment, learning, or contribution to an epic meaning (see Chapter 3).

All tournaments are dyadic ROTs – a crowd worker competes with only one other crowd worker, the winner gets a bonus of USD 1.00; the loser does not receive a bonus. The choice of the smallest possible number of competitors aims at making the competitor salient and allowing crowd workers to most clearly judge their competitive position. In this, this chapter follows the study design by Eriksson et al., (2009a), Eriksson et al., (2009b), Fershtman and Gneezy (2011), and van Dijk et al. (2001) and posits that this design feature (i.e., a dyadic ROT design) carves out the relation between tournament competition and performance most clearly. To increase experimental control, participants do not compete live but against historic data collected from a previous participant. This is made clear in the instructions of the crowd experiment. Using historical data is a design choice to compensate for unknown arrival times of crowd workers which are given due to the open call nature of crowdsourcing and MTurk, where it is not given that two or more people arrive at the same time to participate in an experiment (Suri and Watts, 2011; Mason and Suri, 2012; Mao et al., 2012; Amir et al., 2012) (see Section 7.2 of Chapter 7 for more details on this design challenge).

All participants are recruited from the general pool of MTurk crowd workers restricted that they can only take part once and in one of the three studies; reside in the US; finished at least 1,000 MTurk tasks (so called HITs) prior to our studies; and 95 % of their prior work was approved by the respective employer (requester) (see Subsection 2.1.2 of Chapter 2 to see more details on crowd worker requirements and the rejection of crowd work). Using MTurk as platform for experimental research is gaining prominence in various disciplines, including economics (Horton et al., 2011), psychology (Buhrmeister et al., 2011), computer science (Chilton et al., 2010), and information systems (Teschner and Gimpel, 2013). For the purpose of this study, MTurk is not merely a platform to recruit and reimburse participants but the natural environment of many crowd workers. In fact it is the

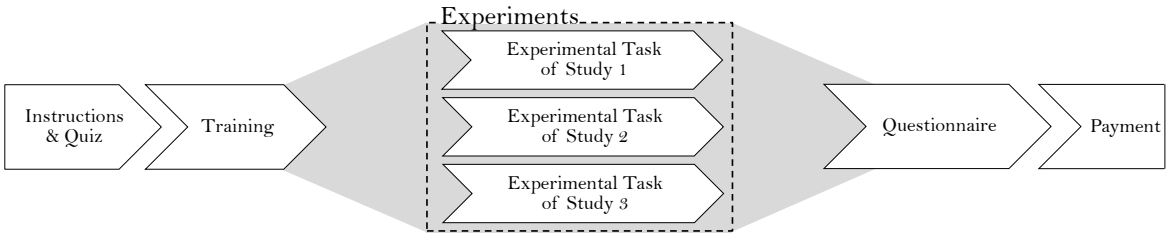


FIGURE 4.2: *Overall experimental process of Chapter's 4 Studies 1, 2, and 3.*

crowdsourcing market with most crowd workers and most tasks. All three studies start out with the instructions and a short quiz to test understanding of these, followed by a training round, the experimental task, a questionnaire on some demographics, and payment of participants according to their respective performance. This overall experimental process is depicted in Figure 4.2.

The studies are conducted with a custom-made web application, using the Grails[8] framework and a model view controller approach. From a technical perspective I follow the guidelines of Mao et al. (2012) and Mason and Suri (2012). The slider task was originally developed in z-Tree (Fischbacher, 2007). A similar version is implemented with Java and integrated in the custom-made web application to be accessible online through MTurk. An 'out' button was added, to allow the crowd workers to quit the task whenever they wanted. Potentially quitting a task is common in crowdsourcing markets: Considering the experience a crowd worker gains during a task and the opportunity costs of time, it might well be rational for the crowd worker to quit by simply abandoning the task. In the MTurk context this is referred to as not returning a HIT. The explicit option to quit aims at reducing experimenter demand effects and the relevance of a potential social norm to never give up.

Figure 4.3 illustrates the task and the feedback for all three studies: Figure 4.3a shows an example for a ROT with feedback. At any time during the ROT a crowd worker sees his own performance so far (here 7 completed sliders), his competitor's performance so far (here 11 completed sliders), and the next slider to be set to 50. In addition, the screen has a timer at the top and a quit button at the bottom. Figure 4.3b exemplifies the no feedback treatments; it is identical except that feedback on the competitor's performance is missing – this information is only disclosed after the ROT when the result is shown. The

[8] see `https://grails.org/` for more details; last accessed: September 6, 2016.

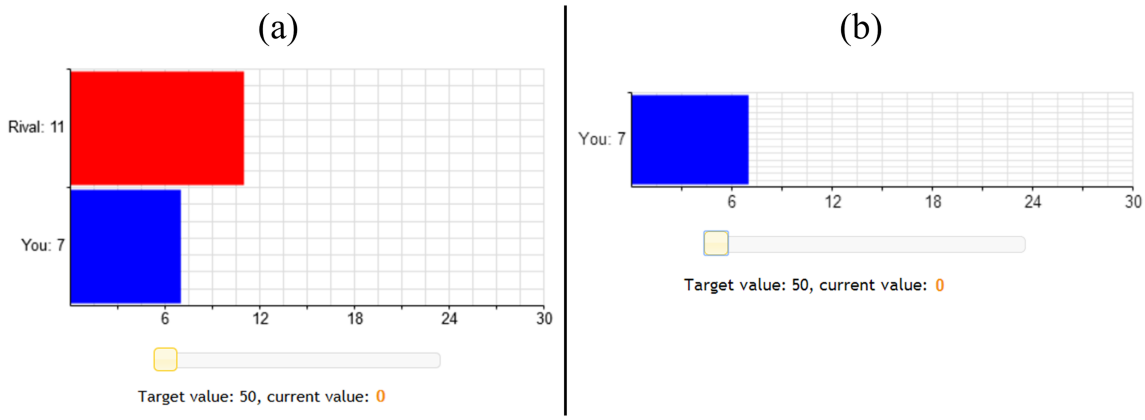


FIGURE 4.3: *User interface: feedback (left image), no feedback (right image).*

user interface for PR treatments is identical to the one for ROTs without feedback (Figure 4.3b); the subsequent payment scheme differs.

In statistical tests, I employ a 0.1 level to decide on the rejection of null hypotheses. More detailed information on p-values is provided in the result Subsections (4.5.2, 4.6.2, and 4.7.2) of this chapter. Design features that differ between the three studies are described below.

4.5 Study 1: Piece Rates versus Rank-Order-Tournaments

4.5.1 Study Design

Study 1 is a comparison of PR payments with ROT payments providing no feedback on the competitor's performance during the study. Presumably, performance depends on various individual characteristics like the individual capability to perform the task and other factors that might partially be captured by observing age, gender, and education (see Section 4.3). To account for this partially unobservable heterogeneity, a within-subject comparison for the two treatments (PR and ROT) is employed: Each participant participates in both payment schemes. Each participant plays a training round of the slider task for 30 seconds to get familiar with the task and the interface followed by two study rounds of

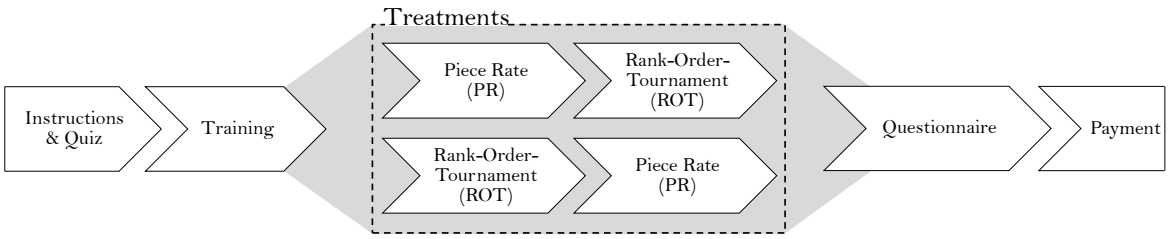


FIGURE 4.4: *Experimental process of Study 1.*

two minutes each, as depicted in Figure 4.4. One of the study rounds is under PR conditions, getting USD 0.02 per finished slider, while the other round is under ROT conditions, winning USD 1.00 if the participant finished more sliders than his competitor. If there is a tie after the experiment (i.e., both finished the same amount of sliders), the winner is decided randomly by the throw of a dice, a so called random tie breaking. Based on pre-tests, payments are calibrated such as participants achieve the same average payment in both mechanisms[9]. Hence, the differences in performance cannot be attributed to different expected or realized payoffs. In both treatments the participants get the same information – their own performance (Figure 4.3a). For the ROT, they are informed after the round if they won. To control for order effects, wealth effects, learning, and fatigue, the order of the two payment schemes is balanced. The number of finished sliders in PR and ROT is used as measure for the participants' performance (i.e., crowd workers' effort) to be compared between payment schemes.

4.5.2 Evaluation

149 participants took part in the first study. General statistics of participants are depicted in Table 4.1. 73 first worked under the PR scheme, then under ROT; 76 first worked under the ROT, then under PR scheme. Participants' age ranges from 19 to 66 years with a mean of 31 years. 41.6 % are female. The task took on average 11 minutes, and the average total payment was USD 1.63. Payment consists of a fix USD 0.50 show-up fee and payments for both incentive schemes. For PR, mean payment was USD 0.55 (standard deviation (SD)

[9]Based on a pre-test with 30 participants payments per finished task in the PR treatment were calibrated to match the expected outcome of the ROT, assuming that 50 % would win. For the PR payment, the assumption was made that participants would finish on average the same amount of tasks as the participants finished on average during the pre-test.

= .17), for ROT it was USD 0.58 (SD = .50). Payments in both incentive schemes are statistically indistinguishable (two-sided t-test, t = -.571, p-value = .569).

The relation of the two payment schemes with the participants' performance is evaluated in three ways: First, we count how many participants performed better in PR than in ROT and the other way round. Of 149 participants, 17 finished the exact same number of sliders under both incentive schemes, 63 performed better in the ROT than with PR and 69 performed better with PR than ROT. This data suggests that both incentive schemes are about equal: Given that one performs differently under PR and ROT, the likelihood of performing better under ROT is 48 % which is statistically indistinguishable from a random 50 % (two-sided binomial test, p-value = .664). Second, we compare the mean number of sliders finished in either treatment: 27.64 for PR (SD = 8.26) and 27.65 for ROT (SD = 8.39), as depicted in Figure 4.5. Again, not statistically significant different (two-sided matched pairs t-test, t = -.028, p-value = .978). Third, an ordinary least squares (OLS) regression is employed with *performance* as dependent variable (DV) while controlling for age, gender, education and the order effects. The binary variable *round*[10] equals zero for the first incentive scheme and equals one for the second. *Age* is measured in years, *education* in the following categories: some high school completed = 0, high school diploma = 1, some college completed = 2, associate's degree = 3, bachelor's degree = 4, master's degree = 5, doctorate = 6. The binary variable *tournament* is one for ROT and zero for PR. This is the focal variable in this study. The results are depicted in Table 4.2. Most importantly – but not surprisingly given the other tests described in this paragraph – a significant correlation between the treatment and performance is not observed.

	Overall	PR First	ROT First
#Participants	149	73	76
Female	186	90	96
Average Age	30.85	32.18	29.57
Average payment (USD)	1.63	1.70	1.57

TABLE 4.1: *General statistics of participants for Study 1.*

[10]The value of *round* depends on the order a participant worked on the treatments. If he worked first on PR then on ROT, PR is coded 0 and ROT 1 (73 participants). If he worked first on ROT and then on PR, ROT is coded 0 and PR 1 (76 participants).

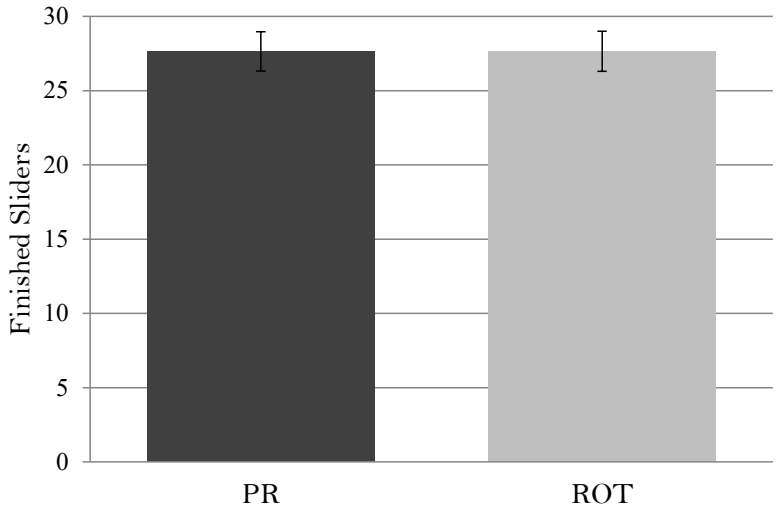


FIGURE 4.5: *Mean finished sliders (performance) by treatment, standard error bars for 95 % confidence interval.*

The absence of significance does not directly imply the absence of a relationship. Thus it is interesting to evaluate the marginal effect size of the tournament in explaining variance in performance. For doing so, a second regression analysis was implemented to obtain the residual R^2, i.e., without *tournament* as independent variable and compare it to the variance explained by *tournament* to calculate the effect size f^2 (Cohen, 1988, p. 407ff.). By convention, $f^2 = .02$ is termed a small, .15 a medium, and .35 a large effect. Here, the effect size f^2 turns out to be merely .00003, i.e., three orders of magnitude less than a small effect. The relation is not only statistically insignificant, it is economically meaningless: The estimated effect of running a tournament equals the estimated effect of increasing the participants' age by about three month which is not substantial given an average age of 31 years. Given the confluence of this evidence, the following result is formulated.

> **Result 1:** *Given equal expected payments, both piece rate and dyadic rank-order-tournament payment schemes without feedback on the crowd worker's competitive position result in equal performance of crowd workers.*

Further relevant results from Study 1 are that performance has a strong relation to age (older crowd workers perform worse than younger crowd workers)[11] and gender (males

[11]Note that participants' age ranged from 19 to 66 years. In our data this results in a linear relation. However, this gives no insight about an overall shape of the relation, since this observation does not span over all ages. It is assumed that younger and older ages, which were not observed, show a different relation.

DV (method)	Performance (OLS)
Intercept	33.886***
Age (in years)	-.319***
Gender (male)	2.905**
Education	.065
Round (second)	3.317***
Tournament (ROT)	.080
N	149
R^2	.242

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

TABLE 4.2: *Regression results for Study 1.*

perform better than females), but performance is not related to the education of participants (Table 4.2). This might be since the real effort task does not need a certain skill or knowledge to master. In the following studies, I consequently continue to elicit demographic information and use it as control variables in the analysis. In addition, participants' performance is strongly associated to the order of tasks (participants perform better in the second round). To avoid any confounding effects from the order of treatments, for the following studies a between-subject design is used and the sample size is increased to control for individual heterogeneity.

A ROT requires more effort and complexity in implementing, communicating, and controlling than PR payment. As this effort does not translate to higher performance, I conclude that the short and simple dyadic ROT studied in Study 1 is – for practical reasons – less suited than PR payments. This might, however, strongly depend on the ROT's design, most prominently the lack of feedback on a crowd worker's current competitive position. Whether such feedback is positively correlated to performance and renders a ROT worthwhile is the focus of Study 2.

4.6 Study 2: Feedback in Rank-Order-Tournaments – Different Strengths of Competitors

4.6.1 Study Design

Study 2 evaluates the relation of competitors' strength and feedback to performance. It is a between-subjects comparison of four treatments. In each treatment crowd workers first work on the slider task for 90 seconds with a PR payment of USD 0.01 per finished slider. The number of finished sliders is the proxy for capability. In addition, it allows crowd workers to get familiar with the task and interface. This data is not used to judge whether PR or ROT lead to higher performance. Second, crowd workers participate in a three minute dyadic ROT. For the ROT, each crowd worker is randomized to either of four treatments: no feedback on the performance of the competitor (NF), feedback on the performance of the competitor in a ROT with a strong competitor (FS), feedback on the performance of the competitor in a ROT with a mediocre competitor (FM), and feedback on the performance of the competitor in a ROT with a weak competitor (FW). The structure of this study is depicted in Figure 4.6.

Data for competitors is retrieved from historic data based on a pre-test; it is constant for each treatment in order to not induce unnecessary variance. The weak competitor finishes 27 sliders in three minutes time; the mediocre competitor 47 sliders; the strong competitor 66 sliders. The number of sliders a crowd worker finishes in the ROT is the measure for his performance. A crowd worker that finishes more sliders than his competitor wins USD 1.00.

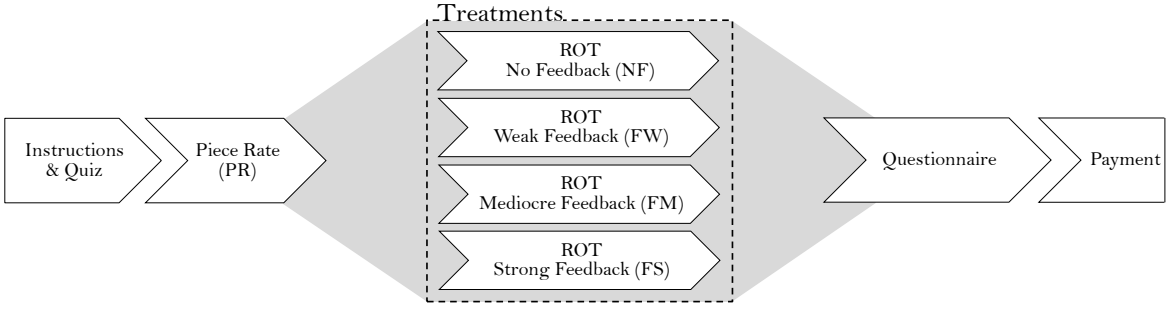


FIGURE 4.6: *Experimental process of Study 2.*

4.6.2 Evaluation

331 crowd workers participated. General statistics are depicted in Table 4.3. 97 participated in NF, 80 in FW, 74 in FM, and 80 in the FS treatment. Participants' age ranges from 18 to 66 years with a mean of 32 years. 39.9 % are female. The task took on average 8 minutes, and the average payment was USD 0.89. Note that differences between average payments were observed (ANOVA, p-value $< .001$). However, this is given by experiment design: In FW, the competitor was weak, in FM he was mediocre, and in FS the competitor was strong. Thus, the rate of participants winning was affected by the strength of the competitor. In the NF treatment the competitor strength was randomized, hence, average payment lies in between the other payments. However, given randomization of the treatments, each participant had equal chances to be in the respective treatments.

The moderated mediation model sketched in Figure 4.1 is evaluated with a set of eight regressions, following the general steps from Hayes' (2009) contemporary interpretation of the mediation and moderation analysis of Baron and Kenny (1986) and a bootstrap test of indirect effects, following Preacher and Hayes (2004). For the causal step mediation analysis, the correlation of the causal variables on the mediator (regression models 1 to 4 in Table 4.4) are established first and then the correlation of causal variables and the mediator on the outcome variable (regression models 5 to 8 in Table 4.4) are estimated. Task Completion is binary (completed $= 1$, not completed $= 0$). Strength of competitors is coded in three levels (weak, mediocre, or strong). In this setting, the statistical consideration of moderation differs from the conventional approach: Conventionally, feedback moderating the correlation of strength of competitors would be modeled by two direct effects (one from feedback, one from strength of competitors) and the interaction of these. In this model and experiment, strength of competitors is, however, not meaningfully defined in the absence of feedback. Without feedback, strength of competitors cannot be

	Overall	NF	FW	FM	FS
#Participants	331	97	80	74	80
Female	132	39	32	28	33
Average Age	31.87	31.08	31.02	35.80	30.05
Average payment (USD)	.89	.95	1.34	.77	.49

TABLE 4.3: *General statistics of participants for Study 2.*

correlated with either task completion or performance. Thus, moderation here results in four combinations: (i) no feedback (irrespective of the strength of competitors), (ii) feedback and a weak competitor, (iii) feedback and a mediocre competitor, and (iv) feedback and a strong competitor. Results of the analyses are provided in Table 4.4.

As expected, *capability* is substantially associated with *task completion* (regression model 1 in Table 4.4; support for H1 in the research model). *Feedback* is a dummy equal to 0 for NF treatment and 1 for FW, FM, and FS. The interaction of *strength of competitors* and *feedback* assesses the moderation. When facing a weak competitor and feedback is given, there is no significant relation to task completion compared to no feedback. On the contrary, when playing against a mediocre or strong competitor, there is a significant relation to task completion. Feedback makes crowd workers quit the task when facing a mediocre or strong competitor. Furthermore, a mediocre or strong competitor makes crowd workers quit more often compared to a weak competitor (regression model 3 in Table 4.4, significant effect of a mediocre or strong competitor interacted with feedback on performance). In total, feedback moderates the relation of strength of competitors and task completion (support for H6). The stronger the competitor the more likely a crowd worker quits the task resulting in a negative correlation (support for H4).

> **Result 2:** *Capability is correlated to task completion in a rank-order-tournament. Capable crowd workers finish the task more often.*

> **Result 3:** *Task completion is correlated to mediocre and strong competitors when feedback is given in a rank-order-tournament; it is not related to task completion when strength of competitors is weak. Crowd workers quit the task more often when facing stronger competitors.*

After establishing the correlations on the mediator task completion, we now turn to the correlations with the outcome. The results of ordinary least squares regressions (OLS) are depicted in columns (5) to (8) of Table 4.4. As expected (H3), task completion has a strong relation to performance. Crowd workers who complete a task also finish more sliders correctly. Capability has a direct relation with performance (support for H2). Capable crowd workers perform better than those who are not. The correlation of capability on performance is mediated by task completion. The more capable a crowd worker is the more likely he will complete the task, which will result in better performance. Giving

DV (method)	Task Completion (Logit regression)				Performance (OLS)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	1.885^{+}	1.899^{+}	1.899^{+}	1.899^{+}	5.455^{+}	4.608	4.608	4.608
Age (in years)	.017	.018	.018	.018	-.146**	-.140**	-.140**	-.140**
Gender (male)	-.396	-.314	-.314	-.314	1.326	1.356	1.356	1.356
Education	-.328*	-.327*	-.327*	-.327*	.103	.164	.164	.164
Task Completion (compl.)					20.009***	20.442***	20.442***	20.442***
Capability	.144***	.149***	.149***	.149***	1.236***	1.239***	1.239***	1.239***
Feedback	-1.264*		-.325	-1.438*	-.980		-2.278^{+}	-1.188
Weak x Feedback		-.325		1.113^{+}		-2.278^{+}		-1.091
Mediocre x Feedback		-1.438*	-1.113^{+}			-1.188	1.091	
Strong x Feedback		-1.765**	-1.440*	-.327		.598	2.876*	1.786
N	331	331	331	331	331	331	331	331
R^2	.206	.242	.242	.242	.707	.711	.711	.711

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$; for logit regressions, Cragg and Uhler's R^2 (Cragg and Uhler, 1970)

TABLE 4.4: *Regression results for Study 2.*

feedback about a weak competitor is correlated with performance compared to no feedback (regression model 6 in Table 4.4). Crowd workers who are informed about facing a weak competitor perform worse than without this information. We conclude that indeed frontrunners lay back when they know that they are frontrunners. On the contrary, giving feedback about facing a mediocre or strong competitor leads to no different performance than no feedback (regression model 6 in Table 4.4). The difference between a weak and a strong competitor is significant (regression model 7 in Table 4.4). The difference between a weak and a mediocre competitor (regression model 7 in Table 4.4) and between a mediocre and strong competitor is not significant (regression model 8 in Table 4.4). As hypothesized, a moderating relation of feedback on the relation of strength of competitors on performance is found. H5 is, however, only partially supported: As expected, with given feedback, playing against a weak competitor decreases performance; contrary to the expectation, when playing against a mediocre or strong competitor feedback does not increase performance. These associations are not associated with fatigue of crowd workers who play longer than those who quit the task, since it is controlled for task completion in the regressions.

> **Result 4:** *Capability is related to performance in a rank-order-tournament. Capable crowd workers perform better than those who are not. The relation is partially mediated by task completion.*

> **Result 5:** *Strength of competitors is related to performance in a rank-order-tournament. When feedback is given, there is a direct, unmediated negative correlation of weak competitors with performance. With mediocre or strong competitors, the negative correlation with performance is mediated by task completion.*

After the causal mediation analysis steps we now turn to the indirect effect and the effect sizes using Preacher and Hayes' (2004) bootstrap test. To do so, the dataset with four treatments is modified in seven sets of pairwise treatment comparisons to assess the analysis (Pedersen et al., 2011). All results are based on 10,000 bootstrap simulations with a sample size of 331. *Feedback* implements treatments FW, FM, and FS. The results are depicted in Table 4.5.

First, the NF treatment that resembles the tournament used in Study 1 is compared with all three feedback treatments (FW, FM, FS; model 1 in Table 4.5). When feedback is given,

Treatments	(1) NF – Feedback	(2) NF – FW	(3) NF – FM	(4) NF – FS	(5) FW – FM	(6) FW – FS	(7) FM – FS
Average mediation effect (95 % CI)	-1.940* [-2.715, -.132]	-.248 [-1.731, 1.212]	-1.425* [-3.951, -.122]	-2.417** [-4.354, -.608]	-1.301 [-3.507, .311]	-2.107* [-4.234, -.109]	.496 [-2.814, 1.640]
Average direct effect (95 % CI)	-.882 [-2.800, 1.074]	-2.218* [-4.383, -.051]	-1.057 [-3.625, 1.501]	.721 [-1.807, 3.278]	1.161 [-1.341, 3.760]	2.939* [.471, 5.420]	1.778 [-.922, 4.424]
Total effect (95 % CI)	-2.822^{+} [-4.729, .203]	-2.466^{+} [-5.195, .241]	-2.483^{+} [5.833, .276]	-1.696 [-4.815, 1.551]	-.141 [-3.325, 2.627]	.832 [-2.125, 3.841]	2.274 [-2.191, 4.490]
Proportion mediated	.687^{+} [-.427, 2.451]	.101 [-1.665, 1.221]	.574^{+} [-.705, 2.759]	1.425 [-8.683,10.292]	9.258 [-14.144,15.219]	-2.533 [-22.687,26.780]	.218 [-8.562, 9.023]
N	331	331	331	331	331	331	331

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

TABLE 4.5: *Mediation analysis results for Study 2.*

the significant total negative effect on performance is mediated by task completion. In comparison to giving no feedback, feedback on a weak competitor leads to no significant mediation but significant negative direct and total effects on performance (model 2 in Table 4.5). For a mediocre competitor, the significant negative total effect is mediated by task completion (model 3 in Table 4.5). For a strong competitor, interestingly, there is a significant negative mediation effect on performance via task completion; the total effect is, however, not significant as the mediation effect is partially offset by an (insignificant) positive direct effect (model 4 in Table 4.5).

Looking at the differences between the different strengths of the competitors (models 5 to 7 in Table 4.5) only the comparison of the two extremes – FW and FS – shows significant relations. Compared to facing a weak competitor, a strong competitor makes some crowd workers quit the task (mediation effect) while incentivizing others to perform better (direct effect). Both effects cancel each other about out, thus leading to an insignificant total effect. These results further underpin and detail our findings so far: The stronger the competitor, the more likely a crowd worker quits the task. Task completion thereby partially mediates the negative correlation of feedback and performance. When a crowd worker decides to complete the task, the weaker the competitor, the lower the crowd worker's performance.

In summary, Study 2 suggests that feedback is related to one's performance. No matter how strong or weak the competitor in a dyadic ROT performs, on average over all crowd workers it decreases performance. The mechanism of this negative correlation is either the mediation by task completion or a direct negative effect on performance. This result seems disillusioning for short dyadic ROTs showing leaderboards. It might, however, be driven by averaging over crowd workers facing competitors of different strength.

It still might be the case that a clever matching of crowd workers yields higher performance. Specifically, it is hypothesized that a competitor not substantially stronger or weaker but about on par with the crowd worker itself should result in the fiercest competition that does neither discourage continuation not allow to relax. This issue is addressed by Study 3.

4.7 Study 3: Feedback in Rank-Order-Tournaments – Equal Strength of Competitors

4.7.1 Study Design

Study 3 – evaluating the relations of an equally well performing competitor – consists again of a PR round measuring *capability* and a dyadic ROT measuring *performance*. Three treatments for the ROT phase are compared – no feedback about the competitor's performance (NF3; with a suffix 3 to denote Study 3), feedback about a mediocre competitor (FM3), and feedback about an equally good competitor (FE), as depicted in Figure 4.7. The first two exactly replicate the respective treatments from Study 2 (see Subsections 4.6.1 and 4.6.2). The new aspect of Study 3 is FE: Knowing a crowd worker's capability from the PR phase, a competitor for the ROT who – in the available historic data[12] – is closest to him in terms of capability is picked.

4.7.2 Evaluation

Overall, 394 crowd workers participated. General statistics are depicted in Table 4.6. 131 participated in NF3, 128 in FM3, and 135 in FE. Participants' age ranges from 18 to 66 years (mean 34 years, 47.7 % female). The task took on average 11 minutes, and the average payment was USD 1.69.

To assess the moderation mediation model, similar to Subsection 4.6.2, we first evaluate the results of a set of six regressions, following the general steps from Hayes (2009) con-

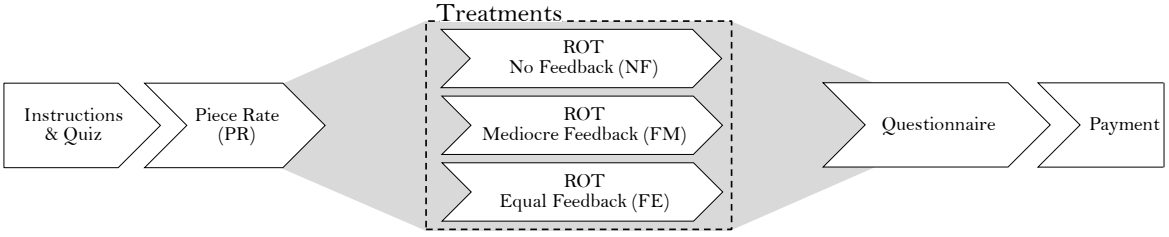


FIGURE 4.7: *Experimental process of Study 3.*

[12]The used historic data consists of the pre-test data and the data from Study 2 (see Subsections 4.6.1 and 4.6.2).

temporary interpretation of Baron and Kenny (1986) mediation and moderation analysis. Again, the correlation of the causal variables on the mediator (regression models 1 to 3 in Table 4.7) are established first and then the correlation of the causal variables and the mediator on the outcome (regression models 4 to 6 in Table 4.7) are implemented. Strength of competitors is coded in two levels (mediocre and equal). Second, we use Preacher and Hayes (2004) bootstrap test method, resulting in four dichotomous comparisons (NF3 – Feedback, NF3 – FM3, NF3 – FE, FM3 – FE) with each 10,000 bootstrap simulations. Results from the six regressions are summarized in Table 4.7 and results from the bootstrap analysis are summarized in Table 4.8.

First, reinforcing the results of Study 2 (see Subsection 4.6.2), *feedback* affects the mediator (*task completion*) negatively (regression model 1 in Table 4.7). The interaction of *strength of competitor* and *feedback* assesses the moderation. Again, facing a mediocre competitor negatively affects *task completion*. This confirms the findings of Study 2. Turning to the new aspect of Study 3 – a matched, equally well performing competitor – however, contrary to the expectations, negatively affects *task completion* (regression model 2 in Table 4.7). Moreover, differences between a mediocre and an equal competitor regarding their effect on *task completion* are not observed (regression model 3 in Table 4.7).

> **Result 6:** *The existence of an equally good competitor is negatively correlated with task completion in rank-order-tournaments. When feedback is given, crowd workers facing an equally good competitor quit the task more often than without feedback.*

Second, confirming the results of Study 2 as well (see Subsection 4.6.2), *feedback* affects the outcome variable (*performance*) negatively as well (regression model 4 in Table 4.7). The mediator (*task completion*) again effects performance (regression model 4 in Table 4.7). Hence, the effect of *feedback* is partially mediated by *task completion*. When

	Overall	NF3	FM3	FE
#Participants	394	131	128	135
Female	188	71	59	58
Average Age	33.79	33.72	33.30	34.32
Average payment (USD)	1.69	1.72	1.68	1.67

TABLE 4.6: *General statistics of participants for Study 3.*

DV and method	Task Completion (Logit regression)			Performance (OLS)		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	.212	.214	.214	3.889^{+}	3.886^{+}	3.886^{+}
Age (in years)	.002	.002	.002	-.125***	-.126***	-.126***
Gender (male)	.309	.312	.312	1.303^{+}	1.294^{+}	1.294^{+}
Education	.048	.044	.044	.485^{+}	.497^{+}	.497^{+}
Task Completion (compl.)				20.082***	20.088***	20.088***
Capability	.220***	.221***	.221***	1.255***	1.255***	1.255***
Feedback	-1.069*		-1.162*	-1.500*		-1.296
Mediocre x Feedback		-.966^{+}	.196		-1.710*	-.414
Equal x Feedback		-1.162*			-1.296	
N	394	394	394	394	394	394
R^2	.344	.345	.345	.777	.777	.777

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$; for logit regressions, Cragg and Uhler's R^2 (Cragg and Uhler, 1970)

TABLE 4.7: *Regression results for Study 3.*

Treatments	(1) NF3 – Feedback	(2) NF3 – FM3	(3) NF3 – FE	(4) FM3 – FE
Average mediation effect (95 % CI)	-1.219** [-2.382, -.237]	-.203^{+} [-2.428, .087]	-1.270* [-2.798, -.106]	-1.321 [-1.874, 1.245]
Average direct effect (95 % CI)	-1.541* [-2.954, -.153]	-1.757* [-3.399, -.177]	-1.331 [-2.985, .349]	.426 [-1.314, 2.139]
Total effect (95 % CI)	-2.761*** [-4.617, -1.029]	-1.960** [-5.028, -.818]	-2.602** [-4.926, -.595]	-.896 [-2.247, 2.449]
Proportion mediated (95 % CI)	.442* [.126, .881]	.104^{+} [-.048, .850]	.488* [.061, 1.290]	1.475 [-7.574, 6.444]
N	394	394	394	394

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

TABLE 4.8: *Mediation analysis results for Study 3.*

facing a mediocre competitor there is a significant negative effect on *performance*, when facing an equally good competitor there is not (regression model 5 in Table 4.7). Hence, I conclude that the negative effect from *feedback* on *performance* is partially mediated by *task completion* when facing a mediocre competitor and completely mediated when facing an equally good competitor.

> **Result 7:** *The existence of an equally good competitor is negatively correlated with performance in rank-order-tournaments. When feedback is given, crowd workers facing an equally good competitor perform worse than without feedback. This relation is completely mediated by task completion.*

Turning to the results of the bootstrap analysis presented in Table 4.8, the NF3 control treatment is first compared with the two feedback treatments (FM3, FE; model 1 in Table 4.8). Feedback has a significant negative total effect on performance. This negative effect is partially mediated by task completion. This confirms the results that giving feedback is negatively correlated with crowd workers' performance, in some cases through the mediation effect that crowd workers quit the task and in other cases because crowd workers do not quit but still do less compared to when not getting feedback.

The effects of a mediocre competitor on performance are replicated (model 2 in Table 4.8): A mediocre competitor leads to a total negative effect on crowd workers' performance with a comparatively low but significant mediation through task completion.

The new aspect of Study 3 is evaluating equally good – group matched – competitors in treatment FE. Compared to no feedback, FE leads to a significant negative total effect on performance which is, completely mediated by task completion, as there is no significant direct effect (model 3 in Table 4.8). Even though crowd workers have a reasonable chance to win at all times, since their competitor has about equal strength, they still quit the task resulting in lower performance. The correlation seems to be stronger (more negative) than the correlation induced by a mediocre competitor, but there is not a significant difference between FM3 and FE (model 4 in Table 4.8). Contrary to the expectations, group matching shows no positive or less negative correlation with performance, but rather a comparable negative correlation. Regarding the implementation overhead, it is therefore not recommended to implement such a matching, since it does not boost crowd workers' performance in a short term dyadic ROT. Reasons for this could be that feedback may just be a distraction or excels arousal and performance decreases in an aroused state.

4.8 Summary

Financial incentive schemes and their relationship with feedback and crowd workers' performance, have gained new relevance with the omnipresence of digital work places and crowdsourcing human work. In this chapter, dyadic rank-order tournaments (ROTs) and piece rates (PRs) as incentive schemes for short crowdsourcing tasks are investigated. Furthermore, their relationship to task performance in an anonymous digital workplace for activities that can be divided into small pieces that can be done (mostly) independently of each other is in focus of this chapter. A model on the correlates of crowd workers' performance in ROTs is introduced and tested with a series of empirical studies on MTurk – the most popular crowdsourcing workplace.

The best dyadic ROT in this study does not outperform a simple PR in terms of performance elicited from participants (RQ 1). Not all dyadic ROTs are equal, however: A relation to performance from giving feedback about the competitor's strength is found (RQ 2). Feedback that a crowd worker is performing comparatively well does not show a relation to his tendency to complete the task but tends to reduce his performance. Potentially, as feedback signals that the crowd worker does not have to excel to win the competition, or it signals that low performance is the norm, or both. Feedback that shows that a crowd

worker trails behind increases his likelihood to quit the task. Underlying reasons could be that the crowd worker knows that he is about to lose (and hence the financial reward) and he cuts his losses in terms of time invested or he aims to work on tasks where he has a structural advantage over other crowd workers. Mediocre competitors lead to correlations in between. When competitors are group matched and hence compete against an equally strong competitor, performance is reduced as well. Reasons could be that crowd workers perform worse under pressure or are distracted by constantly checking the feedback on whether they are winning or not. However, this is to be evaluated in future work. Performance of crowd workers who obtain the feedback that they are comparatively weak but who nevertheless continue to work on a task, do not change their effort compared to receiving no feedback.

In summary, this results in a clear guidance how to set up the two studied incentives in an anonymous crowd labor market for distributable work: A simple piece rate payment is better than a short dyadic tournament as incentive for simple short crowdsourcing tasks, as it is easiest to implement and unbeaten in terms of crowd workers' performance. Holding a short dyadic tournament does not offer performance benefits – if one does it anyways, during the tournament one should not provide a leaderboard or feedback on crowd workers' relative performance. Selectively matching crowd workers to homogeneous groups seems not to be worth the effort, as it decreases their performance in such a tournament setting.

The main contribution of this chapter is threefold: First, it implements existing evidence of incentives and feedback in tournaments via a theoretical model. Second, it studies the model and compares two common incentive schemes used in crowdsourcing in a series of three studies. Third, it provides guidance for crowdsourcing practitioners on how to structure payment schemes for their crowd workers. It thereby partially answers the practitioner question on how to design crowd labor tasks and contributes to theoretical discussion of designing and developing digital workplaces in general.

The limitations of this chapter are straightforward and include the following: First, three discrete levels of the strength of a competitor (Study 2) and equally strong competitors (Study 3) are evaluated but a continuous competitor strength is not observed. Expanding the analysis in this direction might show that moderation of the effect between strength of competitors and performance by feedback is non-linear. Second, even though

the experiment was applied on a crowdsourcing platform (MTurk), the slider task (chosen to provide experimental control on incentives to perform the task) is a rather unnatural task and the short 3-min dyadic tournaments are rather small tournaments. Results might not extend to longitudinal and more complex ROT settings with many participants. In order to increase external validity even further, a next step might be to explore tasks more common to crowdsourcing, to scale up the tournaments (length and participants), and to camouflage the experimental context. Results might differ when tournaments are played over a longer timeframe with more participants. Third, the feedback system was rather simple. More complex leaderboard and feedback designs might induce different results. Fourth, crowd workers' characteristics like, e.g., personality traits are not in focus of this work. However, they might show, that for some parts of the population, ROTs indeed spur performance. Last, the scope of this chapter are tasks that aim at the collection and subsequent aggregation of crowd input; settings in which the employer is interested in only a single best solution are not studied and the results might not extend to such settings.

Future research might investigate causality among the constructs studied in this paper. In the three empirical studies, several constructs depicted in Figure 4.1 are either controlled by the researchers (strength of competitors, feedback), or they are given exogenously by the nature of participants and vary at most marginally during the short duration of studies (age, gender, education). Pooled with random assignment of participants to treatments, it appears reasonable to hypothesize causation where these constructs correlate with capability, task completion, and performance. Testing for such causation and further investigating the underlying mechanisms is up to future research. In addition, the analysis could be extended to more complex tasks with a longer duration (see Chapter 5, which focuses on long lasting ROTs). Other crowdsourcing settings, specifically tasks where the employer is only interested in the single best solution and tasks that require collaboration among crowd workers should be evaluated (see Chapter 5). Furthermore, it might be fruitful to design tournaments which invoke intrinsic motivation to increase performance (see Chapter 6, which focuses on non-monetary incentives). In addition, future work should disentangle the effects of social norms and financial incentives on crowd workers' performance.

To sum up, this section focused in particular on systematically investigating the effect of different payment schemes and feedback on a crowd worker's effort. Especially in ROTs

crowd workers' effort is influenced by the feedback scheme the platform designer uses. In contrast to this chapter, the following chapter focuses on the effect of ROTs on a crowd worker's risk attitude in a long lasting tournament (see Chapter 5). Thereby the overall understanding of a crowd worker's behavior is improved and the insights can be used by platform designers. Not at least, risk-taking is believed to be an underlying reason of different crowd workers' efforts in ROTs (Eriksson et al., 2009a).

Chapter 5

Risk-Taking and Crowdsourcing: Tournament Modes and Rankings

> “ Competition is not only the basis of protection to the consumer, but is the incentive to progress.”
>
> HERBERT HOOVER, 1930

IN contrast to the last Chapter, which focused on crowd worker effort, this chapter focuses on the risk-taking of crowd workers in different ROT setups. It thereby evaluates a subsidiary behavioral aspect of crowd worker behavior. Risk-attitude could be an explanation for high effort variances in crowd tournaments (Eriksson, Teyssier, and Villeval, 2009). In addition platform designers can benefit from these new insights as a crowd worker's risk attitude can improve outcomes in certain settings. This chapter thereby evaluates Research Question 3 and its Sub-Questions 3.1, 3.2, and 3.3 (see Section 1.2 of Chapter 1).

Research Question 3: *How is risk-taking behavior affected in long lasting rank-order-tournaments?*

Research Question 3.1: *Does the tournament mode (teams or individuals) affect risk-taking behavior?*

Research Question 3.2: *Does the tournament progress affect risk-taking behavior?*

Research Question 3.3: *Does the ranking position affect risk-taking behavior?*

This chapter builds on and extends the collective work: Straub, Teubner, and Weinhardt (2016). It starts with introducing and motivating the context and importance of risk-taking in crowdsourcing scenarios (Section 5.1). Section 5.2 presents related work and posits this chapters' experiment in the literature on ROTs and risk-taking. In Section 5.3 the experimental design is described. Results are presented in Section 5.4. Section 5.5 provides concluding remarks and shows pathways for future research.

5.1 Introduction

Electronic markets are developing rapidly. New and IT-enabled business models, features, and standards arise, for instance recommender systems and crowdsourcing platforms. Such platforms use the input of workers all over the globe to co-create value. There are several types of such platforms with different aims: Prediction markets aim to forecast uncertain events (Kranz et al., 2014), innovation contests generate creative solutions for a given problem (Terwiesch and Xu, 2008), design contests are used to create aesthetic and meaningful graphical designs (Araujo, 2013), and crowdsourcing platforms such as MTurk[1] try to outsource "manual" tasks or are being used for experimentation (Paolacci et al., 2010) (see Chapter 2 for more details).

One dimension of crowdsourcing is given by the incentive schemes (Kittur et al., 2013). The most common types here are PRs and ROTs (see Chapter 4 and Subsection 2.1.1 of Chapter 2). In ROTs, companies usually only pay for satisfying solutions (see Subsections 2.1.1 and 2.1.2 of Chapter 2 for more details). Therefore they face essentially no risk of failure.

[1] `https://www.mturk.com/mturk/welcome` last accessed: September 6, 2016.

ROTs often use rankings as feedback mechanism to inform participants about their current evaluation (see see Subsection 2.1.1 of Chapter 2). Such feedback schemes are often thought of as a supporting element to motivate (continued) participation, whereas there is also evidence that introducing competition – or even any rewards at all – can also have detrimental effects, especially for creativity-centered tasks (Frey and Jegen, 2001; Frey, 2012).

Crowdsourcing tournaments may be differentiated with respect to whether and how the crowd workers are grouped. In order to harness potentially beneficial group effects (positive feedback, motivation, cohesion, belongingness, social presence, etc.), the issuer may group the crowd workers into teams. In contrast, to stronger emphasize individualistic aspects, participants may compete individually.

Although crowdsourcing is very attractive for companies, there is only little research on online crowdsourcing ROTs design itself. A key challenge here is to find an adequate operational mode, incentives, and feedback mechanisms for such ROTs (Archak and Sundararajan, 2009; Shaw et al., 2011). ROTs, independent from online and crowd-based scenarios, have previously been analyzed in economic laboratory experiments (van Dijk et al., 2001; Eriksson et al., 2009a; Eriksson et al., 2009b; Bracha and Fershtman, 2013). Much of the existing literature focuses on factors directly effective with respect to effort (van Dijk et al., 2001; Eriksson et al., 2009a; Eriksson et al., 2009b; Bracha and Fershtman, 2013) (see Chapter 4). The role of risk-taking behavior in tournaments and its interplay with feedback (i.e., ranking) and grouping (teams or individuals) has experienced much less attention (Nieken and Sliwka, 2010) (see Chapter 4). Risk-taking, however, may determine a crowd worker's contribution quite significantly. On the one hand, high risk-taking may be reflected in crowd workers handing in low-cost-low-quality work, which does not require much timely or cognitive effort. Such work of course has a higher probability of being rejected (and hence not paid) by the issuer – but this is the risk the crowd worker is willing to take. On the other hand, especially for tasks involving creativity, high risk-taking may be desirable as it may result in a variety of more diverse, offbeat, and extraordinary contributions.

Thus, risk-taking in tournaments is a factor worth considering. Issuers of crowdsourcing tasks may well benefit from better understanding how different incentive schemes and feedback mechanisms interact with human risk-taking behavior. High risk propensity may,

in this sense, help to unleash the crowd's creative potential and to avoid trodden paths and conformist solutions. For more straightforward tasks, risk aversion may be the stimulus of choice to ensure and increase output quality. Thus, this chapter introduces a factor potentially influencing a crowd worker's behavior in crowdsourcing ROTs – risk-taking. Thereby this chapter extends the findings of Chapter 4.

In this work, it is thus considered how (i) tournament mode (teams or individuals), (ii) tournament progress, and (iii) the position in the ranking affect an individual's risk-taking behavior. Overall, this is reflected in the three main research questions of this chapter presented in Section 1.2:

> **Research Question 3.1:** *Does the tournament mode (teams or individuals) affect risk-taking behavior?*
>
> **Research Question 3.2:** *Does the tournament progress affect risk-taking behavior?*
>
> **Research Question 3.3:** *Does the ranking position affect risk-taking behavior?*

These questions are addressed by means of an online experiment along with the FIFA World Cup 2014. Participants bet on the outcomes of all 64 matches in order to earn credit. The likelihood of the outcome of the matches is reflected by odds which account for the possible credit participants can earn. Thereby these odds reflect the risk-taking of participants. Hence, contrary to Chapter 4, this chapter abstracts from the factor effort and concentrates on risk-taking behavior specifically.

5.2 Background

Crowdsourcing tournaments (e.g., 99designs[2], threadless[3]) offer a mechanism to distribute work to an open workforce for comparatively low cost (Leimeister, 2010; Paolacci et al., 2010). Two of the main challenges of crowdsourcing in general and its subgroup of crowdsourcing tournaments are to incentivize crowd workers adequately and thus secure quality and cost effectiveness (Shaw et al., 2011; Krause et al., 2016). Theories to assess behavior

[2] `http://en.99designs.de/` last accessed: September 6, 2016.
[3] `https://www.threadless.com/` last accessed: September 6, 2016.

in these environments must be seen in front of the background of more fundamental aspects such as motivation and competition, as well as their interplay in online settings (see Chapters 2 and 3).

The intuition to this chapter's research can be located in the streams on cognitive dissonance theory (Festinger, 1957; Cooper, 2011), responsibility diffusion theory (Wallach et al., 1964), and mechanisms of group conformity (Stallen et al., 2012). Cognitive dissonance theory posits that participants experience conflicting cognitive states as unpleasant and take actions in order to resolve or reduce such mental conflicts (Cooper, 2011). For the case of rankings, a (positive) self-image is certainly in conflict with being ranked low, while this conflict is less likely to occur when being ranked at or near the top of the list. Behavior with effect on a participant's score can hence be expected to be different for different positions in the ranking. In particular, cognitive dissonance suggests that falling behind others alters behavior compared to its "natural" form, and it does particularly more so than for being on top, by stressing the importance of catching up. While the general presence of a tournament mode was found to result in higher effort levels, rankings often prove to have a detrimental effect (Straub et al., 2015; Barankay, 2011) (see Chapter 4). With respect to risk-taking (chances on high rewards), rankings may be pointing to a possible path for conflict resolution, if ranked at an unsatisfying position. Hence, it may be hypothesized that positions at the bottom of the ranking entice higher levels of risk-taking.

With regard to actions relevant to the entire group, responsibility diffusion theory states that individuals feel less responsible for their actions (and their consequences) in groups compared to acting individually (Fraser et al., 1971; Stoner, 1968; Zajonc et al., 1968). Effort in groups was shown to often suffer due to this diffusion effect (Petty et al., 1977), also referred to as "social loafing" (Latane et al., 1979). Different levels of perceived responsibility may also lead to a shift of risk-taking behavior. This can be due to a lack of motivation to determine the group-optimal risk strategy, as this may require a lot of effort itself. Subjects may thus take larger risks as they would shine in the light of success but presumably rely on the others in case of failure. In contrast, in groups (compared to individual decisions) there are – after all – others for which participants may feel responsible and accountable and thus prefer a more moderate, i.e., less risky, behavior.

Moreover, mechanisms of group conformity may entice some team members to adapt the strategies of other members within the group (Stallen et al., 2012), e.g., of those with

particular good performance. To the contrary, preferences for individualism may result in choice differentiation (Ariely and Levav, 2000). This may result in a plethora of different constellations, depending on which strategies were successful early on. In this chapter, this effect is ruled out by making the group members' decisions unavailable to others.

In the following subsections, some light will be shed on the factors risk-taking in ROTs and the relationship between risk-taking and creativity by illustrating experimental and empirical evidence on these matters.

5.2.1 Risk in Rank-Order-Tournaments

In crowd tournaments (e.g., design contests, open innovation contests, etc.) a crowd worker's effort may be affected by the crowd workers' risk-taking behavior. Many studies in this regard observed high variance in effort (Bull et al., 1987; Harbring and Irlenbusch, 2003; van Dijk et al., 2001). Risk-taking may be one explanation for high effort variance in tournaments (Eriksson et al., 2009a) (see Chapter 4). In order to investigate participants' risk taking behavior, it is therefore crucial to better understand the relation between tournaments and rankings. Hvide (2002) find that individual risk-taking depends on the modification of the tournament: High rewards for the highest performance led to high risks, but not explicitly to hard work.

In the domain of sport psychology and motor racing, Bothner et al. (2007) consider the impact of positions during races on risk-taking. Using data from the NASCAR professional sports car racing series, they find that "pressure from below" induces risk-taking. "Pressure from above", in particular when the participant sees the opportunity to advance in rank, in contrast, has a smaller impact on risk-taking.

As Taylor (2003) shows, the gap from sports to corporate decision making is not too large. He examined risk-taking behavior in a mutual funding competition, in which two fund managers with unequal midyear performances compete against each other for new cash inflows. In short, leading managers avoid risks whereas those behind take risks. The rationale is that losing managers try to catch up. Qiu (2003) examined the behavior of mutual fund managers as well. The results show similar findings. Managers of funds who are close to top managers have greater incentives to increase their portfolios, i.e., taking a

greater risk, than top managers who are more likely to lock in their positions, i.e., reduce their risk.

Nieken and Sliwka (2010) find that risk-taking depends on the correlation of the outcomes of risky strategies. In their study, two agents with different scores simultaneously decide between a risky and a safe strategy. The safe strategy guarantees an outcome, while the risky strategies' outcome varies with a certain probability, but might lead to a higher payout. The experiment involves different correlations of the outcome of the risky strategy, if it is chosen by both players, i.e., the outcome is the same for both players if both choose the risky strategy (correlation $= 1$), it is the same half of the time both choose the risky strategy (correlation $= 1/2$), or they always get a different outcome when both choose the risky strategy (correlation $= 0$). The authors find that the leading player chooses the safe strategy more often whereas the trailing player nearly always plays risky when there is no correlation. However, when the outcome of the risky strategy is perfectly correlated, the leading player chooses the risky strategy more often than the competitor. This means that as long as risk-taking might lead to a competitive advantage, leading players will decrease their risk-taking, while trailing players increase their risk-taking. When there is no competitive advantage, i.e., when the risky strategy leads to same outcome for both players, leading players take risk as well.

Bracha and Fershtman (2013) find that people rather tend to make riskier decisions under tournament conditions than under performance-pay conditions. Participants make the decision whether to play a lottery (50:50 chance to win USD 35 or USD 10) or to get a guaranteed fixed amount (USD 22) two times. The first time the decision simply pays the payoffs. The second time the decision is set up as a head-to-head tournament. Only the player with the higher points receives USD 40. Contrary to the expectation that participants would choose the safe option more often under tournament conditions, the opposite occurred: Participants choose the riskier option more often.

Kräkel (2008) investigates how risk influences effort and the probability of winning a tournament. While contestants try to minimize their effort, they are willing to take risks when the chance of winning is increased. In uneven tournaments (one or several contestants are more capable to win) the underdogs always prefer risky choices to increase the likelihood of winning. This finding is supported by the results of Becker and Huselid (1992), who examined the risk effect of the contestants' "spread" in a tournament.

Panel data from different auto racing tournaments (i.e., NASCAR and International Motor Sports Association) showed that drivers take more risks when the spread to the first driver increased. In addition with the results from (Bothner et al., 2007) it seems, that "pressure from below" and "pressure from above" are not only influenced by the rank in a tournament, but the distance, and hence the likelihood, to the next rank as well.

In addition, Grund and Gürtler (2005) analyze the risk-taking behavior of professional soccer coaches. Their key finding is that during a match, risk-taking of the leading coach decreases with the goal difference. Pull et al. (2013) found that in tournaments with high contestant heterogeneity, an increase of the individuals risk will first lead to an enhancement of incentives. After reaching a critical risk level, it will then weaken incentives.

5.2.2 Risk-Taking and Creativity

Crowdsourcing works well for simple and repetitive tasks. Extending crowdsourcing to creative tasks on a large scale is more difficult to achieve (Kittur, 2010). However, creativity is an important factor in several crowdsourcing tasks (e.g., design contests, open innovation contests, etc.). While several crowdsourcing markets exist focusing on creative tasks (e.g., InnoCentive[4], threadless[3], 99designs[2]), it is not clear if they are optimally designed for tapping the crowd workers' creative potential. Crowdsourcing literature so far focuses on the collaboration of crowd workers to facilitate structures for collaborative creativity processes (Kittur, 2010; Yu and Nickerson, 2011; Nickerson et al., 2011). Though, it may be beneficial for creative crowdsourcing tasks to incentivize creativity, on an individual as well as on a collaborative level.

Literature on creativity suggests that risk-taking and intrinsic motivation are the main drivers for creativity (Hossain, 2012; Dewett, 2007; Amabile et al., 1986). Amabile et al. (1986) show in their model that intrinsic motivation is benefiting creative tasks. Compared to offering no reward (i.e., intrinsic motivation) offering a reward (i.e., extrinsic motivation) decreases creativity. Dewett (2007) analyzed the link from intrinsic motivation to creativity in more detail and found that risk-taking mediates this relation. Their survey data shows that risk-taking affects one's creativity positively. It is therefore vital to induce or reduce participants' risk-taking behavior based on the desired outcome on creativity.

[4] `https://www.innocentive.com/` last accessed: September 6, 2016.

5.3 Study Design

Based on the theory and the empirical insights illustrated in the last section (5.2), the design for an online experiment is presented, which systematically investigates the impact of tournament mode (individuals or teams competing), position in the ranking, and tournament progress in this section.

The experiment was conducted as an online experiment using a ROT payout scheme as incentive. The experiment consisted of two treatments, which varied the way the ROT was executed, i.e., the *team* (*TEAM*) and the *individual* (*IND*) treatment. In the TEAM treatment participants were ranked as a team competing against other teams. In the IND treatment, participants competed individually against other participants.

The experiment was conducted along with the FIFA World Cup 2014 in June and July 2014, using an online interface. Participants placed bets on the 64 matches of the tournament, could receive points if bets were placed correctly, and competed for payoffs, based on their total team's or individual points. A total of 72 participants took part in the study. Participants were recruited from a voluntary student pool using the Online Recruitment System for Economic Experiments (ORSEE, Greiner, 2004) at the Karlsruhe Institute of Technology (KIT). Participants were assigned either to the TEAM or the IND treatment, using a between-subject design, i.e., in each treatment there were 36 participants.

After an initial online registration on the experiment website, participants received instructions and user credentials via email (the instructions are provided in the Appendix's Chapter B). The participants' task then was to place bets on the 64 matches over the course of the World Cup using the website interface. The bets reflected the possible outcomes for each match – home win, draw, or away win in regular time (3-way-bets)[5]. Participants could discontinue participation at any time during the experiment, like in other online or crowdsourcing tournaments where the open call characteristic usually does not enforce participation due to anonymity and the lack of contracts.

Odds were retrieved from the sports bookmaker bwin.com[6] 72 hours prior to match

[5] 3-way-bets standardly close after regular game time. In tournaments where a winner needs to be decided, such as final rounds of the FIFA Wolrd Cup, the only assessable way to implement 3-way-bets is to bet on the outcome after regular game time. Otherwise, the draw outcome could not be realized, since overtime or a shootout always decides a winner of the game.

[6] `https://www.bwin.com/` last accessed: September 6, 2016.

start. The odds reflected the inverse probability of an outcome. Bookmakers try to secure that they earn money. Hence, they calculate, based on their insights, how likely the outcome of an event in their eyes is. To secure their outcome, more likely outcomes result in lower odds. Thereby, they try to minimize the possible profit a participant can earn by placing his bet. More unlikely outcomes are reflected by higher odds. Odds are the multiplier with which a participant's investment is multiplied if the outcome occurs. Following this, outcomes are more likely when odds are small, thereby less risky and a good proxy for risk-taking. Note that to the best of our knowledge the "true" probabilities of how the matches may end, (e.g., based on a model incorporating all relevant factors) are simply not available – the bookmakers odds are society's best guess.

Figure 5.1 shows a screenshot of the betting environment. Participants earned points according to the selected odd if their bet turned out to be correct. Wrong bets resulted in zero points. Forgetting to bet on a specific match resulted in zero points for that match. Placing bets was free of charge. Therefore, unlikely events yielded higher potential payoffs (points) than ex ante rather likely events. For example, the participant who placed bets in the betting environment, depicted in Figure 5.1, decided to bet on Brazil in the first match and therefore got the respective selected odd as points (1.30 points), since Brazil actually won against Croatia (match result – 3:1). However, this participant received zero points for the wrong bet on Spain against Netherlands.

Participants were able to bet on the outcome of a match between 72 hours and 5 minutes before the match started. Betting on all 64 matches was incentivized with an additional payoff of 5€, to prevent dropouts from the experiment.

FIGURE 5.1: *Betting environment.*

rank		name	points	#correct
1	(50€)	der_bomber_der_nation	69.06	35
2	(25€)	der_metzger	68.61	34
3	(15€)	kaiser	64.24	32
4		der_titan	64.00	29
5		flipper	62.84	32
6		tante_käthe	47.36	26

FIGURE 5.2: *Individual ranking.*

rank		community	points/player	points	#player
1	(6 x 50€)	Group E	63.92	383.54	6
2	(6 x 25€)	Group D	63.25	379.47	6
3	(6 x 15€)	Group B	56.07	336.40	6
4		Group C	55.11	330.68	6
5		Group A	53.33	320.00	6
6		Group F	52.70	316.18	6

FIGURE 5.3: *Team ranking.*

Half of the participants, 36 students, were randomly divided into 6 groups of 6 participants each. These participants competed against each other *individually* within their group (*IND*). Each participant received points according to the respective selected odd for each correct bet. After each match, the online ranking was updated (Figure 5.2). The own position was highlighted (position 1 in Figure 5.2). After the experiment participants were paid according to their final individual rank – 1st rank 50€; 2nd rank 25€; 3rd rank 15€. Additionally, participants were able to see an alphabetic list containing their own group and the other groups. Please note that this information was not relevant for payoff.

The other half was also randomly divided into 6 groups of 6 participants each. However, these groups competed as a *team* against the other teams (*TEAM*). The individual points of each participant in a team were added up to a team score. After each match the team ranking was updated showing payoff, the total amount of points, average points per participant, and positions of all six teams (Figure 5.3). The own team position was highlighted (here position 6). Participants were paid according to their team's final rank – 1st rank 50€ for each team member, 2nd rank 25€ for each team member; 3rd rank 15€ for each team member. In addition, participants were able to see an alphabetic list showing their own team's individual performance. Again, please note that this information was not relevant for payoff. Hence, both treatments were identical – except with respect to the payoff rules. In both treatments, participants could only see the (assigned) username of the other participants[7], whereas communication among participants was not enabled. The same 6 names were used within all teams.

Before the actual experiment started, participants were asked to answer comprehension questions by e-mail in order to ensure proper understanding of the payoff rules and the entire procedure. Four participants were replaced in response to not answering the questionnaire. After the experiment, participants filled out questionnaires assessing individual risk aversion (Holt and Laury, 2002), competitiveness (Griffin-Pierson, 1990), and the "Big Five" personality traits (Rammstedt and John, 2007). Moreover, they were asked to assess their experience during the experiment with regard to perceived social presence (Gefen and Straub, 2004), perceived competition (Byun and Mann, 2011; Vandercruysse et al., 2013), group belonging (Bollen and Hoyle, 1990; Greene, 2004), arousal (Liu et al., 2013), and enjoyment (Hassanein and Head, 2005) (see Appendix B). Afterwards they received their payment. This experimental process is depicted in Figure 5.4.

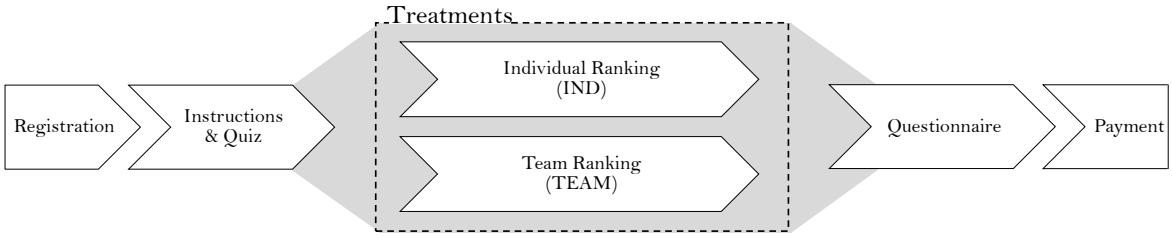


FIGURE 5.4: *Experimental process of Chapter's 5 study.*

[7]Nicknames of famous soccer players were used as usernames. Each group consisted of the same 6 usernames.

5.4 Evaluation

Risk-taking is assessed by the odds that were selected by the participants as described in Section 5.3. Assuming that the provided odds incorporate all relevant available information, each choice (home, draw, away) has the same expected outcome, but higher odds are more risky than lower odds as they represent a lottery with a higher variance. Moreover, odds represent the inverse likelihood of an event. Events more likely are represented by small odds of bookmakers, while high odds reflect unlikely events. Hence, odds represent a good proxy for how risky it is to bet on an outcome of an event. The higher the odd is, the more unlikely is the outcome of the event and in conclusion it is more risky to bet on this event since chances to win are more unlikely. Following this, the selected odd by the participant proxies a participant's risk-taking. In statistical tests, I employ a 0.1 level to decide on the rejection of null hypotheses. More detailed information on p-values is provided below.

Table 5.1 shows general statistics about participants. Overall 72 participants (56 male, 16 female) participated in the study. Average age was 21 years. All participants were enrolled in university and the majority of participants had a background in economics (62.5 %). All participants completed the experiment, but not all participants placed their bets on all games. Participants placed a total of 4,401 (out of a possible $72x64 = 4,608$) conclusive bets, representing a betting rate of 96 % of all possible bets. Moreover, some participants placed several bets on the same game, thereby changing their opinion over time. Overall 6,380 bets were placed – 4,401 conclusive and 1,979 non-conclusive bets.

	Overall	IND	TEAM
#Participants	72	36	36
Female	16	7	9
Economics	45	24	21
Average Age	21.44	21.81	21.08
#Bets	6,380	3,439	2,941
Conclusive Bets	4,401	2,193	2,208
Non-Conclusive Bets	1,979	1,246	733

TABLE 5.1: *General statistics of participants.*

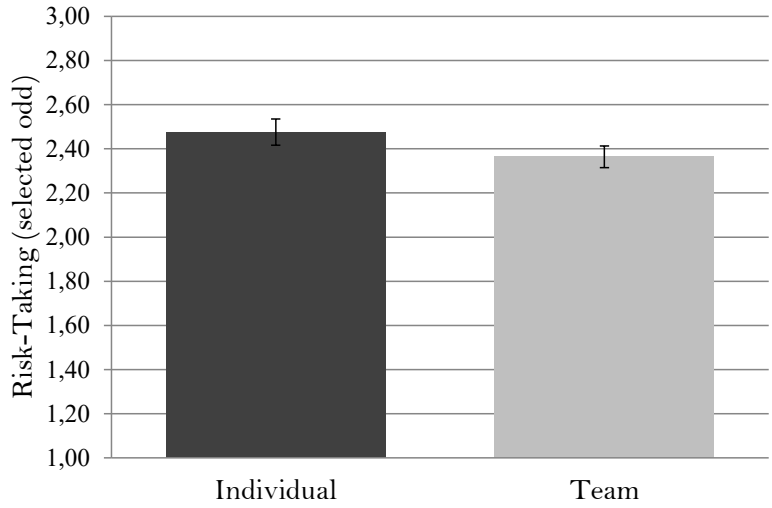


FIGURE 5.5: *Mean risk-taking (selected odd) by treatment, standard error bars for 95 % confidence interval.*

The main focus of this research lies on betting behavior, i.e., which degree of risk participants are willing to take in their respective treatment condition, with their respective position in the ranking, and tournament progress. As a proxy for risk, the selected odd (out of 3) for a given participant and match is used.

Overall, taking both treatments into account (IND and TEAM), the mean risk-taking (selected odd) is 2.42 (SD = 1.310). Moreover, we can observe higher mean risk-taking (selected odd) in the IND treatment (mean = 2.476; SD = 1.423) compared to the TEAM treatment (mean = 2.364; SD = 1.185), as depicted in Figure 5.5 (two-sided t-test, t = -2.838, p-value < .01). However, this analysis does not control for the timely sequence, which is addressed in later analyses.

> **Result 1:** *Participants in the individual treatment take higher risks compared to participants in the team treatment.*

Table 5.2 shows descriptive statistics of risk-taking (selected odd) per treatment and rank. Overall, taking both treatments into account (IND and TEAM), the payout-relevant ranks (rank 1-3) we can observe lesser risk-taking (selected odd) (mean = 2.32; SD = 1.13) than in non payout-relevant ranks (rank 4-6: mean = 2.56; SD = 1.51). While this effect is the same in both treatments, we can observe, independent of (non) payout-relevant ranks (Figure 5.5), higher risk-taking (selected odd) in the IND treatment (rank 1-3:

	Overall		IND		TEAM	
	Mean	SD	Mean	SD	Mean	SD
All	2.42	1.31	2.48	1.42	2.36	1.18
Rank 1	2.33	1.26	2.28	1.04	2.31	1.11
Rank 2	2.34	1.26	2.44	1.49	2.22	.91
Rank 3	2.44	1.13	2.40	1.16	2.35	1.15
Rank 4	2.47	1.33	2.60	1.52	2.44	1.12
Rank 5	2.49	1.23	2.66	1.32	2.32	1.08
Rank 6	2.53	1.66	2.75	2.13	2.64	1.67
Rank 1-3	2.32	1.13	2.35	1.19	2.29	1.07
Rank 4-6	2.56	1.51	2.67	1.69	2.46	1.32

TABLE 5.2: *Mean and standard deviation of risk-taking (selected odd) by treatment and rank.*

mean = 2.35; SD = 1.19 – rank 4-6: mean = 2.67; SD = 1.69) than in the TEAM treatment (rank 1-3: mean = 2.29; SD = 1.07 – rank 4-6: mean = 2.46; SD = 1.32).

However, the data was collected as a time sequence over the course of the tournament. Thus, the data has a panel structure as it faces a timely sequence of 64 decisions (corresponding to the 64 matches) for 72 participants. Hence, to evaluate this descriptive finding and to address the Research Questions 3.1, 3.2, and 3.3, a generalized least squares (GLS) regression analysis with subject random effects is implemented. The dependent variable is the *selected odd*.

As independent variables, a dummy for the *IND* treatment is used, i.e., whether a participant was allocated in the *IND* or in the *TEAM* condition. Moreover, to capture tournament progress, the number of remaining matches in the tournament (*#remaining matches*) is used as temporal panel variable. It states how many open matches were left at the time of placing the bet. This measure is more accurate than using the respective match number since placing bets was allowed within a time frame of 72 hours to 5 minutes before match start, which mostly included several other matches. By using the number of remaining open matches, this variable refers to a fixed state of information at the time of placing the bet. Furthermore, the product term *IND x #remaining matches* is used to capture possible interaction effects between treatment and tournament progress. Indicators are used for the position (1 to 6) in both the main ranking (actual relevant for payoff) and the position in the respective other ranking which was irrelevant for the experiment payoffs.

DV (method)	Risk-Taking – Selected Odd (GLS)		
	(1)	(2)	(3)
Intercept	2.641***	2.629***	2.727***
Treatment (IND)	.243^{+}	.244*	.241^{+}
#Remaining Matches	-.007***	-.007***	-.007***
IND x #Remaining Matches	-.004*	-.004*	-.004*
Rank (Relevant)		.003	.010
Rank (Irrelevant)			-.031*
N	4401	4401	4401

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{+}p < .1$

TABLE 5.3: *Mixed effects regression results.*

In order to provide some sense for the data and robustness, three specifications of regression models, building up on each other, are presented in Table 5.3[8]. Moreover, the relation between tournament progress, treatment, and risk-taking (*selected odd*) is illustrated in Figure 5.6.

As can be seen in Figure 5.6, Research Question 3.1 and 3.2 must be approached in a joint manner, as tournament mode and progress interact. Whereas overall risk-taking increases in the *TEAM* condition (by .007 units per match, p-value $< .001$), this increase is even higher in the *IND* condition ($.007 + .004 = .011$ per match, p-value $< .05$). Starting from a negligible difference at the beginning of the tournament, this treatment difference amounts to (marginally significant) .2432 units towards the end. In conclusion, participants increase their risk-taking behavior over the course of the tournament, where this increase is significantly more pronounced in the *IND* condition.

> **Result 2:** *Participants increase their risk-taking behavior over the course of the tournament.*

> **Result 3:** *Tournament mode and -progress interact. Participants in the individual treatment increase their risk-taking behavior more over the course of the tournament than participants in the team treatment.*

[8]Note that R^2's are not reported. Random effects models try to reduce variance of the data by default. Hence, there is no unified way of reporting the goodness of fit, researchers agreed on (Nakagawa and Schielzeth, 2013; Snijders and Bosker, 1994). Based on the possible goodness of fit models, the overall explained variance might be over- or underrepresented. Hence, R^2 or pseudo-R^2 are not reported.

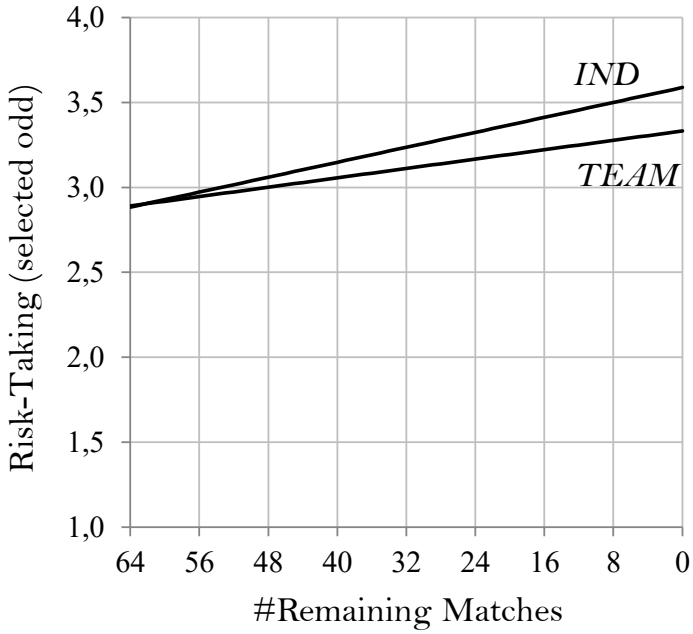


FIGURE 5.6: *Comparison of risk-taking (selected odd) by individual and team ranking over the course of the tournament (GLS regression estimates).*

With regard to the position in the ranking, the results are a little less obvious. While the actual relevant payoff in model (2) does not show to have any effect, including the irrelevant ranking position in model (3) changes the picture: Positions at the bottom of the ranking are associated with less risk-taking (by -.031 units per payoff irrelevant rank, p-value $< .05$). This observation is persistent in direction and magnitude with regard to removing the relevant rank variable and also to adding other control variables for ranking position (e.g., position in the *IND* or *TEAM* ranking). Controlling for gender and risk aversion did not reveal any significant effect or model improvement.

> **Result 4:** *Payoff-relevent ranks have no effect, while the payoff-non-relevant ranks have an effect on participants' risk taking behavior. Participants who are placed on positions at the bottom of the payoff-non-relevant ranking take less risks.*

Results from the questionnaire show that participants in the *TEAM* condition felt higher group belonging (Bollen and Hoyle, 1990; Greene, 2004) – mean scores from group belonging: $IND = 29.47$, $TEAM = 33.41$ (two-sided t-test, $t = 23.380$, p-value $< .001$) – and social presence (Gefen and Straub, 2004) – mean scores from perceived social presence:

IND = 15.81, *TEAM* = 16.31 (two-sided t-test, t = 3.780, p-value < .001). This might evidence that participants felt less responsible for their actions like the responsible diffusion theory states as introduced in Section 5.2 of this chapter (Fraser et al., 1971; Stoner, 1968; Zajonc et al., 1968). This is in line with the results of the questionnaire regarding the sensed competition. Participants experienced the *TEAM* condition as less competitive – mean scores from competitiveness (Griffin-Pierson, 1990): *IND* = 74.04, *TEAM* = 70.82 (two-sided t-test, t = 14.340, p-value < .001); mean scores from perceived competition (Byun and Mann, 2011; Vandercruysse et al., 2013): *IND* = 27.54, *TEAM* = 24.31 (two-sided t-test, t = -28.642, p-value < .001). One explanation for this could be, that the overall result is shared with the other team members and thus participants feel less responsible for their individual decisions. Hence, this might reduce perceived competition and consequently result in less risk-taking behavior.

> **Result 5:** *Participants in the team treatment felt higher group belonging and social presence, while participants reported that the treatment is less competitive than the individual treatment.*

5.5 Summary

In this chapter results of an experiment, evaluating the interplay of individual and team rankings with risk-taking behavior are presented. Results indicate that risk-taking increases over the course of a tournament (RQ 3.2) and that this increase is stronger for individual (compared to team) rankings (RQ 3.1). The ranking position itself exhibits an unexpected effect: Whereas the payoff relevant position does not appear to impact risk-taking, the complementary ranking does, decrease risk-taking for worse positions (RQ 3.3). It might be that participants compare their own strategy with their peers within their team (TEAM treatment) or with other teams (IND treatment) and adjust to perform better.

This chapter contributes to the discussion on how to design online crowdsourcing tournaments. Results indicate that implementing crowdsourcing tournaments should be assessed carefully with regard to risk-taking. Fundamental design variables like payoff rules and feedback information were found to impact risk-taking behavior. Besides higher levels of effort and variance (Bull et al., 1987; Harbring and Irlenbusch, 2003; van Dijk et al.,

2001), an issuer of a crowd task may hence also deliberately control risk-taking behavior. This might be especially helpful for tasks requiring high levels of creativity such as design and innovation contests. Adequate payoff and feedback schemes then help to tap the crowd workers' creative potential by inducing, encouraging, and emphasizing risk-taking behavior.

With regard to the shift in risk-taking behavior over time, Effron et al. (2015), for instance, found that participants in a self-reported coin-flip experiment tended to cheat on their last (expected) trials, represented by statistically unnatural high fractions of paying coin flips, which the authors related to the participants' anticipated regret over missing the (last) chance to enrich themselves. For the case of betting, cheating can be ruled out as an issue as there is simply no way to cheat. Hence, the increase in risk-taking towards the end rather stems from the increasing awareness that only so many matches remain. With less matches left, the chances (matches to bet on) to have a substantial lucky hit and advance in the ranking by earning points decrease. To some extent this is comparable to long-shot bets at the end of the day in horse races (Camerer, 2004). This finding is in line with the result that this increase in risk-taking is more pronounced for individuals than for teams, as the impact of such a bet is potentially mitigated by the other four group members' behaviors in the team treatment.

For platform operators, this means that creating long lasting tournament situations may be a suitable way to induce risk-taking behaviors and – implicitly linked to this – unleash silent creative potentials, since risk-taking is one of the drivers of creativity (Hossain, 2012; Dewett, 2007; Amabile et al., 1986), as already discussed in Subsection 5.2.2 of this chapter. Note that I do not claim that individually increasing participants' creativity is possible, as creativity was often found to be inhibited by extrinsic motivators (Amabile et al., 1986). Rather one should create a platform environment which encourages creative participants to actually pursue their "wild" ideas and not to hand in main stream – and hence presumably more "secure" – draftings.

Several limitations of this study exist. Since the World Cup is associated with a lot of emotion to many viewers, it might be that the participants' betting and hence risk-taking behavior is biased. Preferences for certain teams might have distorted behavior. Controlling for the participants' favorite team did not reveal interference with risk-taking. Preferring a second over a third team might have still had an influence, which cannot

be accounted for. Another issue stems from the fact that odds from bwin.com 72 hours before the match started were used. Of course, odds may have changed between 72 hours to 5 minutes before the start of the match – the time-frame participants could place bets. Differences here could have been used for arbitrage betting.

Future work might further disentangle the effect of a participant's current position in the ranking and risk-taking. Hence, the study should potentially be transferred to a setup where the position in rankings can be investigated in a more isolated manner such as a laboratory experiment. This might give more insights for potentially unobserved correlations in the current setup. In addition, future work should concentrate on the effect of non-monetary incentives on crowd workers, which platforms like Wikipedia[9] already put into practice (see Chapter 6). Thereby a broader picture of incentives for crowdsourcing would be given, as already outlined in Chapter 2. Furthermore, to increase external validity even more, further studies might focus on natural field experiments in this domain. Similar approaches without taking risk into account are provided by (Mason and Suri, 2012; Sorokin and Forsyth, 2008) and could be used as orientation of the task setup. To conclude, crowdsourcing has experienced rapid growth and can be expected to do so in the future, not only as a means of outsourcing, but also as complementary base of intelligence for machine learning systems. The underlying level of risk-taking may be valuable information and even a set screw for task quality. Hence, future studies should transfer the current insights to a setup where quality is the outcome, to further strengthen the insights for practitioners in the field of crowdsourcing.

To conclude, this chapter particularly focused on evaluating crowd workers' risk-taking over the course of different ROTs. In long lasting crowd tournaments risk-taking can be influenced by the ROT setup. Platform designers thereby have a strong tool at hand to influence a crowd workers behavior. With the complementary results of Chapter 4, platform designers thereby can influence a crowd workers behavior in monetary incentivized crowdsourcing settings based on the desirable outcome. However, not all crowdsourcing cases allow monetary incentives. Conclusively, the focus of the next chapter will be non-monetary incentives (see Chapter 6).

[9] `https://www.wikipedia.org/` last accessed: September 6, 2016.

Chapter 6

Non-Monetary Incentives for Crowdsourcing: Framing and Monetary Compensation

> “ Don’t think money does everything or you are going to end up doing everything for money.”
>
> VOLTAIRE

CONTRARY, to Chapters’ 4 and 5 focus – monetary incentives – this chapter focuses on non-monetary incentives. As outlined in Section 1.2 of Chapter 1, crowdsourcing platforms use both monetary- and non-monetary incentives. Hence, to give the full picture of crowdsourcing incentives as outlined in Subsection 2.1.1 of Chapter 2, this chapter addresses Research Questions 4 and 5 (see Section 1.2 of Chapter 1), focusing on non-monetary incentives.

Research Question 4: *Is framing capable of influencing behavior in non-monetary settings towards a certain goal?*

Research Question 5: *Can a customized framing for a non-monetary setting lead to similar behavior as in monetarily incentivized settings?*

This chapter builds on and extends the collective work completed during the course of the German government financed project, the so-called "Planspiel Flächenhandel"[1]. The remainder of this chapter is structured as follows. Section 6.1 focuses on introducing and motivating non-monetary incentivized crowdsourcing scenarios. Section 6.2 gives insight in related work (see Subsection 6.2.1) and explains the context of the project (see Subsection 6.2.2), this study is embedded in. Section 6.3 describes the study design. Section 6.4 then discusses the results. Section 6.5 provides concluding remarks and outlines possibilities for future work.

6.1 Introduction

Crowdsourcing does not always need monetary incentives (Horton and Chilton, 2010; Hammon and Hippner, 2012) (see Subsection 2.1.1 of Chapter 2). Many prominent crowdsourcing examples, such as Wikipedia[2], Google ReCAPTCHA[3], Stack Overflow[4], OpenIDEO[5], and many more, show that collaborative online work can be incentivized without monetary compensation. People working on such platforms are, for example, driven by intrinsic motivations, social status, enforcement to get access to a certain web page, and meaningful tasks (Huberman et al., 2009; Kaufmann et al., 2011; Pilz and Gewald, 2013). Therefore, to broaden the picture of crowdsourcing incentives, it is necessary to better understand these non-monetary incentives, as a platform designer might benefit from incorporating such.

Besides the already discussed different monetary incentive schemes (see Chapters 4 and 5), this chapter focuses on the motivational factor of non-monetary incentives. As classical motivation theory tells us (see Chapter 3), besides extrinsic motivational factors like money, intrinsic motivational factors are an important part of participants' motivation to participate (Ryan and Deci, 2000). However, it is difficult for crowdsourcing platform designers to specifically target intrinsic motivational factors, since a person is moved to take part out of his personal interest, fun, or curiosity (see Subsection 3.2.3 of Chapter 3).

[1] `http://www.flaechenhandel.de/` last accessed: September 6, 2016.
[2] `https://www.wikipedia.org/` last accessed: September 6, 2016.
[3] `https://www.google.com/recaptcha/intro/index.html` last accessed: September 6, 2016.
[4] `http://stackoverflow.com/` last accessed: September 6, 2016.
[5] `https://openideo.com/` last accessed: September 6, 2016.

Reminding people of their initial motivation to take part might reinforce their participation again. Explaining to them the scenario in which they take part and encouraging their altruistic behavior might spur their motivation to take part initially. Some platforms already do this by publicly sharing the overall goals they want to achieve, for example, OpenIDEO[5] deals with meaningful societal and ecological questions and Wikipedia[2] transports the idea of distributing knowledge. Both platforms make that clear on their websites and hence frame participants. Therefore this raises the question whether framing participants can be a motivational factor for participation and desirable behavior. Moreover it raises the question if a framing incentivizes comparable behavior to monetary based incentives. To be specific this chapter considers if similar behavior to monetary incentivized scenarios can be induced by a framing, even when monetary compensation lacks. Overall, this is reflected in the two main research questions of this chapter presented in Section 1.2 of Chapter 1:

> **Research Question 4:** *Is framing capable to influence behavior in non-monetary settings towards a certain goal?*
>
> **Research Question 5:** *Can a customized framing for a non-monetary setting lead to similar behavior as in monetarily incentivized settings?*

Research Question 4 is addressed by means of four online experiments – with and without a framing. Research Question 5 is addressed by means of four experiments (two online experiments and two controlled lab experiments) – with and without a monetary compensation (see Figure 6.1 for more details). All experiments were implemented in the context of a German government financed project, which tests a cap and trade system[6] with tradeable development rights (TDRs) to reduce land consumption in Germany, similar to the CO_2 emission trading used in the European Union[7] (United Nations, 1998) – the so-called "Planspiel Flächenhandel".

[6] A cap and trade systems is a scenario where the maximum use of a certain good is defined by a "cap" and allowances for using these goods are "traded". The total amount of allowances account for the overall "cap".

[7] `http://ec.europa.eu/clima/policies/ets/index_en.htm` last accessed: September 6, 2016.

6.2 Background

6.2.1 Incentives and Motives for Non-Monetary Crowdsourcing

Crowdsourcing platforms offer a variety of incentives to a potential crowd (Horton and Chilton, 2010). Incentives are monetary and non-monetary (Horton and Chilton, 2010; Hammon and Hippner, 2012) (See Subsection 2.1.1 of Chapter 2). Non-monetary incentives vary from information (Adamic et al., 2008; Jain et al., 2009), fun and entertainment (Von Ahn, 2006), to altruism, social status, and attention (Huberman et al., 2009). Non-monetary incentives are a main driver of motivation in certain crowdsourcing settings and sometimes even more important than classical monetary incentives as applied in standard labor markets (Kaufmann et al., 2011; Rogstadius et al., 2011; Chandler and Kapelner, 2013; Pilz and Gewald, 2013).

Brabham (2008) analyzes motives for participation in iStockphoto, a crowdsourcing applications where non-professional photographers upload and sell images on the platform. In this online survey it is shown that motives for participation range from money to developing skills. Brabham (2010) transfers this study to threadless, a t-shirt manufacturer who crowdsources his designs using ROTs. Using expert interviews he identified, that besides monetary incentives, community based motivations and skill improvement are strong motives for participation as well. Overall, monetary incentives are not the only motivation for participating in crowdsourcing.

Kaufmann et al. (2011) find that for many crowd workers, who work on paid crowdsourcing tasks intrinsic motivation aspects are important. They conducted a survey on MTurk and asked crowd workers for their motives to work on crowdsourcing tasks. Overall, motives of crowd workers range from extrinsic to intrinsic motivations. However, besides the most important incentive of immediate payoffs, intrinsic motivations are the main drivers for participation in crowdsourcing settings. Pilz and Gewald (2013) transfers the study from Kaufmann et al. (2011) to the software development field, in this case MobileWorks a software application development platform using a crowdsourcing approach. The findings show that in this setting extrinsic motivations dominates intrinsic ones. The most important factor is payment. However, extrinsic but non-monetary incentives, like learning and skill improvement are as well strong drivers of crowdsourcing participation.

Chandler and Kapelner (2013) use an experiment to compare intrinsic and extrinsic motivational factors in a crowdsourcing setting. They apply an experiment on MTurk with a labeling task of medical images. They applied three treatments with three different framings before the experiment. Participants were either not told the purpose of the task, additionally told that their results would be discarded, or told that they are labeling tumor cells. Results indicate that participation increases when the task has a meaning. However, quality is not affected.

Contrary to Chandler and Kapelner (2013), Rogstadius et al. (2011) find different results. They analyze intrinsic and extrinsic motivational factors in crowdsourcing settings as well. Using an experiment on MTurk they compared an unpaid scenario with two paid scenarios. Each of these scenarios included two framings varying in the identity of the requester, resulting in a 2x3 design. During the experiments, participants were instructed to count blood cells with a malaria infection. The first framing manipulation was framed as a non-profit health organization, helping malaria patients, while the second was framed as a private profit oriented pharmaceutical company. Results indicate that indeed intrinsic motivation (non-profit scenario) affects the result quality, while no differences between the unpaid and the two paid scenarios were observed in regard of crowd worker effort (amount of finished tasks). However, both Chandler and Kapelner (2013) and Rogstadius et al. (2011) show that framing in crowdsourcing settings influences participants. Somehow, participation and quality is effected by the scenario. Hence, a framing can induce positive effects on participation, besides extrinsic motivation factors like money.

6.2.2 The "Planspiel Flächenhandel"

The "Planspiel Flächenhandel" is a German government financed project. It evaluates a cap and trade system for land consumption with TDRs for land construction. The basic idea is that each new land construction in Germany needs to be confirmed by a TDR, which gives the right to use and construct buildings on this land. Each day 70 hectare of new land are planned to be used in Germany leading to urban sprawl and potential negative economic and ecologic consequences[8]. The German government plans to reduce this to 30 ha/d

[8]Numbers are based on the "Statistisches Bundesamt" and the "Umweltbundesamt" from Germany. For a detailed overview see the project's homepage `http://www.flaechenhandel.de/flaechenhandel/problem-flaechenausweisung` last accessed: September 6, 2016.

by 2020 (Federal Government, 2002; Coalition Treaty, 2013). One possible approach is the use of TDRs where the overall amount is capped, consequently restricting new land consumption. Thereby administrative cost are kept low as only the capped TDRs need to be distributed to the local authorities (Thorsnes and Simons, 1999).

To incentivize local authorities to further reduce their new planned land consumption, a market is implemented where local authorities can trade TDRs between each other. Thereby not used TDRs can be monetized and this money can be used for other purposes. Moreover such a market-oriented instrument can improve the distribution of TDRs as markets allow to transfer these warrants between the local authorities. Hence, economically more valuable building projects lead to a higher demand for TDRs in the market and would continue to be more likely realized, despite an overall TDR cap. Following this, a market-oriented instrument thereby has the potential to still achieve a high welfare in a situation where land use is capped, since such markets allow the shift of TDRs based on the demand, which is influenced by the evaluation of the building projects.

The European Union already uses a cap and trade systems to restrict CO_2 emissions[7] (United Nations, 1998). In the context of the "Planspiel Flächenhandel", a transfer of such a systems to land consumption is evaluated. The project committee of the "Planspiel Flächenhandel" implemented different steps to evaluate this cap and trade system for TDRs and to derive guidance for a potentially real world implementation of such a system (see the project's homepage for further details[1]). First, the projects' committee consist of experts and consultants in the area of land consumption, local authorities, and economists. Together, main aspects of such a system can be discussed from a theoretical and practical level. Second, through different workshops with employees of planning departments from local authorities, insights and important factors are collected about the process of planning and execution of new land consumption. Third, economists work together to design and conduct laboratory experiments with a high abstraction level to secure internal validity (Bizer et al., 2014; Meub et al., 2015; Proeger et al., 2015; Meub et al., 2016; Proeger et al., 2016). Fourth, contrary to the laboratory experiments, economists plan field experiments as well. In these field experiments real data from local authorities is used and planners from local authorities take part. As extension, additional information is shown, which is important for planning actual new land constructions but not necessarily important for the decisions throughout the experiment. Thereby internal validity is not in focus,

but a high external validity is achieved. Fifth, these field experiments are replicated in a laboratory setting to compare the behavior of students and planners from local authorities. Sixth, results are discussed with the local authorities and their feedback is incorporated in a final manuscript.

Overall, these steps give an extensive insight in how a cap and trade system for land consumption would work. The main goals of the project are to test whether planners from local authorities can handle a trading system, how different auction mechanisms affect the distribution of TDRs, if local authorities would pay realistic prices for TDRs and, hence, do not overdraw their accounts. Furthermore it is tested if a cap on TDRs lead to a new evaluation of building projects and in addition, if building projects with higher fiscally evaluation are more likely to be realized.

This chapter reports about the field experiment and its replication with student participants in a laboratory. It focuses if a framing induces similar behavior as monetary incentives. Overall project results, and the implications for a cap and trade system for land consumption can be read in an information paper series on the projects' homepage[1].

6.3 Study Design

The study consists of six experiments as depicted in Figure 6.1. Four field experiments consisting of 38 land construction planners representing overall 97 German local authorities and two laboratory experiments consisting of 76 students – 38 students each – were conducted. The six experiments were conducted in four sessions. Overall, the experiment design is divided in settings (NoFr, Fr, NoMo, Mo), treatments (NA, A), and experiments (1-6) as depicted in Figure 6.1.

Settings vary in whether participation is voluntary, and no monetary compensation is implemented (NoMo setting – NoMo for not monetary), or if participants receive a monetary compensation (Mo setting – Mo for monetary). Furthermore, settings vary in whether participants receive a framing before the experiment (NoFr setting – NoFr for no framing; Fr setting – Fr for framing). The framing includes instructions to behave realistically, as if the decision throughout the experiment would be real and highlighting the importance of the project. In addition the Fr setting includes a budget restriction per year, while the NoFr

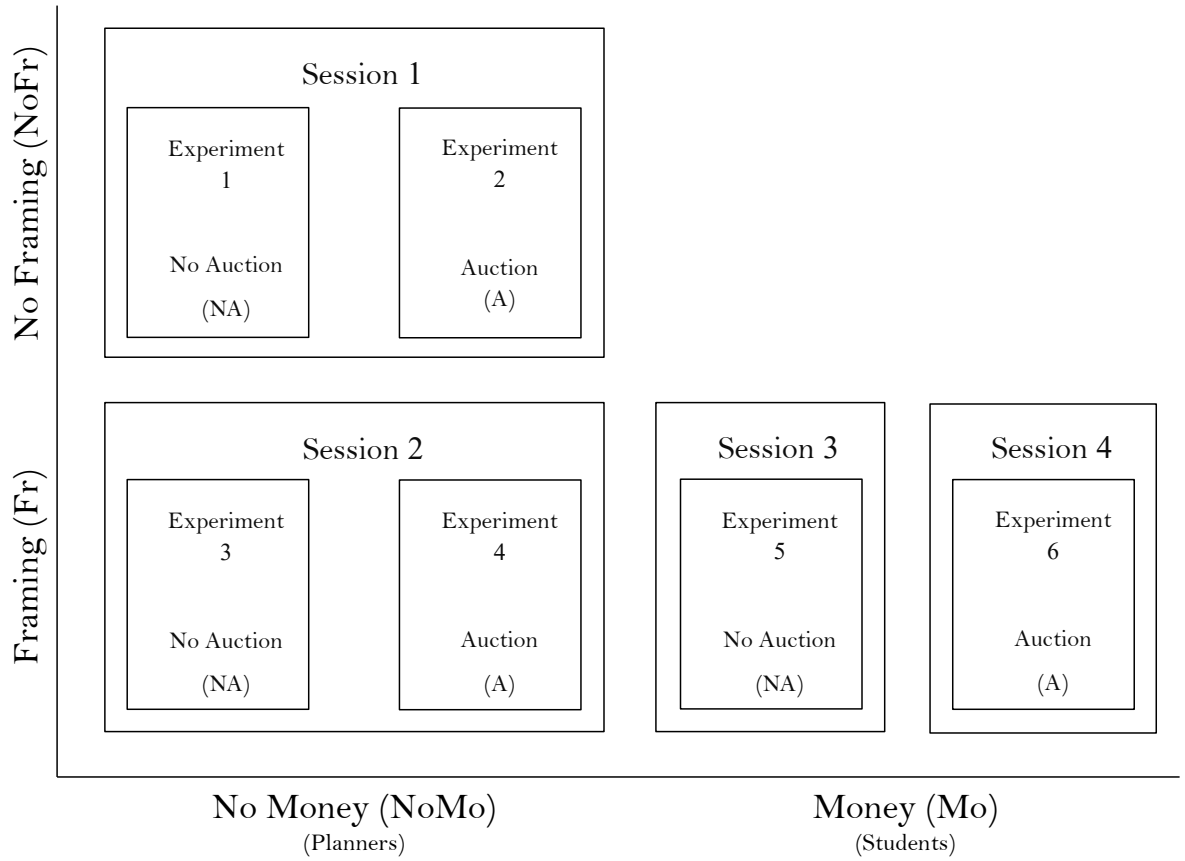


FIGURE 6.1: *Overview of the six experiments conducted in the context of the "Planspiel Flächenhandel".*

setting allows to overdraw accounts without restrictions (see Subsection 6.3.2 for further details).

Treatments vary in whether a continuous double auction (CDA; the NA treatment – NA for no auction) or a combination of a CDA and an uniformed sealed bid auction (USBA; the A treatment – A for auction) is used as market mechanism for trading TDRs (see Subsection 6.3.3 for further details on these mechanisms).

6.3.1 Participants

All 38 land construction planners were employees of the local authorities working in the public sector. They were recruited by the partners of the project "Planspiel Flächenhandel" through personal contact and promotion events during the course of the project. Their

participation was voluntary. Hence, no monetary compensation was given (NoMo). They participated from their office in their local authority's townhall. Access to the experimental platform was given through a web interface with a personal account. The platform was hosted on an online server at the FZI Research Center for Information Technology (FZI) in Karlsruhe (see Subsection 6.3.4 for further details about the experimental platform). Participants of the NoMo setting (land construction planners) were assigned to represent their own local authority and play four field experiment settings (Experiments 1, 2, 3, and 4 in Figure 6.1). The field experiments were designed as a within subject study at two separate days – Sessions 1 and 2 each with two experiments per day (see Subsection 6.3.2 for further details about the session structure). These four field experiments consisted of the two treatments (NA, A) once conducted with the conditions of the no framing setting (NoFr; Session 1 in Figure 6.1) and once with the conditions of the framing setting (Fr; Session 2 in Figure 6.1), resulting in four experiments at two separate session. Due to no shows of four land construction planners during the Fr setting the respective treatments (Session 2 – Experiments 3 and 4 – in Figure 6.1) consisted of only 34 participants.

For Sessions 3 and 4, all 76 students were recruited from a voluntary student pool from the KIT using ORSEE (Greiner, 2004). In contrast to the land use planners (NoMo), students received a monetary compensation (Mo). Students participated at the Karlsruhe Decision & Design Lab (KD^{2}Lab)[9]. Access was given to the same online platform through browsers in isolated cabins to assure comparability between treatments. Participants of the Mo setting (students) were each assigned to represent one of the local authorities, who participated in the NoMo settings, and play one of two lab experiments (Experiments 5 and 6 in Figure 6.1). The lab experiments were designed as between subject designs at two separate days, i.e., in each treatment there were 38 student participants (see Subsection 6.3.2 for further details about the session structure). Hence, the 76 students were assigned to either the NA (Experiment 5 in Figure Figure 6.1) or the A treatment (Experiment 6 in Figure 6.1), both with a restricted budget and a framing (Fr). However, students did not play the setting without restrictions (NoFr). Students were incentivized by a 5€ fixed payment for participation and a ranking-based bonus. The ranking was based on their last play money account balance, which depended on their decisions throughout the experiments. Bonus payment ranged from 10€ to 21€ with an average of 15.57€ plus

[9]For a detailed description of the KD^{2}Lab see `http://www.kd2lab.kit.edu/` last accessed: September 6, 2016.

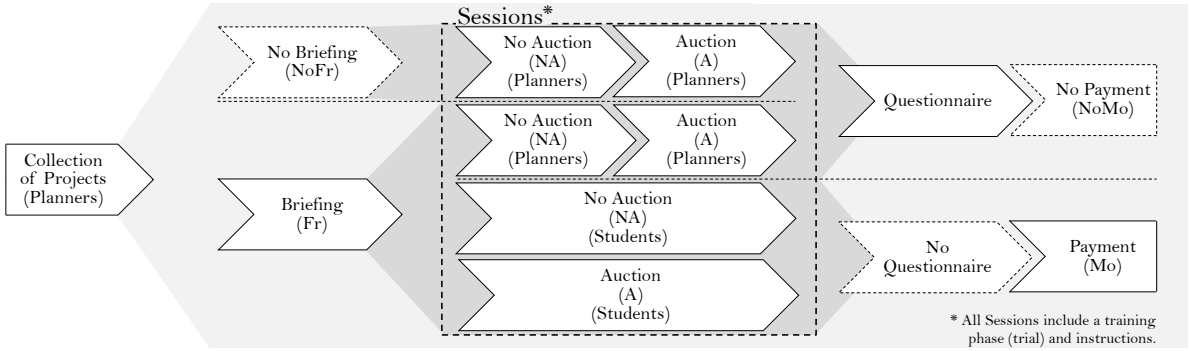


FIGURE 6.2: *Experimental processes of Chapter's 6 study.*

5€ fixed payment. Hence, payments were above an hourly wage of 10€ (see Table C.1 in the Appendix for a detailed overview of bonus payments).

6.3.2 Experimental Process

The six experiments were conducted in four sessions as depicted in Figure 6.1. Overall, the experiment design is divided in settings (NoFr, Fr, NoMo, Mo), treatments (NA, A), and experiments (1-6). Session 1 consisted of two field experiments with 38 land construction planners. Both experiments were conducted on one day – NA in the morning, A in the afternoon. No monetary compensation was given; both were part of the NoMo setting. Play money was not restricted or limited. Participants started with zero play money and could overdraw their account without restrictions. Participants were not framed (NoFr) before the experiments to behave in any certain way. The experiment was conducted remotely. Participants accessed the online experimental platform through a web interface from their own office. After an initial log in, participants were instructed to read the instructions and view a tutorial video of the NA treatment (Experiment 1). Afterwards, all participants played a 15 minute trial round to get familiar with the experimental software. Subsequently Experiment 1 (treatment NA) was conducted. After 75 minutes, participants took an one hour break. Before the second experiment started, participants had to view another tutorial video, explaining differences to the first experiment – the USBA extension (A). Experiment 2 (treatment A) took 90 minutes due to the extension of the USBA. After the second experiment, participants filled out a questionnaire asking for their motives and feedback. The overall experimental processes are depicted in Figure 6.2.

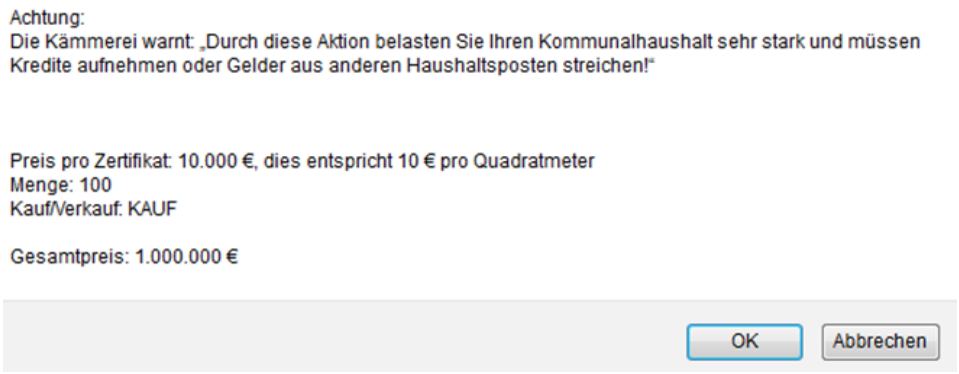


FIGURE 6.3: *Information message that was shown if a participant's action would spend more than 10 % of his yearly budget.*[10]

Overall, Session 2 was conducted in the same sequence as Session 1: instructions part 1, trial round, Experiment 3 (treatment NA), break, instructions part 2, Experiment 4 (treatment A), and a questionnaire (see Figure 6.2). The same 38 land construction planners as in Session 1 were invited to take part. However, due to four no shows only 34 land construction planners took part. Session 2 differs from Session 1 by a framing and a budget restriction (Fr). Two weeks before the experiment, all participants were called by phone and instructed to behave realistically, as they would do if the decisions they face during the experiments were real. Instructions followed a script to assure similar framings. Furthermore, the importance of the experiment within the overall project context was highlighted. The instructions were slightly changed, now including the importance of the project and the instructions to behave realistically. A participant's budget was restricted on a yearly basis. Each year a participant could spend 20 % of the local authority's yearly budget. Information messages were shown, once a participant spend more than 10 % of the local authority's budget as depicted in Figure 6.3. The message stated that the participant would overdraw the budget which the local authority provides for land

[10] Translation of Figure 6.3:

Attention:
The finance department warns: "With this decision you severely stress your local authority's account and have to raise credits or withdraw the money from different budget items!"

Price per certificate: 10,000€, which corresponds to 10€ per square meter
Amount: 100
Buy/Sell: BUY

Total price: 1,000,000€.

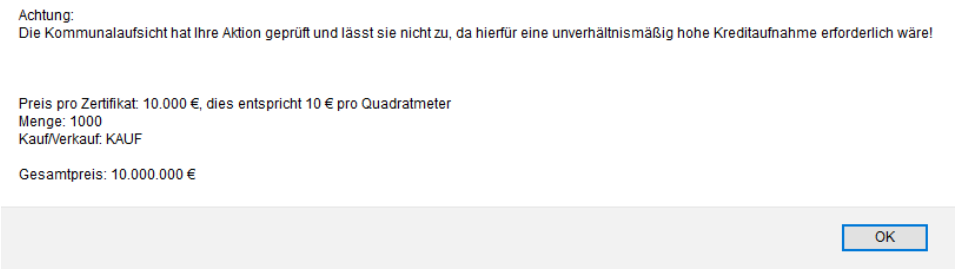


FIGURE 6.4: *Information message that was shown if a participant's action would spend more than 20 % of his yearly budget.*[11]

consumption and if the participant still wanted to proceed, money from other sectors of the local authority had to be used. Thereby the participants were reminded, that spending money has consequences for their local authority's community. If a participant's action would lead to a yearly spending over the 20 % budget limit, another information message was shown (Figure 6.4) informing the participant that it is forbidden to spend this amount in the current year.

Session 3 and Session 4 each consisted of one experiment (Experiments 5 and 6 in Figure 6.1). 76 students of the KIT took part, 38 students in each experiment. Both sessions included a framing and a restricted budget (Fr) (see Figure 6.2). The framing was similar to Session 2, but not done by phone. Instead participants were briefed directly before the experiment started. Participants were invited via ORSEE (Greiner, 2004). The experiment took place in the KD^{2}Lab. Participants were incentivized by a monetary rank-based bonus payment (Mo) as depicted in Table C.1 and a fixed payment of 5€. During the experiments participants could earn play money. Participants were ranked according to their last play money account in the respective experiment. The experiments (5 and 6) started with the briefing including the framing and the payout structure. After the initial

[11] Translation of Figure 6.4:

Attention:
The municipal regulation reviewed your decision and does not allow it, since disproportionate high credits would be needed!

Price per certificate: 10,000€, which corresponds to 10€ per square meter
Amount: 1000
Buy/Sell: BUY

Total price: 10,000,000€.

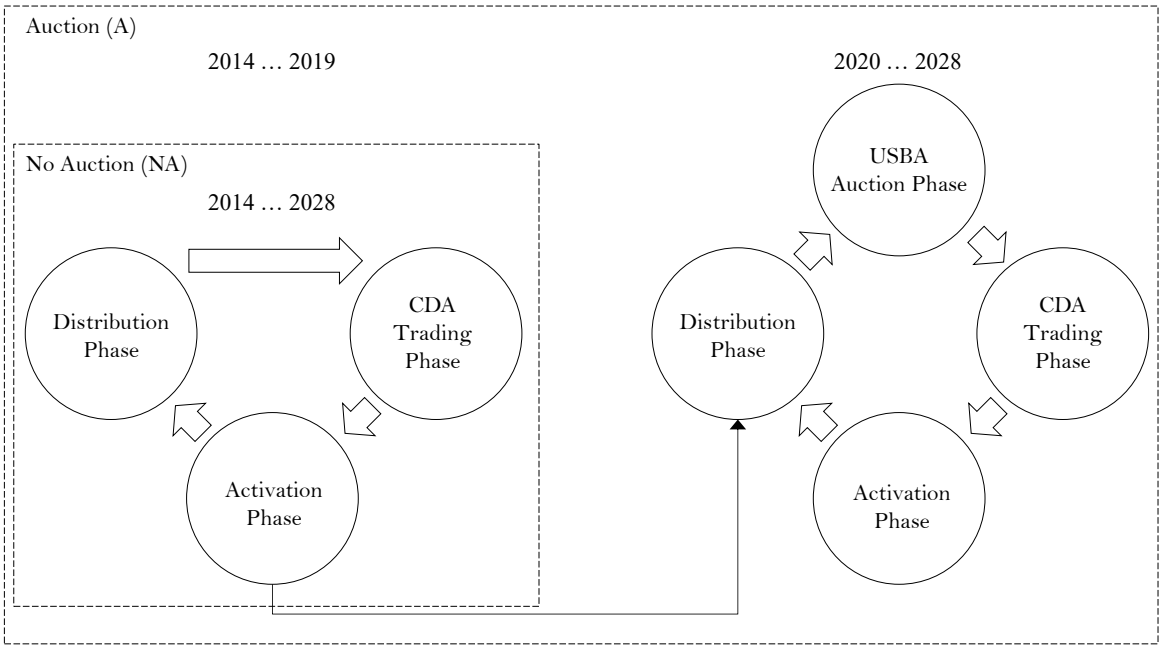


FIGURE 6.5: *Overview of the experimental phases per simulated year.*

briefing, participants were assigned to isolated cabins. Participants then had time to read the instructions, followed by a trial round of the experiment. Afterwards, they played the respective experiment – treatment NA or A depending on the session. After the experiment, participants got their payment and the respective session was finished (3 or 4).

6.3.3 Phases of the Experiments

The experimental processes vary in whether an USBA auction (A) takes place or not (NA), as shown in Figure 6.5. During the experiment 15 years (2014-2028) are simulated. The NA treatment consisted of three phases per year: distribution phase, CDA trading phase, and activation phase. The A treatment consists of an additional USBA auction phase, which is introduced in the year 2020. All phases in the respective experiments are repeated once per year, i.e., 15 times throughout one experiment.

The distribution phase is the first phase of each year. During the distribution phase each participant gets TDRs in the form of certificates for free. The distribution happens instantaneously during the change of the year. The amount of TDRs each participant gets for free depends on the represented local authority's population. Larger local authorities

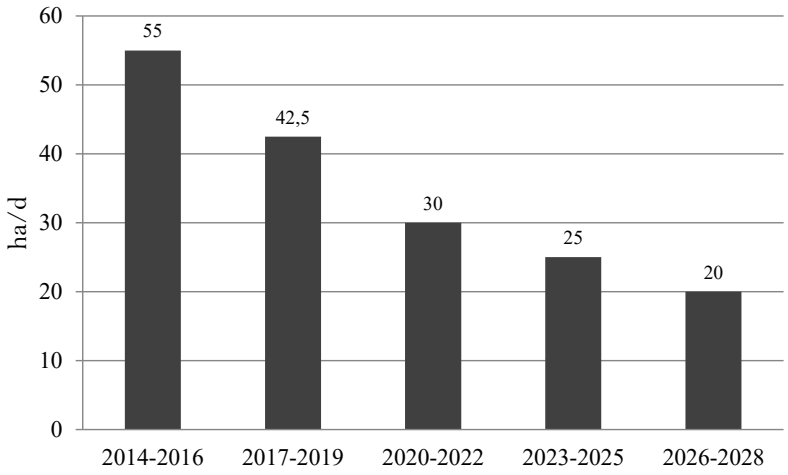


FIGURE 6.6: *Stepwise reduction of land consumption in hectare per day (ha/d).*

get more TDRs than smaller ones[12]. In total 8,467 TDRs are freely distributed throughout the experiment (average = 223; max = 885; min = 5). Overall, the project tests a reduction of land consumption from roughly 75 ha/d to 30 ha/d in the year 2020. Since this is a big gap, the reduction is introduced in five steps over time. In the years 2014-2016 TDRs which account for a new land consumption of 55 ha/d are distributed. Followed by 42.5 ha/d in 2017-2019, 30 ha/d in 2020-2022, 25 ha/d in 2023-2025, and 20 ha/d in 2026-2028, as depicted in Figure 6.6. In the NA treatment the TDRs are completely distributed. In the A treatment, beginning with the year 2020 certificates are partly retained to distribute them via an USBA.

Aktuelle Gebote

Kauf		Verkauf	
Preis	Menge	Preis	Menge
1.200 €	5	2.000 €	25
1.000 €	15	6.000 €	5

FIGURE 6.7: *Orderbook of the experimental platform.*[13]

[12] Details can be found in Henger and Schier (2014) and on the project homepage: `http://www.flaechenhandel.de/flaechenhandel/zertifikate-rechner` last accessed: September 6, 2016.

[13] Translation of Figure 6.7: Markt = market; Aktuelle Gebote = current orders; Preis = price; Menge = amount.

The second phase is the CDA trading phase. Participants can trade their TDRs with each other via a CDA. A CDA is a widely used market mechanism, e.g., in stock exchanges and prediction markets (Wolfers and Zitzewitz, 2004). Participants make buy or sell orders consisting of the amount of TDRs they want to buy or sell and a limit-price (per single TDR). Orders are collected in an orderbook, as depicted in Figure 6.7. In contrast to stock exchanges, no transaction costs for putting orders are present. Thereby the CDA is a zero sum game. Orders can be canceled by participants as long as the order is not executed. Orders will be executed automatically once prices from a sell order and a buy order match. TDRs and money are then exchanged between the respective participants' accounts – the buyer and the seller. The basic idea is that participants who do not need TDRs which they get for free in the distribution phase can sell TDRs to participants who need additional TDRs. For example if one participant relinquishes to realize a building project the participant can sell the TDR contingent needed for the realization. This phase lasts three minutes.

The third phase is the activation phase. During this phase a participant can activate building projects of the represented local authority. Each local authority has a list of projects it plans to hypothetically build until 2028 as depicted in Figure 6.8. Each local authority's projects were collected before the experiment via a project partner's survey census platform as depicted in Figure 6.9. Afterwards, projects were fiscally evaluated, taking, e.g., into account how many workplaces and living quarters they would create, how much taxes they would generate, and how much the land costs.

Dependent on the size of the project, more or less TDRs are needed for realization. Projects firstly can be realized starting at a certain year – their theoretical building year as depicted in the first column of Figure 6.8. Their realization can be shifted to later years as well. Hence, a participant can realize projects depending on the year and his available TDRs. If a participant realizes a project, the fiscally evaluated amount of the project is booked on the respective account.

Besides the realization of new projects, a different set of already existing projects can be depleted. These depletion projects cost money but generate new TDRs – so called "white TDRs". To not flood the market with TDRs and thereby prevent the goal of an overall reduction of new land consumption, the yearly maximal amount of claimable white TDRs is capped. The cap is depending on the size of the local authority. If a depletion project

Entwicklung

Beginn	Name	Typ	WE	AP	Dauer [a]	BBL [ha]	Fiskalwert [€]	Fiskalwert [€/Zert.]	Zertifi-kate	Entwicklung Ja/Nein
2014	Turmstrasse	G	0	15	2	3,0	500.000	-	0	☑
Gesamtbedarf bis 2014 / abzüglich der noch austehenden kostenfreien Erstzuteilungen:									**0 / 0**	
2015	Schillerstrasse	W	30	0	3	6,0	2.000.000	33.333	60	☐ Ab 2015
2015	Goethestrasse	W	40	0	2	8,0	4.000.000	50.000	80	☐ Ab 2015
2015	Poststrasse	M	20	20	3	4,0	1.000.000	25.000	40	☐ Ab 2015
Gesamtbedarf bis 2015 / abzüglich der noch austehenden kostenfreien Erstzuteilungen:									**180 / 0**	
2016	Schlossallee	W	100	0	1	14,0	10.000.000	71.429	140	☐ Ab 2016
2016	Rathausplatz	M	50	50	5	10,0	6.000.000	60.000	100	☐ Ab 2016
2016	Bahnhofstrasse	G	0	60	4	12,0	8.000.000	66.667	120	☐ Ab 2016
Gesamtbedarf bis 2016 / abzüglich der noch austehenden kostenfreien Erstzuteilungen:									**540 / 0**	

Gesamtkosten: 0 Zertifikate
Gesamterlös: 500.000 €.

Entwicklung bestätigen

FIGURE 6.8: *Example project list of a local authority.*[14]

would claim more white TDRs than the yearly cap, the TDR payout is stretched over several years. The implementation of depletion projects and the possibility to claim white TDRs is used as an incentive to over-think the use of fallow building cites. Participants can use generated white TDRs for realizing other projects or sell them in the trading phases. An example list of depletion projects is shown in Figure 6.10. Both, building and depletion projects, can be executed during the activation phase. It runs two minutes per year.

The fourth phase is the USBA auction phase, which only exists in the A treatments. Beginning with the year 2020, TDRs, which were freely distributed in the distribution

[14]Translation of Figure 6.8: Beginn = start; Name = name; Typ = type; WE = living quarters; AP = workplaces; Dauer [a] = duration [years]; BBL [ha] = gross development area [ha]; Fiskalwert [€] = fiscally evaluation [€]; Fiskalwert [€/Zert.] = fiscally evaluation [€/TDR]; Zertifikate = TDRs; Entwicklung Ja/Nein = realization yes/no; Gesamtbedarf bis 2014 / abzüglich der noch ausstehenden kostenfreien Erstzuteilungen = total demand 2014 / deducting the outstanding free of charge TDRs of the distribution phases; Gesamtkosten = total costs; Gesamterlös = total income; Entwicklung bestätigen = confirm realization.

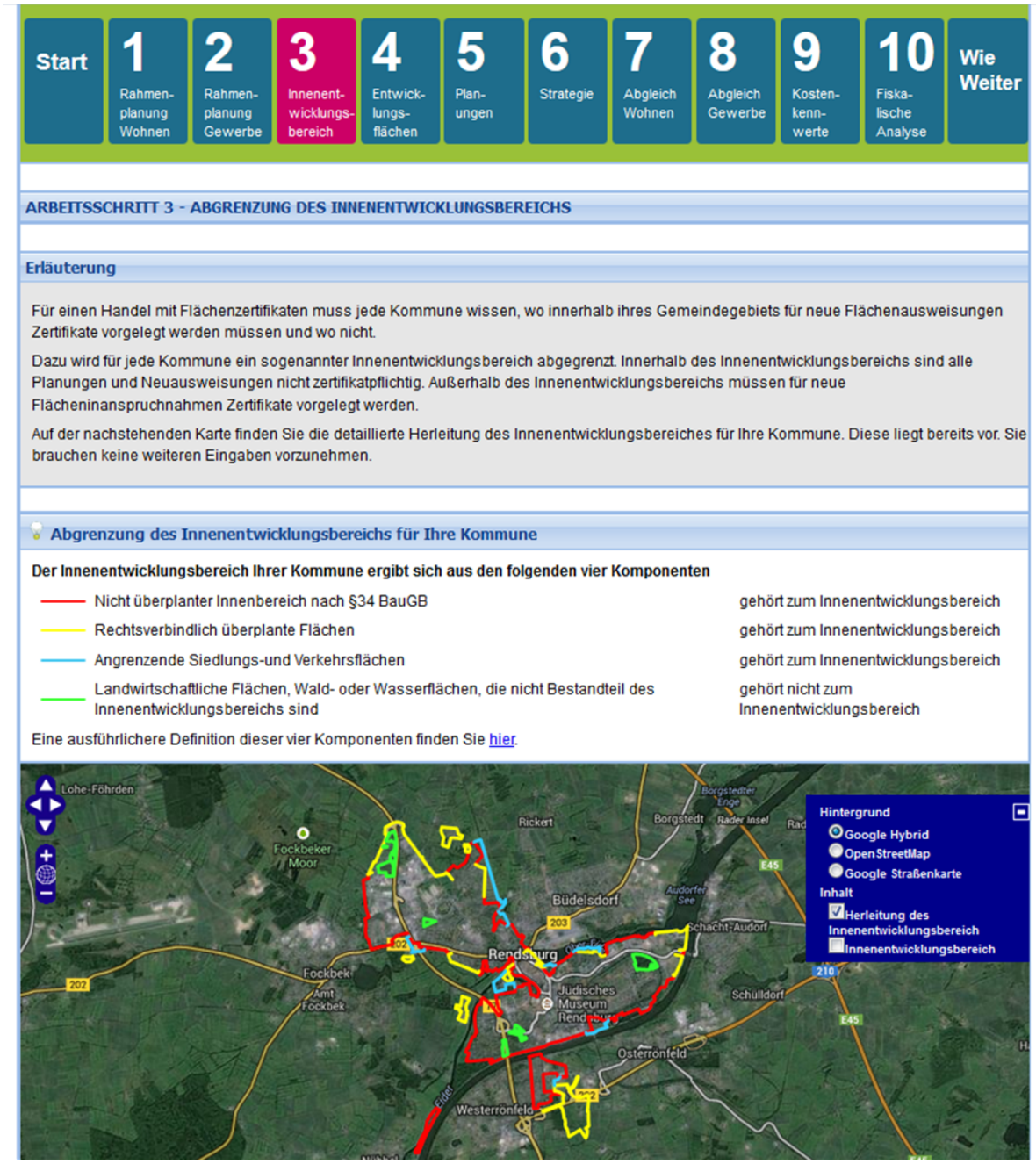


FIGURE 6.9: *User interface of the project census platform by Gertz Gutsche Rümenapp GbR.*

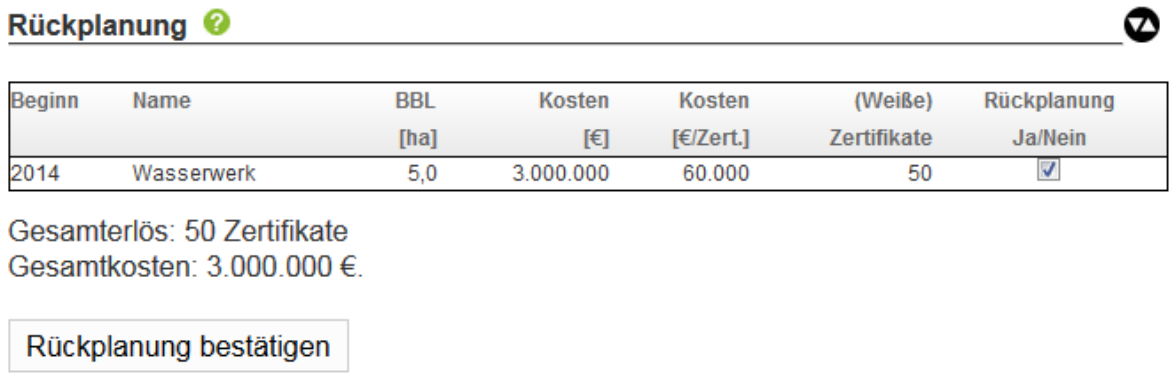


FIGURE 6.10: *Example depletion project list of a local authority.*[15]

phase of the NA treatments, are partly retained to auction them via an USBA in the A treatments. The rest of the TDRs are still distributed freely during the distribution phase. To get used to this new mechanism the amount of auctioned TDRs is increased stepwise over years. In the years 2020-2022, 20 % of the overall TDRs are auctioned, followed by 40 % in 2023-2025 and 60 % in 2026-2028. The phase is active for one and a half minutes directly after the distribution phase. Participants can bid on the auctioned TDRs by handing in the amount of TDRs they want to auction and a limit-price they want to pay per TDR. Handing in several bids is allowed. Participants see their own bids, but bids from other participants are hidden. Own bids can be canceled during the phase and the the overall amount of auctioned certificates is available as an information message in the header. At the end of the phase bids are collected and are served from the highest to the lowest price. This means that a participant who is willing to pay the highest price for TDRs is served first. Afterwards the participant with the second highest price is served, and so on. Participants are served as long as TDRs are available. All participant pay a unified price for the auctioned TDRs. The unified price is the limit-price of the last served bid, i.e., the lowest limit-price of all fully or partly executed bids. Hence, all participants who are being served pay this same price multiplied with the amount of TDRs they auctioned during this phase. All other bids do not get served. The money which is spend during this phase is not re-distributed to the participants. It is kept as hypothetical governmental tax.

[15] Translation of Figure 6.10: Beginn = start; Name = name; BBL [ha] = gross development area [ha]; Kosten [€] = costs [€]; Kosten [€/Zert.] = costs [€/TDR]; (Weiße) Zertifikate = (white) TDRs; Rückplanung Ja/Nein = depletion yes/no; Gesamterlös = total income; Gesamtkosten = total costs; Rückplanung bestätigen = confirm depletion.

6.3.4 Experimental Platform

A customized experimental platform was programmed to implement the described study design. It was implemented using the grails framework[16], following a model view controller approach. An overview of the platform is given in Figures 6.11 and 6.12. The platform consists of features, supporting participants' decisions by potentially making land construction planning easier. The platform is divided into several parts. On the left hand side, bids for the trading and auction phase can be made if the respective phase is active (Figure 6.11: number 3). Before handing in an order a message is shown, reminding participants about how much they would spend in total. Only by accepting this billing orders are confirmed and executed. The list of all building projects (Figure 6.11: number 4; Figure 6.12: number 4) and depletion projects (Figure 6.11: number 5; Figure 6.12: number 5) is always shown, to give an overview of how many TDRs are needed. Building projects are standardly shown, if the activation phase is active and building projects are currently realizable (Figure 6.12: number 4). Building projects are displayed slightly grayed out if other phases are currently active (Figure 6.11: number 4). This feature was implemented to highlight parts of the platform, which are important in each phase, while still having all the information displayed at any given time throughout the experiment.

On the right side of the platform overall goals are shown (Figure 6.11: number 6; Figure 6.12: number 7). In reality, the realization of building projects would create working or living places. The goals show how many of these are currently created through realized projects and how many overall could be created, if all potential building projects would be realized. Below, the participant's deposit is shown (Figure 6.11: number 6; Figure 6.12: number 7). Besides a participant's current money and TDR deposit, potential changes through actions are shown. An additional column shows how orders or planned building projects would affect the deposit. The market information on the right side shows the orderbook (best five buy and sell orders), own orders (not yet allocated orders, which can be deleted), and a price chart (price per TDR and volume) (Figure 6.11: number 7; Figure 6.12: number 8). As for the case of building and depletion projects, this information is displayed slightly grayed out if the CDA trading phase is inactive (Figure 6.12: number 8). For the USBA auction phase the own auction bids are shown as well (Figure 6.11: number 8; Figure 6.12: number 9).

[16] `https://grails.org/` last accessed: September 6, 2016.

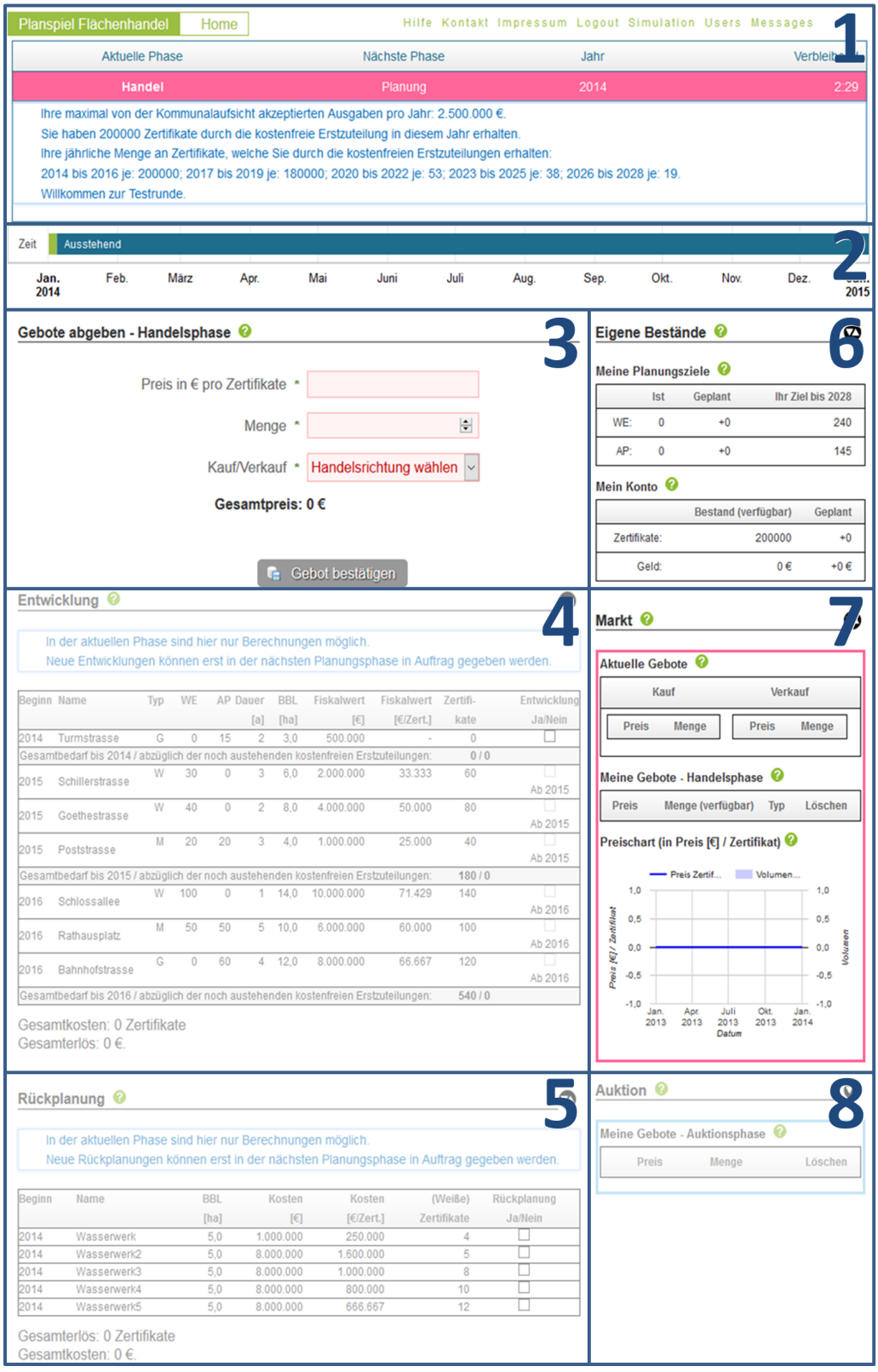


FIGURE 6.11: *User Interface of the experimental platform: Active CDA trading phase.*

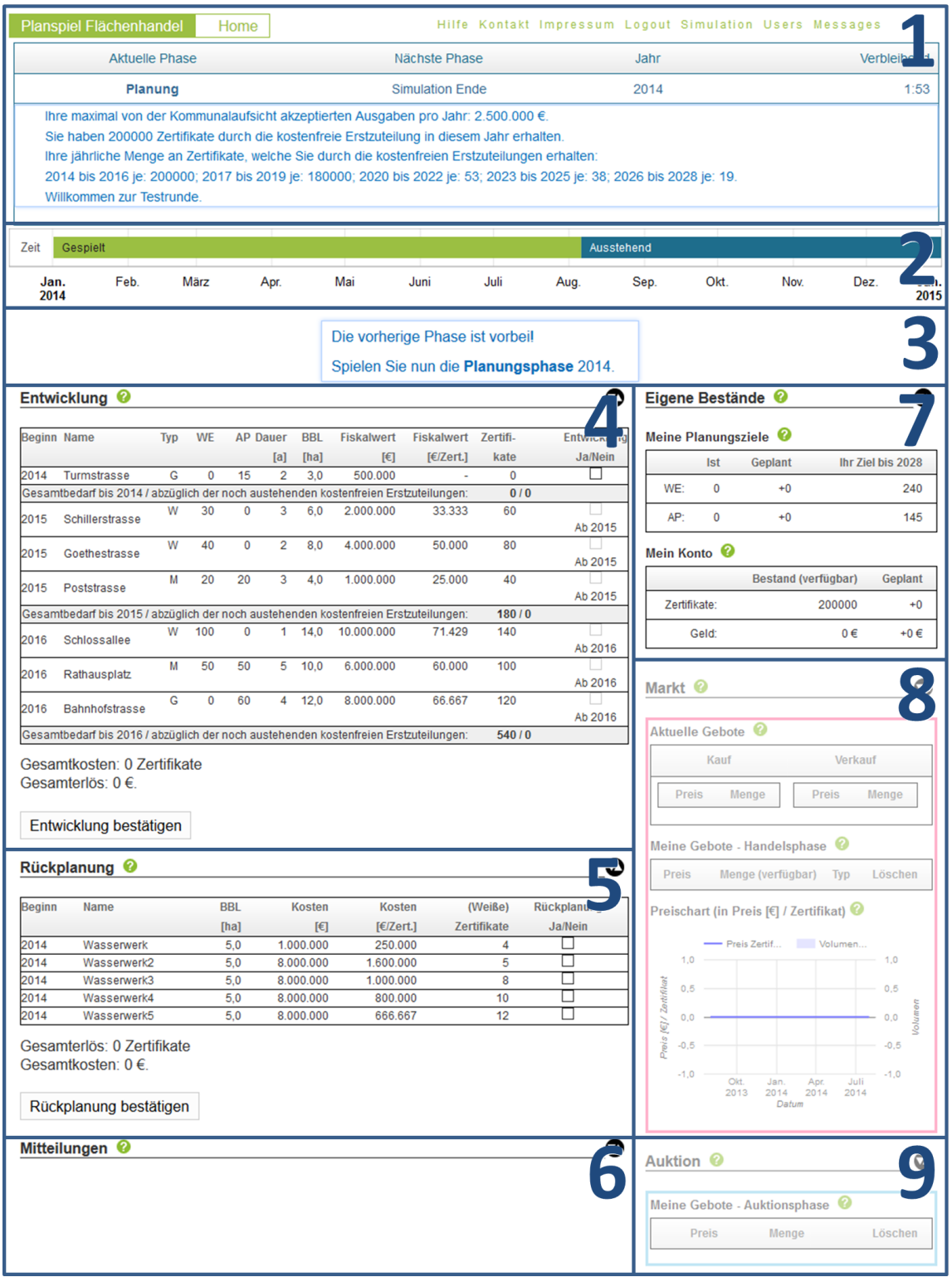


FIGURE 6.12: *User Interface of the experimental platform: Active activation phase.*

On top of the platform messages are shown, which inform participants about changes and events during the experiment, e.g., how many TDRs they auctioned during the last USBA auction phase, how many TDRs they will get over the course of the experiment during the several distribution phases, or that an auction will take place in the next year and how many TDRs will be auctioned (Figure 6.11: number 1; Figure 6.12: number 1). Besides these short informations, participants can see the messages in more detail at the bottom of the page (Figure 6.12: number 6). Furthermore, an overview about the progress of the experiment is shown. A time progress bar shows how much time has already progressed and how much time of the experiment is left (Figure 6.11: number 2; Figure 6.12: number 2). Additionally, the top of the platform informs participants about the currently active phase, which will be the next phase, which year is currently simulated, and how much time is left in the current phase (Figure 6.11: number 1; Figure 6.12: number 1). Phases are color-coded. Pink for the CDA trading phase, white for the activation phase, and blue for the USBA auction phase. Instructions can be accessed at any given time during the experiment by a navigation on top of the page, if participants still had open questions (see Appendix C for further details on the instructions).

6.4 Evaluation

Overall, 114 participants took part in the study – land construction planners (12 female, 26 male)[17] and 76 student participants (19 female, 57 male). The evaluation starts with the differences from the NoFr to the Fr setting, followed by the differences between NoMo and the Mo setting. The differences are evaluated by comparing prices of TDRs, spending of participants, and the conversion rate of projects. In statistical tests, a 0.1 level is employed to decide on the rejection of null hypotheses. More detailed information on p-values is provided in the result Subsections (6.4.1 and 6.4.2) of this chapter.

[17]Note that due to no shows only 34 land construction planners took part in Experiments 3 and 4. 38 took part in Experiments 1 and 2 (see Figure 6.1)

6.4.1 No Framing vs. Framing

To asses Research Question 4 and hence the comparison between the NoFr and the Fr settings, the results of Sessions 1 and 2 are compared (Figure 6.1). In both sessions the behavior of land construction planners is examined. In each session they took part in the NA and A treatments, once without a framing and restrictions (NoFr) and once with a framing and a yearly restricted budget (Fr).

The relation of the two treatments is evaluated in three ways: First, it is evaluated if the prices of the CDA trading phase and the USBA auction phase move in a realistic scope given the German land prices. Second, it is evaluated if the spending of participants behaves similarly. Third, it is evaluated if the conversion rate of projects is different in the two respective settings.

Calculating an exact theoretical average price for a TDR is difficult. It depends on many factors, like which local authority sells a TDR at which point to whom, how many TDRs a local authority banks in his depot to sell later on, and most importantly, how differently local authorities evaluate land prices. However, several land consumption experts and consultant institutions were involved in the project's context. Together they evaluated land prices of the local authorities taking part in the experiments. While it is hard to give an exact average price per TDR, they concluded for the experimental setup that a TDR should not cost more than 200,000€. This price is a bit higher than the average square meter price in Germany. As a reduction in land consumption would lead to a shortage, it is expected to raise prices. Thereby, the first part of the evaluation observes if this goal is achieved.

Session 1 tested the NoFr setting, including two treatments (NA, A) and Session 2 tested the Fr setting[18], including the same two treatments (NA, A). Price statistics are shown in Table 6.1.

The experimental economics literature (Smith, 1976) suggests that, experiments without incentives may not lead to realistic behavior and hence unrealistic results. This is reflected by the high prices in both treatments of the NoFr setting. Especially in the A treatment of the NoFr setting it becomes clear, that prices are far off the 200,000€ upper

[18]Note that due to unrealistically high prices, the A treatment of Session 2 (Experiment 2) was aborted at the year 2020.

TDR price boundary (see Table 6.1). Without incentives, which were linked to their actions and decisions throughout the experiment, participants tested the boundaries of the experiments. They spend unrealistically high prices per TDR, since they had an unlimited budget and were not incentivized to take the money in their depot into account for their decisions. Overall this led to the result that Experiment 2 (NoFr - A) was aborted at the year 2020, once the first USBA auction phase showed, that prices would not relax from this unrealistic level.

These unrealistic price level, however, did not occur in the Fr settings. In both cases prices were under 200,000€. This leads to the conclusion, that this customized framing, reminding participants to behave realistically, had an effect on the participants' behavior towards more realistic decisions.

Besides this framing, a yearly budget was as well implemented in the Fr settings. However, this did not affect the price results, as only one participant in one year in the NA treatment (participant 24 in the top image of Figure 6.13) and one participants in four years in the A treatment (participant 33 in the bottom image of Figure 6.13) came close to their yearly budget limit. The rest of the participants earned money in each year or did not spend more than 75 % of their yearly budget (relates to 20 % of the local authorities yearly budget) as depicted in Figure 6.13.

Overall this shows, that after participants are framed to behave realistically, they do pay prices for TDRs experts and consultants would expect. Hence, the following result can be concluded:

> **Result 1:** *In the no framing setting unrealistically high prices were paid. The assessment of a framing lead to prices which experts and consultants in this area expected.*

Setting	Treatment	Experiment	Participants	Minimum	Median	Mean	Maximum
NoFr	NA	1	Planners	25,000	15,000,000	12,890,000	28,000,000
NoFr	A	2	Planners	1,000,000	500,000,000	497,449,000	999,999,990[19]
Fr	NA	3	Planners	40,000	138,000	133,500	200,000
Fr	A	4	Planners	25,000	100,000	92,870	122,000

TABLE 6.1: *Prices in €per TDR of Experiments 1-4.*

[19]Note that the technical maximum was set to 999,999,990.

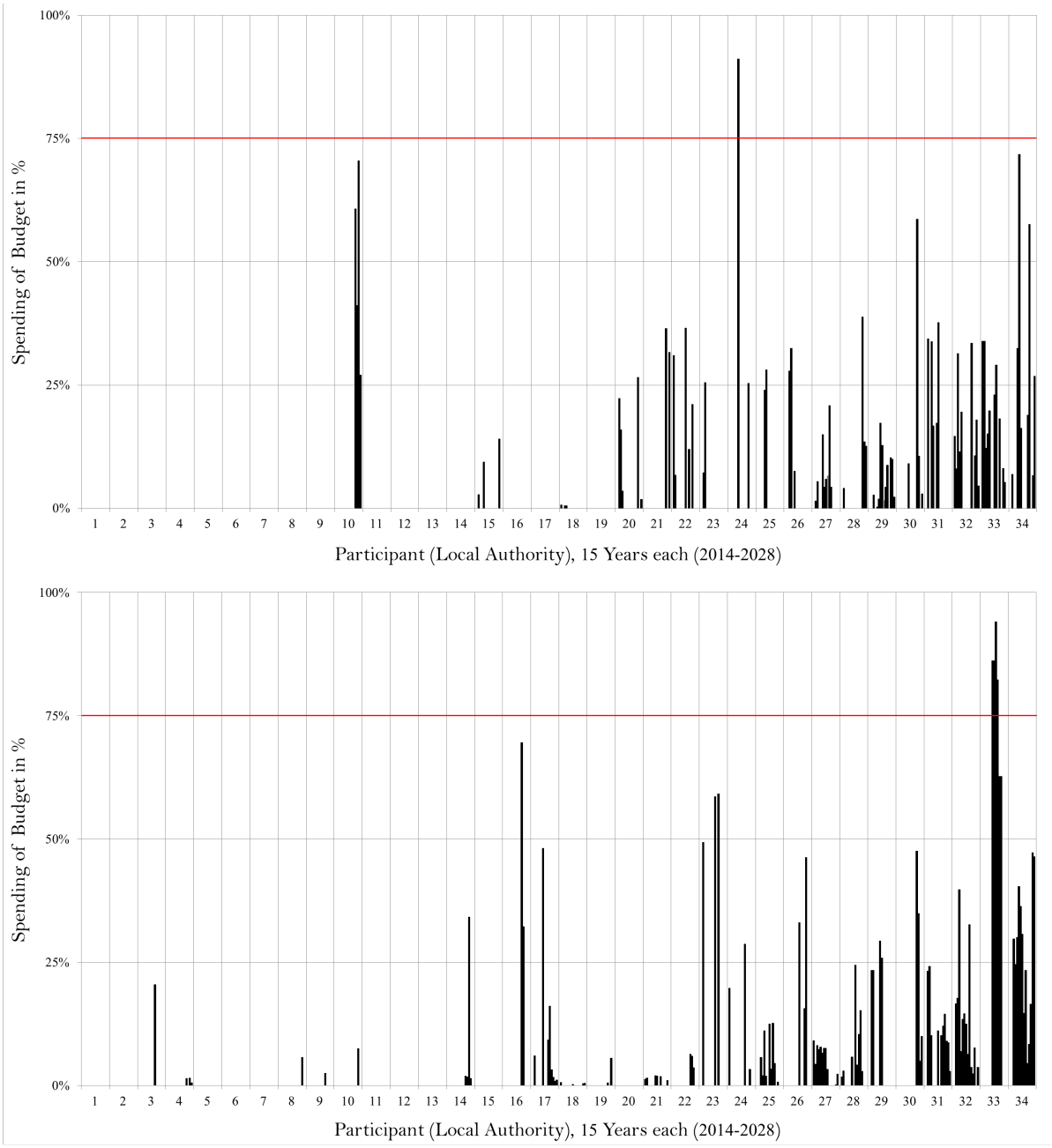


FIGURE 6.13: *Participants' spending of the local authorities' budget for land consumption during the experiments (Fr): NA treatment (top image), A treatment (bottom image). Only spendings are shown on the graph. Local authorities who made money are depicted as 0 % spendings of their yearly budget. The graph is sorted by participants, from lowest to highest average spendings. One x-Axis index consists of 15 yearly spendings (2014-2018) of one participant.*

After looking at the TDR prices the evaluation of the money local authorities spend or made throughout the experiment is now assessed. This gives insight if they managed their resources somehow realistically. The mean last deposit values by experiments are shown in Table 6.2.

Due to the abortion of Experiment 2 (NoFr - A), a comparison of the local authorities' spending is only possible between Experiment 1 (NoFr - NA) and Experiment 3 (Fr - NA). Furthermore, the experimental design lacks of repeated observations of each experiment due to the field experiment character with 38 land construction planners. To not stress their voluntary participation, project partners appointed four experiments at two different days (two experiments per day; see Figure 6.1). This made repeated observations of the same experiment setup (setting and treatment) with same or different participants impossible. Thereby no inference based statistics can be applied on an experiment setup level (setting and treatment). As an approximation an assessment of analyses on a participant level is possible. Hence, a pairwise comparison of the local authorities' behavior between experiments is assessed. However, four participants did not take part in Experiments 3 and 4, thereby this pairwise comparison is only possible with 34 observations.

In the case of the local authorities' last depot values, tests comparing the means and the variance (e.g., t-test) are not suitable. The experimental design of the NA treatments allowed changes in the depot values by building or deploying projects and by trading TDRs during the CDA trading phase. From design the CDA trading phase is a zero sum game. Money one local authority earns by selling TDRs is spent by another who buys the TDRs. Hence, means equal zero and all statistical tests conclude in no differences between treatments, even if variance and overall spending tend to differ. For the A treatments this is different, since spending during the USBA auction phase are kept as a hypothetical governmental tax. As already explained, Experiment 2 was aborted and comparison is only possible for the NA treatments (NoFr – Fr). Hence, the correct way to use inference statistics on a participant level is thereby a test which does not compare means. Following

Setting	Treatment	Experiment	Participants	Minimum	Median	Mean	Maximum
NoFr	NA	1	Planners	-9,992,375,000	57,728,750	28,486,710	6,259,660,000
NoFr	A	2	Planners	-691,546,403,260	3,290,281,990	-9,580,158,760	182,616,536,800
Fr	NA	3	Planners	-11,883,995	2,449,498	34,889,768	440,641,000
Fr	A	4	Planners	-13,194,045	1,058,951	32,627,956	446,016,050

TABLE 6.2: *Last mean depot values in € of Experiments 1-4.*

this a kolmogorov-smirnov test is assessed to test whether the two experiments follow the same distribution.

Results from the kolmogorov-smirnov test show that the two distributions are different. This leads to the conclusion that participants in fact behave differently in the two treatments (NoFr NA – Fr NA) (two-sided kolmogorov-smirnov test: p-value = .013). Descriptive statistics, depicted in Table 6.2 show, that overall in the framing setting participants spend less money.

> **Result 2:** *In the framing setting participants spend less compared to the no framing setting.*

The last part of the evaluation compares the realization of building projects on a participant level. Table 6.3 shows summary statistics about the conversion of building projects, divided by positive and negative fiscally evaluation.

As already explained, only the NA treatments (NoFr – Fr) are being compared. A paired wilcoxon signed-rank test is assessed, since the same participants competed in both experiments (NoFr, NA – Fr, NA; within subject design). Results show that the total amount of realized projects differs in the two experiments (matched-pairs wilcoxon signed-rank test: p-value = .041). Table 6.3 shows that overall more projects were realized in the Fr setting (Total: NoFr = 73.35 %, Fr = 77.45 %). Comparing positive and negative fiscally evaluated building projects results show that only more positive fiscally evaluated building projects were realized in the Fr setting (matched-pairs wilcoxon signed-rank test: p-value = .027; Positive: NoFr = 82.85 %, Fr = 89.41 %). Negative fiscally evaluated building projects do not differ between experiments (matched-pairs wilcoxon signed-rank test: p-value = .86; Negative: NoFr = 56.57 %, Fr = 56.11 %). This leads to the conclusion that the framing

Setting	Treatment	Experiment	Participants	Building Projects			Depletion Projects
				Total	Positive	Negative	Total
NoFr	NA	1	Planners	73.35 %	82.85 %	56.57 %	95.00 %
Fr	NA	3	Planners	77.45 %	89.41 %	56.11 %	95.12 %
N				501	321	180	41

TABLE 6.3: *Conversion rate of building and depletion projects in Experiments 1 and 3.*

led to a different appreciation of the fiscally evaluation leading to a higher share of positive fiscally evaluated projects.

> **Result 3:** *In the framing setting participants realized more positive fiscally evaluated projects.*

In conclusion, with the results of the three evaluations, a framing did work in this setting. Results show that in all three evaluations participants behaved in the directions as they were framed. Hence, regarding Research Question 4 (see Section 1.2 of Chapter 1), it can be concluded that a framing is capable of influencing participants' behavior towards a certain goal. How far this behavior is off from actual monetarily incentivized experiments will be part of the next section.

6.4.2 No Monetary Compensation vs. Monetary Compensation

To assess Research Question 5 (see Section 1.2 of Chapter 1) and hence the comparison of not incentivized planners from local authorities and monetarily incentivized students (NoMo – Mo), comparisons between Experiments 3 and 5 and Experiments 4 and 5 are conducted (see Figure 6.1). In Session 2 an observation of the behavior of land construction planners is conducted, while in Sessions 3 and 4 students took part. All Sessions included a framing but Sessions 3 and 4 were monetarily incentivized. In Session 3, 8 female and 30 male participants took part, in Session 4, 11 female and 27 male participants took part. Overall 29 of 76 student participants studied in fields with an economic background (25 in Session 3, 14 in Session 4). The evaluation is assessed in the same three ways as for the comparison between Sessions 1 and 2.

First, it is evaluated if average prices of TDRs during the experiments cost less than 200,000€– the scope set by experts. Averages prices of the different experiments are depicted in Table 6.4.

By comparing the results of the NoMo and Mo Settings it is shown that all prices move in a scope under 200,000€, besides some outliers in Experiment 6. Hence, the conducted framing for two experiments in the NoMo setting achieved the goal of realistic prices as well as the standardly incentivized monetary lab experiments in the Mo setting.

Setting	Treatment	Experiment	Participants	Minimum	Median	Mean	Maximum
NoMo	NA	3	Planners	40,000	138,000	133,500	200,000
NoMo	A	4	Planners	25,000	100,000	92,870	122,000
Mo	NA	5	Students	1,000	49,000	65,400	130,000
Mo	A	6	Students	10,000	69,000	65,730	300,000

TABLE 6.4: *Prices in €per TDR of Experiments 3-6.*

Result 4: *In both settings, with and without monetary incentives and a framing, prices for TDRs were realized which experts and consultants in this area expected.*

The second part of the evaluation compares how much money is spend by participants throughout the experiments. Table 6.5 shows summary statistics of participants' last mean deposit values.

Like for the comparison of the NoFr and Fr settings the evaluation is assessed on a participant level[20]. Statistical tests based on the mean and distribution are not suitable for the same reasons. Hence, a kolmogorov-smirnov test is assessed. To compare the behavioral differences between the NoMo and Mo settings, Experiments 3 and 5 (NoMo, NA – Mo, NA) and Experiments 4 and 6 (NoMo, A – Mo, A) are each compared pairwise. Otherwise effects of the USBA auction phase would interfere with the differences of the monetary incentive. Results indicate no significant differences for neither the NA (two-sided kolmogorov-smirnov test: p-value = .307) nor the A treatments (two-sided kolmogorov-smirnov test: p-value = .673). This leads to the conclusion that the framing worked regarding the spendings of participants. Planners from local authorities, who did not have a monetary compensation but were framed to behave realistically, behaved similarly to student participants who were incentivized with money in a laboratory experiment setting.

Setting	Treatment	Experiment	Participants	Minimum	Median	Mean	Maximum
NoMo	NA	3	Planners	-11,883,995	2,449,498	34,889,768	440,641,000
NoMo	A	4	Planners	-13,194,045	1,058,951	32,627,956	446,016,050
Mo	NA	5	Students	-19,570,084	7,899,883	23,712,897	240,824,923
Mo	A	6	Students	-15,629,668	6,712,975	30,087,269	449,288,985

TABLE 6.5: *Last mean depot values €of Experiments 3-6.*

[20]Note that 34 observations in each experiment are compared. Since four land construction planners did not show up for the Experiments 3 and 4.

Result 5: *No differences can be observed for the spending of participants in the no money and money settings.*

The last part of the evaluation of the NoMo and Mo comparison again looks at the realization of building projects. Table 6.6 shows summary statistics about the conversion of building projects.

Similar to the evaluation of the participants' spendings Experiments 3 and 5 (NoMo, NA – Mo, NA) and Experiments 4 and 6 (NoMo, A – Mo, A) are compared to assess the evaluation of differences in the behavior between the NoMo and Mo settings. Since different participants are compared (planners – students) an unpaired wilcoxon signed-rank test is used. Results show again no differences in the behavior of the NoMo and Mo settings. The total conversion rate shows no differences for Experiments 3 and 5 (wilcoxon signed-rank test: p-value = .985) and Experiments 4 and 6 (wilcoxon signed-rank test: p-value = .862). Splitting the evaluation by the fiscal evaluation does not show differences as well for Experiments 3 and 5 (wilcoxon signed-rank test: positive – p-value = .932; negative – p-value = .796) and Experiments 4 and 6 (wilcoxon signed-rank test: positive – p-value = .895; negative – p-value = .963). This leads to the conclusion that the framing worked regarding the conversion rate of projects. Framed participants in the NoMo setting behaved similarly to monetarily incentivized student participants.

Result 6: *No differences can be observed for the conversion rate of building projects in the no money and money settings.*

The three evaluations comparing the NoMo and Mo settings overall show, that in this study framing in fact achieved similar behavior of non-monetarily incentivized participants (NoMo) and monetarily incentivized participants (Mo).

Setting	Treatment	Experiment	Participants	Building Projects			Depletion Projects
				Total	Positive	Negative	Total
NoMo	NA	3	Planners	77.45 %	89.41 %	56.11 %	95.12 %
NoMo	A	4	Planners	76.85 %	90.03 %	53.33 %	92.68 %
Mo	NA	5	Students	75.05 %	84.42 %	58.33 %	95.12 %
Mo	A	6	Students	72.85 %	86.29 %	48.89 %	78.05 %
N				501	321	180	41

TABLE 6.6: *Conversion rate of building and depletion projects in Experiments 3-6.*

6.4.3 Summary of Evaluation

In this evaluation two comparisons were assessed. Overall results are depicted in Table 6.7. The first comparison of a non-framed (NoFr) and a framed (Fr) setting showed that for this experiment setup participants' behavior is affected by the assessed framing.

The second part of the evaluation compared the results of the framing and no monetary incentive setting (Fr, NoMo) with the behavior of a classically incentivized student laboratory setup (Fr, Mo). Results indicate that non incentivized field participants who have been framed to behave realistically (NoMo) acted similar during the experiment as the student participants with a monetary compensation (Mo).

In conclusion, for the given scenario of land construction planning with a cap and trade system using TDRs, the introduction of a customized framing achieved comparable behavior to a controlled laboratory environment with monetary compensation in a crowd scenario without monetary compensation.

6.5 Summary

Non-monetary incentives are an important part of the motivation to participate on crowdsourcing platforms. This chapter investigated if a framing, reminding participants of the importance of the context and stating that participants should behave realistically (Fr, NoMo), actually incentivizes participants similar to a monetary incentive (FrMo). For the case of a cap and trade system for TDRs to reduce the overall land consumption, a customized framing affects participants' behavior (RQ 4). Moreover, this behavior is similar to the behavior of monetarily incentivized participants (RQ 5). These results show that a framing aligned with the project context can change the behavior of participants. Thereby

	Level of Analysis		
	TDR Price < 200,000€	Local Authority's Spending	Building Projects
NoFr vs. Fr	$\neq$	$\neq$	$\neq$
NoMo vs. Mo	$=$	$=$	$=$

TABLE 6.7: *Differences in settings of the experiments.*

platform designers of crowdsourcing platforms who are not able to pay their crowd workers, have a strong tool to incentivize and motivate their crowd workers.

The main contribution of this chapter is to give insight in the practitioners question of how to incentivize crowd workers on crowdsourcing platforms. This chapter shows, that a framing aligned to the context of the platform is a valuable addition to monetary incentives. The limitations of this study include the following: First, this study evaluates a specific case – a cap and trade system for TDRs to reduce the overall land consumption. In different settings the findings might not be transferable. Second, this study evaluates the behavior of land construction planners, which participated in this study voluntarily. It might be that the framing in such a setup works better, since the initial motivation of participants was intrinsically driven. In setups were external motivation is standard, findings could be different. Third, due to the project context and the field experiment character, the study lacks of repeated observations of the experiments. Thereby, the evaluation of the price levels is not assessable with inference statistics. Fourth, due to the field experiment character the experiment lacks of internal validity.

Future research should replicate the findings in a laboratory study with a higher focus on internal validity and repeated observations to strengthen the findings. The study should be transfered to a setup where extrinsic motivation is key and evaluate if a framing still affects a crowd workers behavior. Furthermore, it would be fruitful to evaluate different levels of framing, to see how the sentiment of the framing affects participants behavior differently.

To conclude, non-monetary crowdsourcing plays an important role in the context of online work. As this study shows, future work should not only focus on paid crowdsourcing but as well on intrinsic motivational factors. Platform designers can profit from these insights in better understanding how to setup their non-paid and paid crowdsourcing platforms.

To sum up, this chapter enhanced the picture of incentives for crowdsourcing platforms by taking non-monetary incentives into account. It shows, that framing crowd workers can improve the outcome of crowdsourcing platforms. Overall, this thesis thereby evaluated a broad picture of both monetary (see Chapter 4 and 5) and non-monetary incentives (see Chapter 6).

Part IV

Finale

Chapter 7

Conclusion and Future Work

> “ Each problem that I solved became a rule, which served afterwards to solve other problems.”
>
> RENÉ DESCARTES, 1637

INCENTIVE engineering for online work has shown to be an important factor for crowdsourcing platform designers as the behavior of crowd workers can be influenced, consequently affecting the desired outcome. The work at hand documents the implementation of different studies evaluating different incentives and their effect on crowd workers' behavior. This lead to numerous improvements and insights for crowdsourcing platform designers. The full picture of these contributions are presented in the previous Chapters (4, 5, and 6). This chapter summarizes the main contributions in a briefly manner and outlines pathways for potential future research.

The remainder of this chapter is structured as follows. Section 7.1 summarizes the main contributions of this work. Section 7.2 gives an outlook for future research. Section 7.3 closes this thesis.

7.1 Summary and Contributions

The main goal of this work is to address the following five Research Questions:

RQ1 Which incentive scheme – piece rates or rank-order-tournaments – incentivizes crowd workers' effort best?

RQ2 Does feedback in rank-order-tournaments affect crowd workers' effort in crowdsourcing settings?

RQ3 How is risk-taking behavior affected in long lasting rank-order-tournaments?

RQ4 Is framing capable of influencing behavior in non-monetary settings towards a certain goal?

RQ5 Can a customized framing for a non-monetary setting lead to similar behavior as in monetarily incentivized settings?

Contribution 1: *Which incentive scheme – piece rates or rank-order-tournaments – incentivizes crowd workers' effort best?*
It is known that incentives are needed to spur participation and a desirable outcome. Crowdsourcing uses mainly two different payment schemes to incentivize participation, quality of work, and effort – PRs and ROTs (see Subsection 2.1.1 of Chapter 2). However, it is not clear which payment structure incentivizes crowd workers' effort best. Hence, the work at hand specifically evaluated these two payment schemes and their effect on crowd workers' effort, an important outcome of work in general and particularly crowdsourcing platforms. By implementing a field experiment on the most prominent micro-task crowdsourcing platform – MTurk[1] – valuable insights could be gathered. Given balanced average payouts, both payment structures show similar effects on crowd workers' effort. None of the two incentivizes crowd workers' effort better than the other incentive structure. In summary, without further tweaks of the setup, straightforward PRs and ROTs without a feedback scheme incentivize crowd workers' effort equally. For designers of crowdsourcing platforms this means, that they cannot improve a crowd worker's effort by simply changing the payment scheme. Hence, they should stick with their currently implemented payment scheme, if they do not implement further tweaks.

[1] `https://www.mturk.com/mturk/welcome` last accessed: September 6, 2016.

Contribution 2: *Does feedback in rank-order-tournaments affect crowd workers' effort in crowdsourcing settings?*
Most crowdsourcing tournaments use a feedback mechanism to inform crowd workers about their relative performance – a ranking (see Chapter 4 for further details). Such feedback mechanisms are evaluated. First, a theoretical moderation mediation model of the interplay of different feedback schemes on crowd workers' effort was implemented by summarizing existing evidence. This gives practitioners an overview of existing findings and guidance of how feedback influences crowd workers' effort in ROTs. Second, this interplay was evaluated with the implementation of two field experiments on MTurk[1]. With the first experiment, it was shown that giving feedback in a ROT, specifically showing a ranking about the current position in a dyadic ROT, decreases crowd workers' effort. To be more specific, giving feedback about a strong competitor increases the likelihood that a crowd worker quits the task. Feedback about a weak competitor lets crowd workers exert less effort, since they win even if they relax, but do not lead to higher rates of quitting the task. Competitors showing mediocre performance lead to results in between. With the second experiment, the findings regarding the mediocre feedback could be replicated. Furthermore, it was shown that crowd workers competing with equally good competitors reduce their effort as well. Overall, giving feedback, i.e., showing a ranking, about different strong competitors leads to decreased crowd workers' effort. However, underlying reasons are different. Good performing crowd workers, continue to work on the task but do not give their best, while bad performing crowd workers tend to quit the task, leading to less exerted effort as well. Wrapping up, all feedback schemes which were evaluated lead to worse crowd workers' effort than dyadic ROTs without feedback. Furthermore, the best dyadic ROT did not enhance crowd workers' effort compared to PRs (see Contribution 1). This gives a clear guidance for crowdsourcing platform designers. The best incentive for short tasks is a simple PR, as it leads to performance on par with the best ROT design and is easiest to implement (standardly provided by micro-task platforms such as MTurk[1]). If a platform designer wants to implement a short lasting ROT, he is best of by not showing feedback (i.e., a ranking) throughout the task.

Contribution 3: *How is risk-taking behavior affected in long lasting rank-order-tournaments?*
After demonstrating the interplay of payment schemes, feedback within ROTs, and crowd workers' effort another behavioral aspect of crowd workers within ROTs is evaluated – risk-taking. Moreover, the analysis focuses on long lasting ROTs, in contrast to short lasting ROTs (see Contributions 1 and 2). Overall three sub-questions are evaluated. The evaluation was implemented as a field experiment accompanying the FIFA World Cup 2014.

Contribution 3.1: *Does the tournament mode (teams or individuals) affect risk-taking behavior?*
First, it could be shown that in long lasting ROTs the tournament mode affects a crowd worker's risk-taking behavior. Comparing individuals competing against each other (IND) and teams competing against other teams (TEAM) it was shown that individuals competing on their own take higher risks.

Contribution 3.2: *Does the tournament progress affect risk-taking behavior?*
Second, the effect that individuals (IND) take higher risks (see Contribution 3.1) interacts with the tournament progress. Overall, all participants took higher risks over the course of the tournament. However, this effect was stronger in the case that individuals compete against each other (IND).

Contribution 3.3: *Does the ranking position affect risk-taking behavior?*
Third, turning to the ranking within the ROT, results are somewhat surprising. An effect of the payout relevant position could not be observed. It seems, that the effects from the treatment (IND or TEAM) and the progress of the tournament overcompensates a possible effect of the payout relevant rank on a participants risk-taking behavior. Moreover, the complementary non-relevant rank, i.e., a comparison with the peers within the own team or with the group score of other individually ranked groups, affects risk-taking behavior. It decreases risk-taking behavior of participants who are ranked worse. It seems like participants compare their own strategy with their peers when they are performing badly.

Overall, this means that for crowdsourcing platform designers, creating long lasting tournaments is a suitable design option to induce risk-taking behavior of crowd workers. Especially, the design of individuals competing against each other induces high risk-taking

behavior. Crowd workers take higher risks over the course of the ROT. If a platform designer wants to avoid this, he is better of implementing a ROT where teams compete against other teams, as this effect is less pronounced (TEAM). The effects might be explained with the results of the associated questionnaire, which showed that participants sense higher competition in the individual ranking (IND), while higher group belonging and social presence in the team competition (TEAM). Hence, participants felt less responsible for their individual decision, since risk is shared with the other team members. This might have led to the lower sensed competition and resulted in less risk-taking.

Contribution 4: *Is framing capable of influencing behavior in non-monetary settings towards a certain goal?*
Besides monetary incentives (see Contributions 1, 2, and 3) crowdsourcing often uses non-monetary incentives (Wikipedia[2], StackOverflow[3], ReCAPTCHA[4], etc.). In Chapter 6 the focus was on such non-monetary incentives. In particular it was evaluated if a framing can alter a crowd workers' behavior, as it is reminding him of his initial motivation to participate in the crowdsourcing task. The evaluation was implemented as set of two field experiments in the context of the German government financed project "Planspiel Flächenhandel". Results indicate on three levels that a customized framing positively affected the behavior of participants, regarding the outcome. First, average prices in a market implementing a cap and trade system for TDRs are reduced after participants have been framed. Second, average spendings, thereby the indebtedness of participants representing local authorities, are reduced. Third, new land consumption is differently evaluated after the framing. Participants take the value of new land consumption into account and hence, have a higher focus on positively fiscally evaluated land. Overall, it was shown, that a framing, customized to the platforms context and reminding participants of their initial motivation to take part and the importance of the context they work on, affects the outcome positively. For crowdsourcing platform designers this means, that framing people is a beneficial option to induce a certain behavior and improve the outcome of crowd workers.

[2] `https://www.wikipedia.org/` last accessed: September 6, 2016.
[3] `http://stackoverflow.com/` last accessed: September 6, 2016.
[4] `https://www.google.com/recaptcha/intro/index.html` last accessed: September 6, 2016.

Contribution 5: *Can a customized framing for a non-monetary setting lead to similar behavior as in monetarily incentivized settings?*
After showing that framing can positively affect a crowd worker's behavior (see Contribution 4), it was evaluated how far this behavior is off of standardly incentivized laboratory settings. Therefore the field experiment (see Contribution 4) was replicated with students in a laboratory experiment. Results show no differences between the field- and laboratory experiments on all three levels, i.e., prices per TDR, spendings of participants, and evaluation of new land consumption. Hence, the behavior of framed participants is similar to the behavior of monetarily incentivized students. Overall, this means that in the given case of the German government financed project "Planspiel Flächenhandel" a customized framing could incentivize participants to behave as in a laboratory setting, even without monetary compensation. For crowdsourcing platform designers this not only means, that framing people improves the outcome. In a scenario similar to the project "Planspiel Flächenhandel" where crowd workers' participation is based on intrinsic motivation, a framing highlighting this can potentially compensate a monetary incentive.

All contributions were derived from custom made web-applications, that use or mimic crowdsourcing platforms. These web-based platform approaches have the advantage to mimic remote work as it is standard for crowdsourcing. Hence, a higher external validity can be observed, compared to the mainly used laboratory experimental setups. Furthermore, the conducted studies focus on monetary and non-monetary incentive settings. Other design options of crowdsourcing platforms, e.g., appeal of design, user profiles, quality measures, etc., are not focus of this thesis. Hence, generalizability of the results to any crowdsourcing or similar platform might not be given. However, the presented contributions are a step in the direction to answer the practitioners question of how to design crowdsourcing platforms. Thus, the work at hand contributes to better understand the interplay of incentives and crowdsourcing and thereby, outlines a potential pathway for the future of work.

Furthermore, all studies are exploratory studies using experimental procedures. Experimentation serves different roles in different research traditions. In the information systems literature, Boudreau et al. (2001), for example, posit that experiments take place in settings created by the researcher for the investigation of a phenomenon: The researcher controls independent variables (e.g., feedback, tournament mode, framing, etc.), creates

different treatment conditions by varying these independent variables, randomly assigns research participants to these treatment conditions, and measures the impact on one or more dependent variables (e.g., crowd workers' effort, risk-taking, etc.). These experimental techniques are used in the work at hand's studies. In economics, experimental research has a long tradition. It is accepted that experiments can serve multiple purposes. Roth (1986, 1987), for example, distinguishes three classes of experiments under the labels "speaking to theorists", "searching for facts", and "whispering in the ears of princess". Experiments speaking to theorists are designed to test well-articulated formal theories. Experiments searching for facts explore phenomena where existing theory may have little to say; they are "often designed without reference to a specific body of theory" (Roth, 1987, p. 2). Experiments whispering in the ears of princess are designed to resemble natural environments and inform policy decisions. On the backdrop of this experimental economics perspective, this work's exploratory studies are experiments searching for facts (Roth, 1986, 1987), more precisely they are all framed field experiments (Harrison and List, 2004). A contrary perspective common in the social sciences (Stebbins, 2001) and applied to information systems research by, e.g., Briggs and Schwabe (2011, p. 98) suggests that the goal of experimental research "is to test the propositions of a deductive nomological theory. It may also be called confirmatory research." In this perspective, only studies "speaking to theorists" (Roth, 1986, 1987) can be considered experiments. In order to be clear on the exploratory nature of this thesis, these empirical studies are referred as "exploratory studies using experimental techniques".

7.2 Future Work

A Guideline on How to Conduct Crowd Experiments

Lessons learned from the online crowd experiments from Chapters 4, 5, and 6 show, that conducting crowd experiments comes with a vast amount of challenges compared to classical laboratory or field experiments. By conducting several online experiments with different crowds many challenges occurred, which in classical laboratory setups are unknown, since classical experimental economics literature was developed before the rise of crowdsourcing platforms (see Friedman and Sunder, 1994). The validity of such experiments in comparison to economic laboratory experiments was comprehensively discussed in the

literature (Paolacci et al., 2010; Chilton et al., 2010). However, compared to classical lab experiments, researchers are facing several challenges while conducting online experiments with crowds as well, e.g., synchronization and control. To overcome such issues on crowdsourcing platforms, guidelines leading through the process of conducting experiments are needed.

A set of comprehensive guidelines for behavioral research on crowdsourcing platforms have already been published (Mason and Suri, 2012; Horton et al., 2011; Paolacci et al., 2010). However, these guidelines often do not focus on economic experiments but on behavioral research in general. Furthermore, against the background of the fast growing number of experiments, especially on MTurk[1], crowd workers start to get used to certain experiment types (Chandler et al., 2014). As a consequence, guideline-based experimental design approaches become outdated rapidly. Approaches to overcome this issue need to be developed steadily since researchers have to address the new insights of crowd workers. Searching for approaches is exhausting. Researchers do not report their own approaches extensively, since it is mostly not the researcher's focus. Guidelines who do report approaches are often outdated and might not work to the expected extend. This stresses the need for a continuously updating guideline with a specific focus on economic experiments.

Future work should hence develop a state-of-the-art guideline for crowd experiments including common challenges and best practices from the experimental economics literature. A potential approach should use a well known process, which most researchers are familiar with as a starting point, such as the work of Friedman and Sunder (1994), depicted in Figure 7.1, and transfer it to a conceptual framework. Each of these stages should then be extensively described and should highlight typical challenges, which researchers would face when transferring an experiment to a crowd setup, and then highlight approaches and examples to overcome them. Furthermore an architecture of an open platform facilitating continuous updates of the guideline should be developed and implemented.

The process of conducting economic experiments can be divided in (i) the experimental design stage focusing on which experimental setup suits the research question best, (ii) the sampling or recruitment of participants, (iii) conducting the experiment, and (iv) the analysis of the results (see Friedman and Sunder, 1994). When transferring these steps to crowd experiments, researchers have to address certain challenges to ensure result quality.

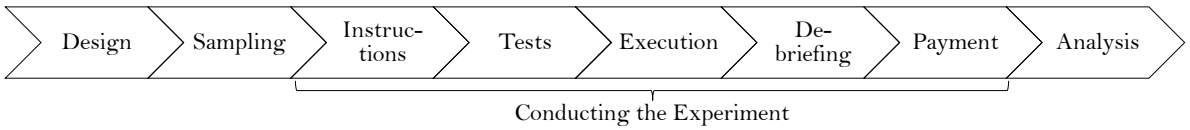


FIGURE 7.1: *Process of conducting experiments based on Friedman and Sunder (1994).*

A guideline should hence highlight these challenges and suggest possible approaches retrieved from literature. In the following, such a possible approach is exemplarily presented for the experimental design stage, by first highlighting the challenges and then showing approaches.

One of the first steps in experimental research is to decide which experimental design suits the research question best. Common design decisions comprise whether an experiment requires (i) asynchronous or synchronized decision-making and if it should be conducted as (ii) a field or a laboratory study (Friedman and Sunder, 1994).

(i) In asynchronous as opposed to synchronous experiments participants do not compete simultaneously against each other. Since in laboratory studies participants usually are in the same room during the observation, it is easy to implement both setups. On MTurk and other online crowd scenarios however, the implementation of synchronous experiments is challenging. Usually tasks are designed as open calls (Howe, 2006b). It is unclear if or when a crowd worker starts completing the task. This makes it difficult to realize experiments where two or more participants have to compete simultaneously against each other due to unknown arrival times of crowd workers (Mason and Suri, 2012; Mao et al., 2012).

First, it should be checked, whether the underlying research question can be addressed with an asynchronous experiment design as well. If possible, the experiment can be redesigned in an asynchronous setup. Second, if participants do not have to compete live against each other, playing against historical data from an earlier observation is possible (Amir et al., 2012; Suri and Watts, 2011; Straub et al., 2015) (e.g., the studies of Chapter 4). Third, if the live interaction and reaction between participants is indispensable a " virtual waiting room" can be implemented (Mao et al., 2012; Mason and Suri, 2012; Paolacci et al., 2010). However, a shortcoming of this approach is that it is unclear how long a participant has to wait, since arrival times vary. One possible approach to address

this is paying a fixed fee after a certain amount of waiting time or using bots in case that waiting times are too long (Horton et al., 2011).

(ii) In a field experiment participants usually do not know that they are being observed and the experimental setup is camouflaged. This leads to a high external validity but lower internal validity. In laboratory studies the actions of a participant are observed in a controlled environment – the laboratory. Isolated cabins prevent unregulated contact and ensure that participants are not influenced by uncontrolled stimuli. Variance in noise, light, and technical factors like input devices, monitors, etc. can be suppressed. Therefore external confounding factors are minimized and internal validity is higher (Friedman and Sunder, 1994). In theory both, field and laboratory setups, can be realized on crowdsourcing platforms like MTurk. However, the internal validity of experiments on MTurk compared to laboratory settings might be lower due to less possible control.

To be more specific, crowd workers might not pay attention during the observation (Paolacci et al., 2010; Horton et al., 2011; Mason and Suri, 2012). Since participants on MTurk work usually from their own computer it is nearly impossible to control their environment during the observation (Chandler et al., 2014; Crump et al., 2013; Rand, 2012). Chandler et al. (2014) finds that participants are watching TV or listening to music while working on MTurk. Another problem with crowdsourcing is that some of the crowd workers try to maximize their payout by finishing as many tasks as possible, often just clicking through tasks. So called “malicious” crowd workers or “spammers” are not paying attention and jeopardize the overall data quality of results (Krause et al., 2016).

First, the overall task design can be aligned to incentivize crowd workers to take the task seriously. Tasks that are fun, interesting, and meaningful incentivize participants to pay attention (Kittur et al., 2013; Crump et al., 2013). Layman and Sigurdsson (2013) show that tasks designed as a game are more satisfying for a participant and thereby motivate to pay attention. Furthermore, researchers could state the expected result quality and give context about the overarching goal in the instructions to give the task a meaning (Oh and Wang, 2012; Oppenheimer et al., 2009). Stating that the task is an experiment and participation helps research, can as well give the task a meaning, if experimenter bias is not a problem (Orne, 1962).

Second, besides redesigning the overall experimental task, experimenters can try to exclude participants who do not pay attention before the actual observation. Many re-

searchers test if a participant is paying attention during the instructions and exclude those who fail the respective test from the sample (Paolacci et al., 2010; Peer et al., 2013; Oppenheimer et al., 2009; Paolacci and Chandler, 2014). Oppenheimer et al. (2009) introduced the instruction manipulation check, which was recently applied by many researchers (Hall and Caton, 2014), as well as in Chapter 4. The fundamental idea of the instruction manipulation check is to trick inattentive participants with a question or free text field which is easy and at best straightforward to answer, e.g., "What is your age?" The instructions state at one point that this particular question should be ignored. Consequently participants who do not read the instructions carefully, e.g., stating their age, can be excluded from the task (Goodman et al., 2012).

A continuously updating guideline can be successfully put into practice within a collaborative process. Hence, researchers should work together to integrate their (new) insights to an online platform. A possible framework for such a process could be structured as follows: First, if researchers decide to conduct a crowd experiment they should have access to the most recent version of the guideline from the platform as depicted in Figure 7.2. Second, challenges that apply to the experimental setup should be possible to identify by looking up approaches in the registered insights from other researchers. Third, researchers should incorporate these approaches to their experimental setup or develop new approaches based on the challenges and hints from the platform. Fourth, after the researchers conducted their experiment they should update the platform based on their findings, e.g., if and how good the applied approaches worked. Through such a process continuous updates to the guideline are facilitated based on collaborative input from researchers and the community.

However, certain requirements should be implemented. Malicious crowd workers might try to trick such a system by accessing the platform in order to get defective insights. Therefore access should be restricted to researchers and practitioners. Login systems only giving access to confirmed users and other security measures must be developed.

Overall the most important factor for a continuously updating guideline is the integration of new insights and results. Therefore researchers must be incentivized or enforced to incorporate their knowledge to the platform. Social rankings raising reputation and chances for citations might incentivize participation. Another conceivable way would be to enforce participation by restricting access with a fee, which a researcher gets back once

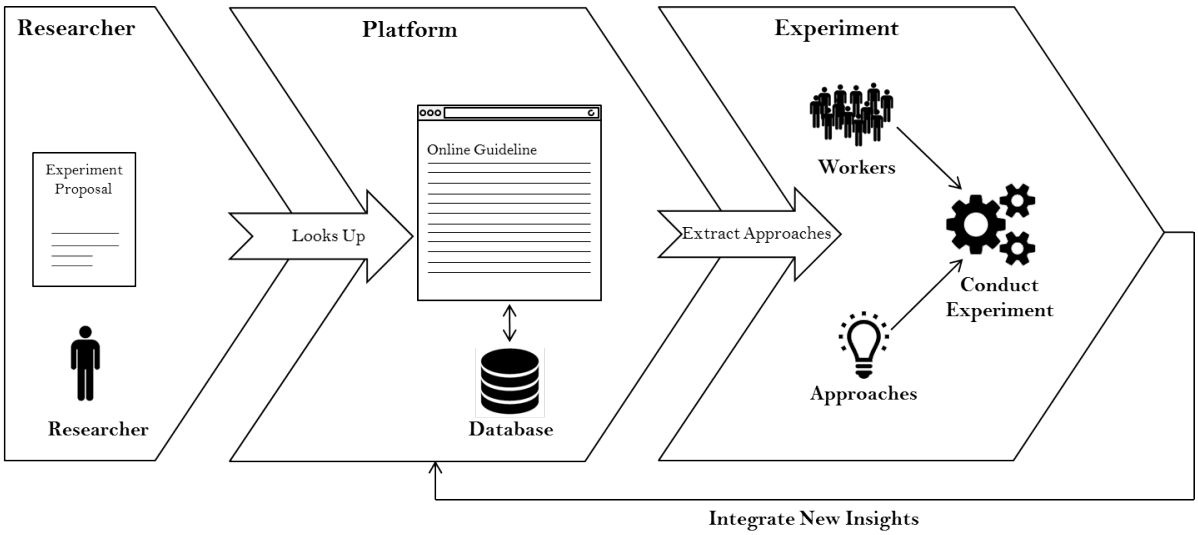


FIGURE 7.2: *Framework of an online platform for a collaborative and continuous crowd experiment guideline.*

he updates the platform. However, finding and developing the correct incentives is one of the research challenges this future work proposes.

So far this future work is a conceptual proposal. Following the notion of collaborative online work and crowdsourcing, a demonstration version should be developed to incorporate practitioners and community feedback in future iterations. Hence, first steps should elaborate and finalize the guideline concept and integrate it in a platform to facilitate a live deployment. Possible future extensions of such a platform include crowd worker databases to block users who already participated in similar experiments as proposed by Chandler et al. (2014) and experimental databases to look up which experiments other researchers already conducted.

NeuroIS and Crowdsourcing

In Chapter 4 the results showed that effort levels of crowd workers decrease in a dyadic ROT with an equally well performing competitor. A potential explanation could be that crowd workers perform worse under pressure. In a scenario, where the competitor is always on par with a crowd worker might be more stressful and hence increases pressure and arousal. A future study should thus shed light on this assumption by transferring this study to a laboratory setup where internal states such as the heart rate are measured as a proxy for a subjects' arousal level. Todays measurement tools, such as biosensors, give

access to the physiological state of a person. Potential applications and further studies could then involve tools for crowd workers to access live bio feedback about their current physiological state. In combination with nowadays affordable biosensor tools, such as webcams and fitness trackers measuring the heart rate, and instructions for self regulation of the internal state, the overall performance of the crowd could be potentially improved.

Selective Rankings and Quality in Crowdsourcing

Results from Chapter 4 showed that feedback in ROTs decreases performance of crowd workers compared to PRs. Hence, the best ROT design is without feedback, leaving the crowd workers in the dark if they are winning or not, when it comes to effort. However, it is straightforward to assume that in long lasting tournaments for creative tasks, such as creating logos and designs (e.g., 99designs[5], threadless[6], etc.), feedback might increase the quality of results.

In relations between the requester and the crowd workers, where the crowd workers have to meet the specific expectations of the requester, communication and feedback might help to guide the workers in the right direction. However, from Chapter 4 we know that giving such a feedback decreases performance. Future work should hence further analyze the relation between ROT feedback and quality. It might still be vital to give feedback in a ROT when a high quality of results and not performance is the desired outcome of the market. A future study could potentially analyze further design tweaks of ROT feedback, as depicted in Figure 7.3. Especially not giving a public feedback, as this decreases crowd workers' effort compared to not giving feedback (see Chapter 4), but giving private feed-

No Feedback

Rank	Name	Score
1.	_____	__
2.	_____	__
3.	_____	__
4.	_____	__
5.	_____	__

Public Feedback

Rank	Name	Score
1.	Player 3	4.5
2.	Player 2	4.0
3.	You	3.8
4.	Player 1	2.4
5.	Player 4	1.7

Private Feedback

You	3.8

FIGURE 7.3: *Potential treatments of a future study analyzing feedback mechanisms and quality.*

[5] `http://en.99designs.de/` last accessed: September 6, 2016.
[6] `https://www.threadless.com/` last accessed: September 6, 2016.

back to each crowd worker separately might increase quality, as it is proven to affect a crowd worker's behavior (Mazarakis and Van Dinther, 2011).

Adaptive Feedback

One limitation of Chapter's 4 studies is that the given feedback was from (only) one competitor. If you beat this competitor a participant received a bonus. However, before the experiment and to some extend during the experiment it was unclear how many tasks this competitor will finish (see Chapter 4). This might have lead to a demotivational factor of some participants who, not knowing how good they have to perform, thought that they will not be able to win. Hence, stating how many tasks a participant has to finish to receive a bonus before the experiment might increase a participant's effort level.

A future study could, hence, analyze if, by stating the amount of tasks a participant has to finish given a certain time frame, the overall effort level can be increased. The setup should analyze repeated tournaments. In each new stage (tournament round) the amount of tasks to finish in this time frame should be adopted based on the result of the last round. If the participant has achieved the goal, the next round's goal should be slightly higher. If the participant fails, the next round's goal should be slightly lower, as depicted in Figure 7.4. Doing so, the goal is achievable for every participant, while still getting harder as long as the participant can do better. Once the goal reached a level higher than his maximum effort level and he cannot beat it, the challenge will become easier. Such an adaptive setup should lead to a higher motivation in long lasting tasks, while still the highest effort of each individual can be potentially achieved. A control scenario could be a tournament where you always have to beat a median score from pre-tests, which will not adopt over the rounds. The analysis could then answer on the one hand, if an adaptive feedback leads to higher effort levels. On the other hand, it could evaluate if such an adaptive feedback leads to continued participation in long lasting tournaments, i.e., lower quitting rates. Therefore, a participant should be able to exit the experiment after each tournament round or to continue and earn more money, as depicted in Figure 7.4.

Crowdsourcing Trust

Through recent developments crowdsourcing applications now face the challenge of the crowd participating and cooperating as a network (Prpić, 2016). In such P2P networks, like

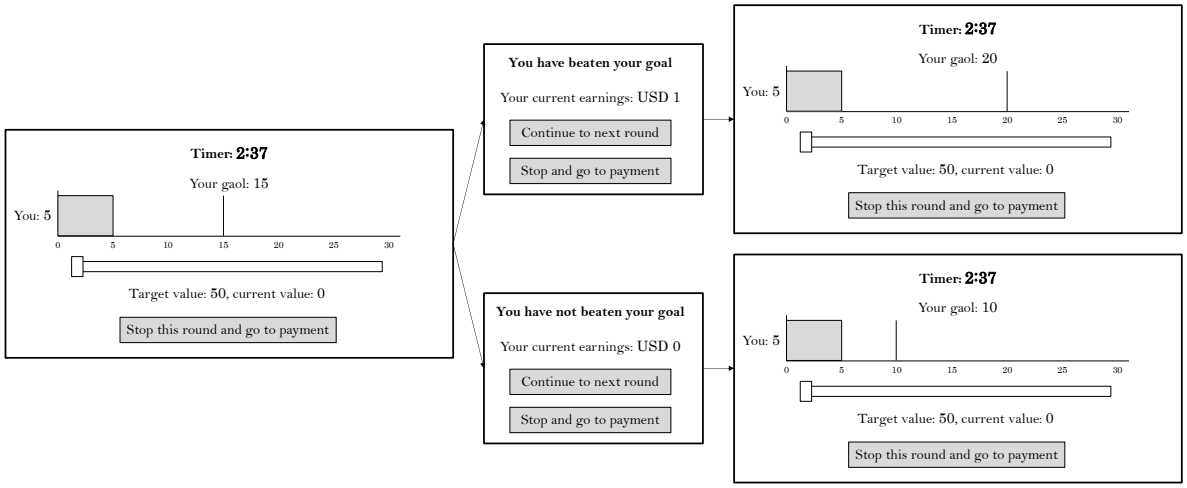


FIGURE 7.4: *Potential treatment process of a future study analyzing adaptive feedback, effort, and motivation in long lasting tasks.*

the sharing economy, the crowd shares goods, such as flats (Airbnb[7]) or cars (BlaBlaCar[8]) with each others. In such situations where the crowd faces each other and wants to exchange goods it is important to trust the respective peer of the transaction (Hawlitschek et al., 2016). However trust to a specific peer is a phenomena which a person develops over time based on past transactions. Hence, such platforms have to overcome the problem of building an initial trust before the first transaction of the respective peers.

One already used mechanism of such platforms are user profiles. However, the question arises if such profiles can be optimized to increase this initial trust. Especially user profile pictures seem to be a promising starting point as they give a first impression of the respective peer and are known to increase trust (Teubner et al., 2014). A future study might examine if crowd workers of an online labor market (e.g., MTurk[1]) are capable of

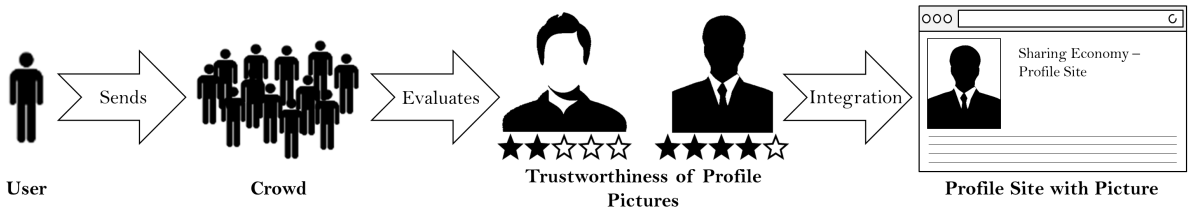


FIGURE 7.5: *Potential process of crowdsourcing trust-scores for profile pictures of sharing economy platforms.*

[7] `https://www.airbnb.com/` last accessed: September 6, 2016.
[8] `https://www.blablacar.de/` last accessed: September 6, 2016.

evaluating how trustworthy a person appears on his profile picture. In such a setup several crowd workers could rate user profile pictures with a likert scale implementing the trustworthiness of the user's profile picture. The mean score could then serve as a proxy for the trustworthiness of the user. Potential research questions involve if profile pictures with a higher trust-score show higher trust in a trust game experiment and how many reviews from crowd workers are needed to build a suitable trust-score. A future application could then, for example, be an integrated service of sharing economy platforms where a user can evaluate his profile picture by a crowd for a small monetary compensation, to increase his likelihood for transactions on such a platform, as depicted in Figure 7.5. This could potentially offer a cheap and fast mechanism given the small monetary compensations on online labor markets, the large subject pools, and the fast execution times of such a task.

Prediction Markets and Rankings

Another prominent crowdsourcing application is a prediction market (e.g., Economic Indicator Exchange[9], intrade[10], or PredictIt[11]) which aim to forecast future events (Kranz et al., 2014). On such markets crowd workers trade stocks which constitute the outcome of future events, such as on real stock markets. The current price on a prediction market usually accounts for the likelihood of the outcome. Thereby, these markets give predictions for future events at any given point in the lifespan of such a market.

A potential research question could be which incentives are best suited to get the attention of the crowd, which could evaluate the findings from Chapter 4 in a real world crowdsourcing application – a prediction market. Furthermore it could be evaluated if increasing the risk-taking behavior of crowd workers could increase the precision of predictions. Thereby the ranking design of Chapter 5 could be further extended and tested on such a prediction market. In addition, the evaluations of Chapter 4 are missing a longitudinal study if these incentives motivate crowd workers over time for a continuous participation. Rankings however, offer a variety of design tweaks, such as selective feedback and information messages after special events, such as increasing a rank after a good trade. Prediction markets offer a tool to analyze if such design tweaks motivate crowd workers to participate in the long run.

[9] `http://www.eix-market.de/` last accessed: September 6, 2016.
[10] `https://prev.intrade.com/` last accessed: September 6, 2016.
[11] `https://www.predictit.org/` last accessed: September 6, 2016.

Standard List

Rank	Suggestion	Votes
1.	Name 3	3500
2.	Name 1	3000
3.	Name 4	2000
4.	Name 5	800
5.	Name 2	150

Random List

Rank	Suggestion	Votes
1.	Name 4	2000
2.	Name 5	800
3.	Name 1	3000
4.	Name 2	150
5.	Name 3	3500

Featured List

Rank	Suggestion	Votes
feat.→	*Name 2*	*150*
1.	Name 3	3500
2.	Name 1	3000
3.	Name 4	2000
4.	Name 5	800

Reversed List

Suggestion	Votes
Name 2	150
Name 5	800
Name 4	2000
Name 1	3000
Name 3	3500

FIGURE 7.6: *Potential treatments of a future study analyzing user interface designs for crowd-voting settings.*

User Interface Design for Crowd-Voting

In crowd-voting, crowdfunding, and crowd-donating settings the crowd is used to decide which is the best out of several options. In such settings it is important to identify the "best" solution (e.g., name, project, etc.). In crowd-voting scenarios, usually, the crowd is used to propose solutions and later to vote on these to identify a winning solution. Often these crowd-voting tournaments are implemented only for a given scenario, such as naming a boat[12]. However, several recent examples have proven, that crowd-voting does not always lead to a satisfying outcome for the requester. For example, a crowd-voting tournament of the Natural Environment Research Council led to the outcome to name their new boat "Boaty McBoatface"[12]. Clearly the institute was not satisfied with this name and named the boat after Sir David Attenborough instead. This example shows, that these scenarios need some control mechanism. A potential future work could, hence, focus on different user interfaces to identify a design which facilitates a different outcome.

Most crowd-voting scenarios usually show just a list of the different suggestions ranked from most liked (most votes) to least liked (least votes). Such a visualization surely facilitates that prominent suggestions like "Boaty McBoatface"[12] are seen by every voter, since it always shows up on the front-page as most popular. The visibility of such a "joke" contribution might spur its popularity. Evidence for this is given by studies analyzing social influences, social biases, and herding effects in such scenarios (Lorenz et al., 2011; Muchnik et al., 2013; Wang et al., 2014). Hence, a potential experiment could analyze different visualizations and their impact on the outcome – here identified winning suggestion. Possible visualizations incorporate (i) a standard list (most votes to least votes) as a basic scenario, (ii) a random list, (iii) a standard list including features (promotion) of

[12]For more details see: `http://www.nerc.ac.uk/press/releases/2016/18-ship/` last accessed: September 6, 2016

currently unpopular suggestion on top of the page, and (iv) a reversed standard list (least votes to most votes), as depicted in Figure 7.6. Such a study could shed first light on the question, if the outcome of crowd-voting scenarios could be influenced by only changing the user interface of such platforms. Together with a customized framing which tells people not to suggests "jokes" (see Chapter 6) and further tweaked visualizations this future work could potentially lead to new crowd-voting platforms where the crowd more likely suggests satisfying solutions.

7.3 Final Remark

In this chapter I presented the five main contributions of the work at hand. Thereby I evaluated several incentives for crowdsourcing applications and their interplay with crowd workers' behavior. First, two common payment schemes which are used in crowdsourcing practice – PRs and ROTs – were compared focusing on effort of crowd workers (RQs 1 and 2). Second, crowd tournaments were further investigated based on their influence on risk-taking behavior (RQ 3). Third, it was evaluated how a framing can suit as non-monetary incentive for crowdsourcing applications (RQs 3 and 4). These contributions (see Section 7.1 of this chapter) give practitioners insights and clear guidance of how to set up incentives for crowdsourcing applications. To sum up, the represented contributions and collected insights about incentives for crowdsourcing provide important details about the effect of different incentives on the user behavior. Platform designers can now incorporate these insights to implement the best possible crowdsourcing applications (see Section 7.1 of this chapter).

Then I shed light on future directions for this research domain, which is built up on lessons learned from the experiments conducted in this thesis. In particular, a framework was presented to collaboratively build an update a guideline for conducting crowd experiments. In addition, six main challenges for future research were derived which build upon and extend the present work (see Section 7.2). The outlined research directions are appropriate to get insights about open questions and extend the results to newly developing crowdsourcing domains. I would be delighted if the contributions and directions for future work help to broaden the understanding of this research domain and potentially inspire future researchers to work in this vastly developing field.

Part V

Appendices

Appendix A

Texts and Platform Screens of Chapter 4

This appendix lists the instructions, platform screens, and questionnaires from the experiments conducted in Chapter 4. Instructions and questionnaires from the experiments are presented in its original English version. Instructions of Study 1 represent the treatment where the PR round is played first. For the case that the ROT round is played first, the order of the rounds in the instructions are changed. Instructions of Studies 2 and 3 are represented by the instructions of Study 2. Instructions of Study 3 were changed where appropriate.

Instructions

Study 1: Piece Rates versus Rank-Order-Tournaments

Please read these instructions carefully:
You participate in an academic experiment on performance in groups.

You will play 3 rounds of a slider task game. A round proceeds as follows:

- At the top there is a feedback of your performance.
- Beneath there is a slider.
- Your task is to position the slider to 50 (it ranges from 0 to 100) as often as possible.
- A new slightly different slider will appear every time you positioned the slider correctly.

- It is possible to stop the round and go to the next round/payment if you click on the button on the bottom of the page.
- **If you do so this will not influence the acceptance of your HIT!** You will always get the earnings you have gathered.

Round 1 will proceed as follows:

- This is a test round.
- You play alone.
- You get used to the task.
- You do not get a payment for this round but for the next 2 rounds.
- The task will last 0:30 minutes.

Round 2 will proceed as follows:

- You play alone.
- You get $0.02 per finished slider.
- The task will last 2:00 minutes.

Round 3 will proceed as follows:

- You play against 1 other participant.
- To prevent any delays, the other participant already finished the task and you play against his historical data.
- If you finish more sliders than your rival you will get a $1.00 bonus. Else you get $0.00 in this round.
- We will inform you if you performed better than your rival after all 3 rounds in the final payment screen.
- If there is a tie we will randomly assign the bonus to one of you.
- The task will last 2:00 minutes.

Your bonus is calculated as follows:

- You get a fix payment for completing this HIT of $0.50.
- You get a bonus payment for every finished slider in round 2 of $0.02.
- You get another bonus of $1.00 if you finished round 3 better than your rival.
- The sum of these 3 will be your final payment.

Do not use a tablet, smartphone or touchpad. The task will be much easier using a mouse.

Please note: Most modern theories of decision making recognize the fact that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables can greatly impact the decision process. In order to facilitate our research, we are interested in whether you actually take the time to read the directions; if not, then some of our research design that relies on changes in the instructions will be ineffective. So, in order to demonstrate that you have read the instructions, please ignore the input field and the continue button below. Instead simply click on the words “please note” at the beginning of this paragraph to proceed to the next screen. Thank you very much.

What is your age in years?

Please insert your age in years in this textfield.

Continue

Study 2 and 3: Feedback in Rank-Order-Tournaments – Different (Equal) Strength(s) of Competitors

Please read these instructions carefully:
You participate in an academic experiment on performance in groups.

You will play 2 rounds of a slider task game. A round proceeds as follows:

- At the top there is a feedback of your performance.
- Beneath there is a slider.
- Your task is to position the slider to 50 (it ranges from 0 to 100) as often as possible.
- A new slightly different slider will appear every time you positioned the slider correctly.
- It is possible to stop the round and go to the next round/payment if you click on the button on the bottom of the page.
- **If you do so this will not influence the acceptance of your HIT!** You will always get the earnings you have gathered.

Round 1 will proceed as follows:

- You play alone.
- You get $0.01 per finished slider.
- The task will last 1:30 minutes.

Round 2 will proceed as follows:

- You play against 1 other participant.
- To prevent any delays, the other participant already finished the task and you play against his historical data.
- If you finish more sliders than your rival you will get a $1.00 bonus. Else you get $0.00 in this round.
- You will see a ranking which informs you about your current performance and payment.

- If there is a tie we will randomly assign the bonus to one of you.
- The task will last 3:00 minutes.

Your bonus is calculated as follows:

- You get a fix payment for completing this HIT of $0.30.
- You get a bonus payment for every finished slider in round 1 of $0.01.
- You get another bonus of $1.00 if you finished round 2 better than your rival.
- The sum of these 3 will be your final payment.

Do not use a tablet, smartphone or touchpad. The task will be much easier using a mouse.

Please note: Most modern theories of decision making recognize the fact that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables can greatly impact the decision process. In order to facilitate our research, we are interested in whether you actually take the time to read the directions; if not, then some of our research design that relies on changes in the instructions will be ineffective. So, in order to demonstrate that you have read the instructions, please ignore the input field and the continue button below. Instead simply click on the words "please note" at the beginning of this paragraph to proceed to the next screen. Thank you very much.

What is your age in years?

Please insert your age in years in this textfield.

Continue

Platform Screens

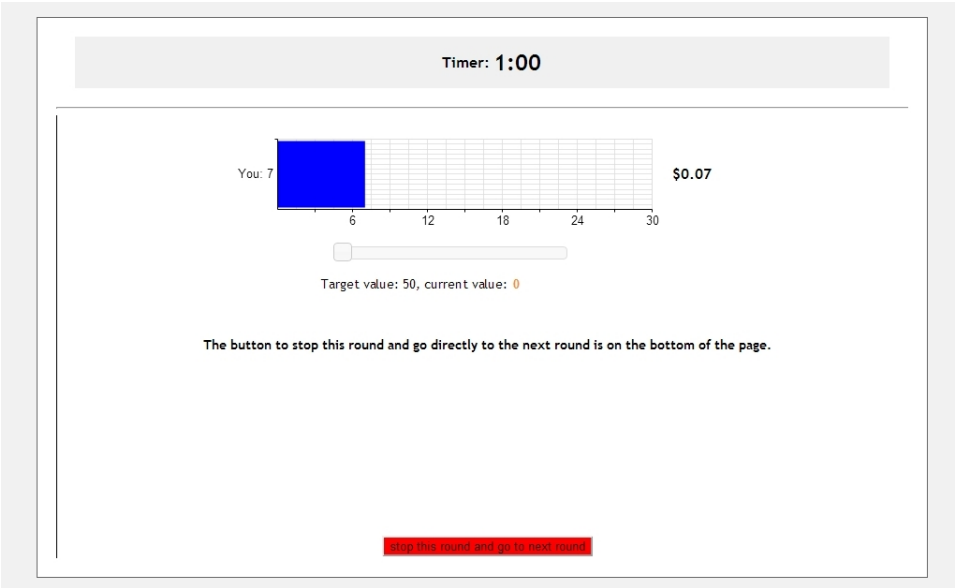


FIGURE A.1: *Web interface of piece rate treatment of Chapter 4.*

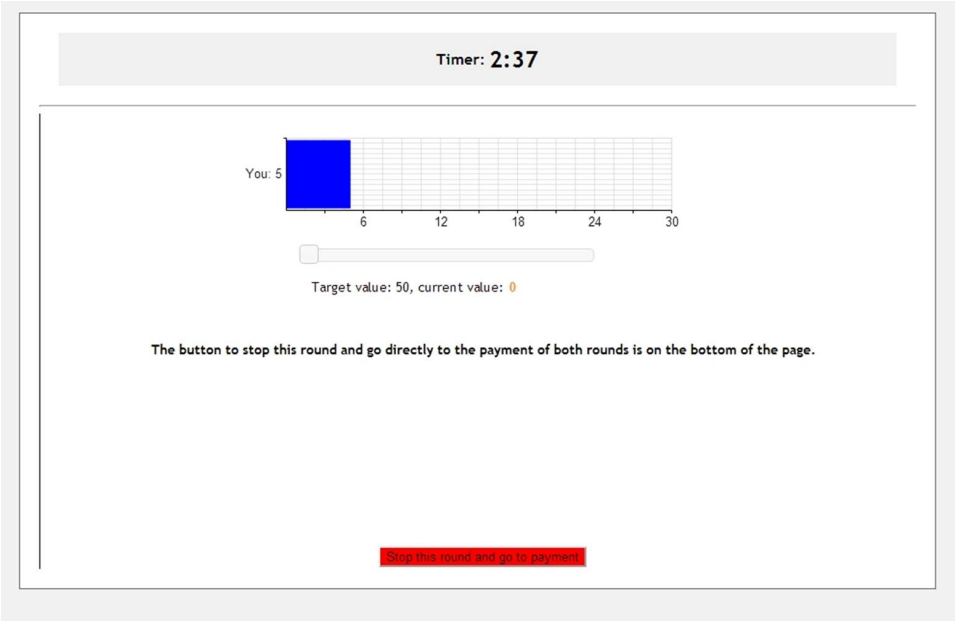


FIGURE A.2: *Web interface of no feedback treatment of Chapter 4.*

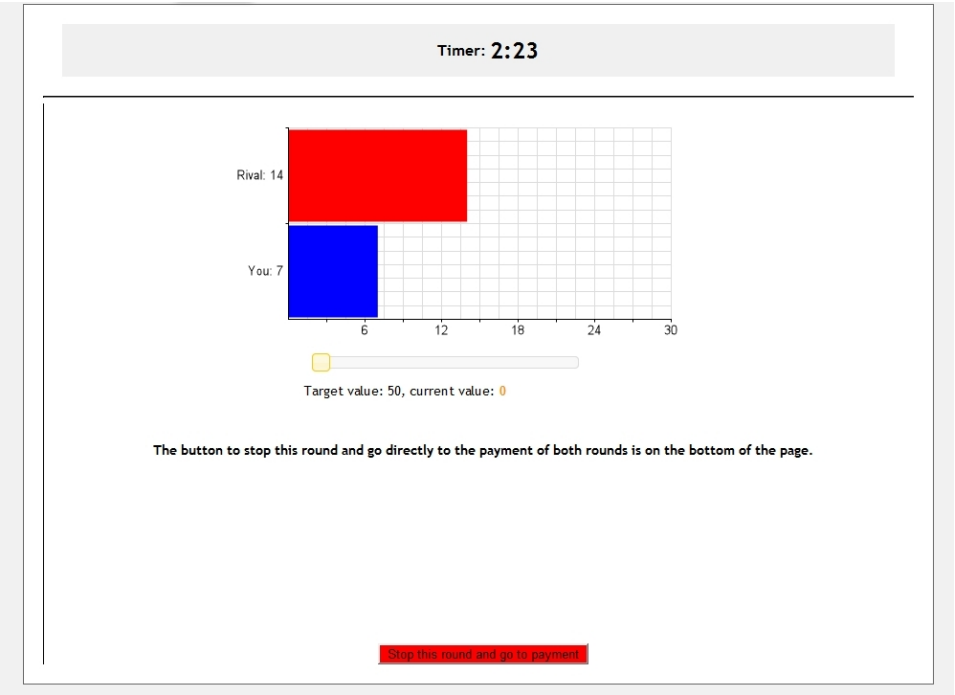


FIGURE A.3: *Web interface of feedback treatments (FW, FM, FS, and FE) of Chapter 4.*

Questionnaires

Instructions Test

We would like to ensure your understanding of the instructions to allow you to participate.

Please answer the 5 following quizzes carefully. If you make too many mistakes, you will not be allowed to participate[1].

In case of doubt, you can review the instructions in a new browser window any time by clicking this link.

1. What is your task?
   a) Estimate the average of a set of numbers
   b) Count the occurrence of a certain number in a given dataset
   c) Adjust a slider to a certain position
   d) Copy a text
   e) Solve a simple maze game

2. Against whom your playing in round 3 (2 for study 2 and 3)?
   a) No one, I'm playing alone
   b) Against historical data from another participant
   c) I'm playing live against another participant
   d) I'm playing live against 3 other participants

[1] After three mistakes a participant was kicked out of the study. They still got the fix fee as reward.

3. To which value do you have to adjust the slider?

a) 50

b) 20

c) To a new value each round

d) 80

e) I don't have to adjust a slider

4. If you are reading this, please answer the third option.

a) Grey

b) Dark Blue

c) Both

d) None of the above

5. How is your bonus calculated?

a) I get a fix bonus. The same as everybody.

b) The bonus is equally distributed between the participants.

c) My bonus will be calculated according to the completed sliders in round 2 (1 for study 2 and 3) and according to my final rank in round 3 (2 for study 2 and 3).

Questions after study 1, 2, and 3

1. What is your gender?

a) male

b) female

2. What is your age?

Please insert your age in years in this textfield.

3. What is your education?

a) Some High School completed

b) High School Diploma

c) Some college completed

d) Associate's degree

e) Bachelor's degree

f) Master's degree

g) Doctorate

h) None

4. How distracted have you been during the experiment?

a) not at all

b) a little

c) very

5. Which input device did you use?

a) Mouse

b) Touchpad

c) Tablet

d) Smartphone

e) Trackball

f) Pointing stick (Lenovo/IBM)

g) Other

6. Which hand do you use to control your input device?

a) left

b) right

How well do the following statements describe your behavior?[2]

- I enjoy competition because it gives me a chance to discover my abilities.
- Competition does not increase my awareness and understanding of myself and others.
- Competition can lead to the formation of friendship with others.
- Competition is not a means of motivating me to bring out the best in myself.
- I enjoy competition because it tends to bring out the best in me rather than as a means of feeling better than others.
- I do not find competition to be a very valuable means of learning about myself and others.
- I like competition because it teaches me a lot about myself.

[2]Questions are based on Griffin-Pierson (1990) and Ryckman et al. (1996). All questions were implemented with a Likert scale from 1 (strongly disagree) to 5 (strongly agree). Questions were randomized for the questionnaire.

- I value competition because it helps me to be the best that I can be.
- I find competition enjoyable because it lets me express my own potentials and abilities during competition.
- Competition does not help me develop my abilities more.
- Without the challenge of competition I might never discover that I had certain potentials or abilities.
- I enjoy competition because it brings me and my competitors closer together as human beings.
- I enjoy competition because it helps me to develop my own potentials more fully than if I engaged in these activities alone.
- I enjoy competition because it brings me to a higher level of motivation to bring the best out of myself rather than as means of doing better than others.
- Through competition I feel that I am contributing to the well-being of others.
- I perform better when I am competing against someone rather than when I am the only one striving for a goal.
- I do not feel that winning is important in both work and games.
- When I win an award or game it means that I am the best compared to everyone else that was playing. It is only fair that the best person wins the game.
- In school, I always liked to be the first one finished with a test.
- When nominated for an award, I focus on how much better or worse the other candidates' qualifications are as compared to mine.
- I would want an A because that means that I did better than other people.
- I have always wanted to be better than others.
- Because it is important that a winner is decided, I do not like to leave a game unfinished.

Appendix B

Texts and Platform Screens of Chapter 5

This appendix lists the instructions, platform screens, and questionnaires from the experiment conducted in Chapter 5. Instructions and questionnaires from this experiment are translated to English from the original German version. The presented instruction is from the IND treatment of the experiment of Chapter 5. The instruction of the TEAM treatment was changed where appropriate.

Instruction

Hello “first name”,

thank you for your participation in this experiment. Here are some relevant informations.

General

This experiment is a betting game. You bet on tendencies (win home, draw, win away) of all 64 games of the FIFA World Cup. You bet on the outcome of the games after regular time. This means that overtime or shootouts are not included (important: This applies as well for the final rounds!).

Each of these events has a quota, similar to commercial sport betting platforms. The quota corresponds to the points you can win.

Placing your bets

Please note that you are only allowed to bet on games 72 hours until 5 minutes before the game starts. You can place your bet by clicking on the respective box (left box = win home, middle box = draw, right box = win away).

If your bet is correct, you will receive points accordingly to the quota. If your bet is wrong or you did not place a bet, you will receive no points.

Group & payment

You are in a group with 5 other participants. Overall there are 6 similar groups. Decisive for the ranking and hence your payment is the overall sum of your points and how you rank compared to the peers in your group. Therefore you are not in competition with the other groups, but with the participants in your group.

The ranking is updated after each game. Your own position is highlighted with yellow.

The payment at the end of the experiment is structured as follows:

- The participant on the first place receives 50€, the second place receives 25€, and the third place receives 15€.
- Each participant who bets on all 64 games receives in addition 5€.
- After the experiment you have to fill out a questionnaire. You can earn additional money dong so.

Besides the ranking of your own group, you can access an alphabetical list of all groups taking part in this experiment. This group comparison is irrelevant for your payment, since you only compete against the participants in your group.

Login

To place bets and to view your ranking, please visit www.golero.de and log in with the following user account:

Username: “username”

Password: “password”

Please confirm that you are able to successfully log in with a short e-mail. Thank You.

If you have additional questions, you can contact us via e-mail (Tim Straub “email” or Dr. Timm Teubner “email”).

Best regards and good luck during the experiment!

Your IISM-Team

Platform Screens

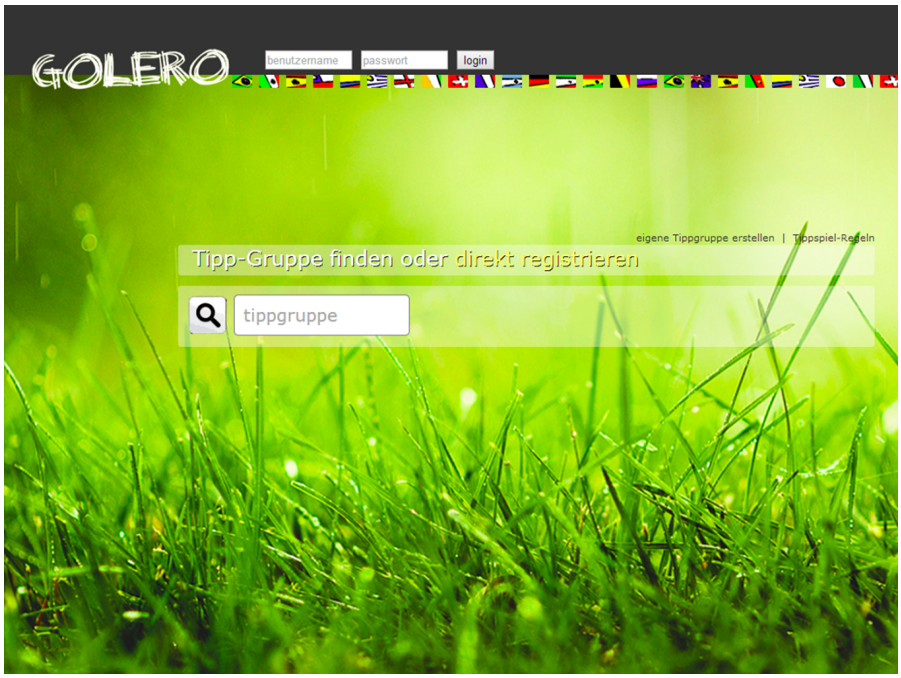


FIGURE B.1: *Log in of the web interface for the experiment of Chapter 5.*

FIGURE B.2: *Inactive betting phase (active 72 hours until 5 minutes before the start of a game) of the web interface for the experiment of Chapter 5.*

FIGURE B.3: *Active betting phase (active 72 hours until 5 minutes before the start of a game) of the web interface for the experiment of Chapter 5.*

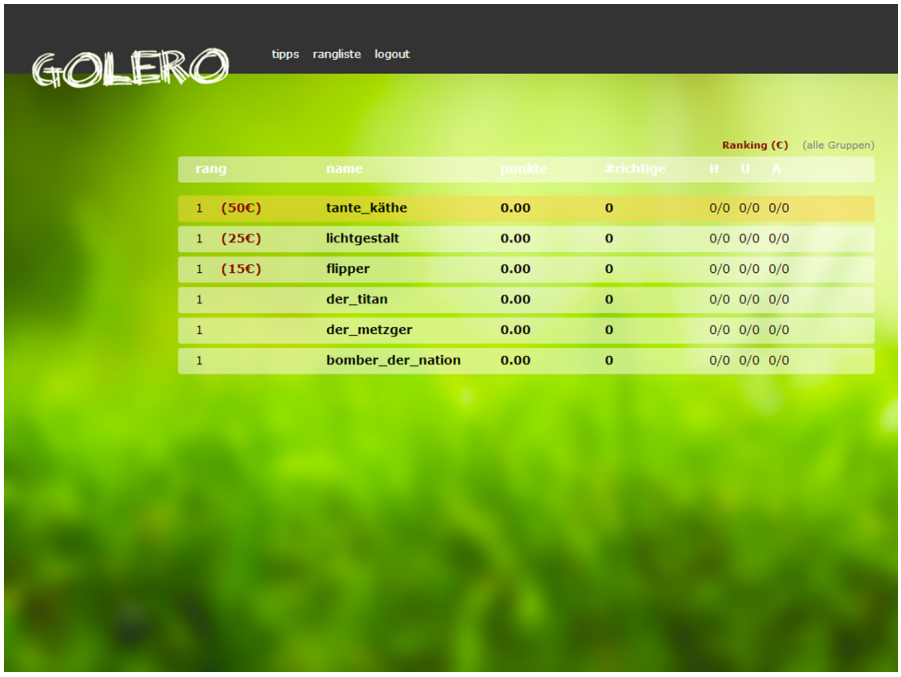


FIGURE B.4: *Ranking of the IND treatment of the web interface for the experiment of Chapter 5.*

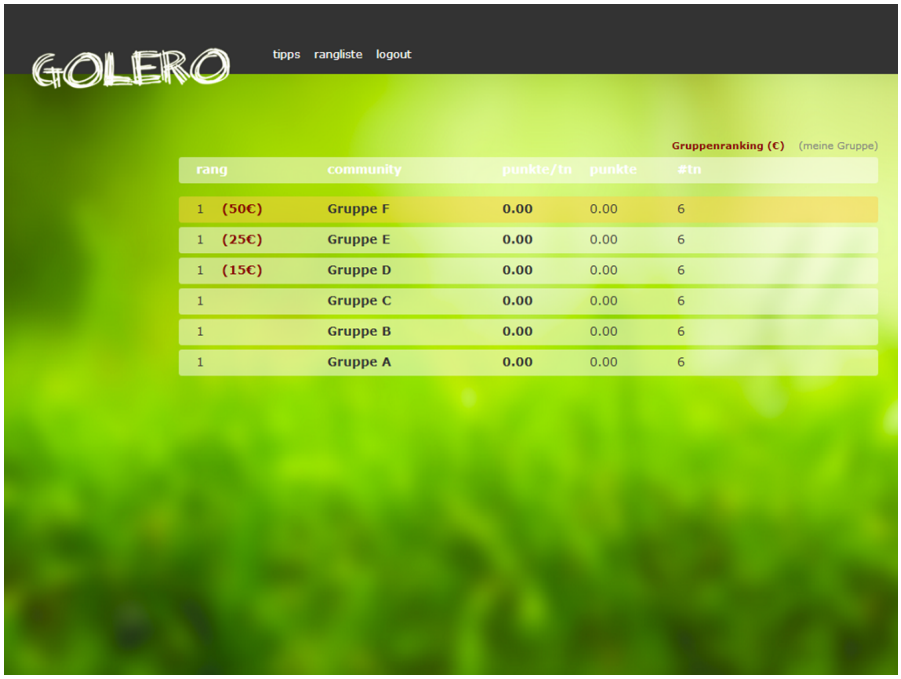


FIGURE B.5: *Ranking of the TEAM treatment of the web interface for the experiment of Chapter 5.*

Questionnaires

Questions after the experiment

First name

Please insert your first name.

Last name

Please insert your last name.

Age in years

Please insert your age.

Citizenship

Please insert your citizenship.

Which is your favorite team of the FIFA World Cup?

Please insert your favorite team.

Did you follow a specific strategy throughout the experiment? If so, what strategy?

Please insert your strategy.

To what extent do the following statements apply to you?[1]

- There is a sense of human contact in this experiment.
- There is a sense of personalness in this experiment.
- There is a sense of sociability in this experiment.
- There is a sense of human warmth in this experiment.

[1]Questions are based on Bollen and Hoyle (1990), Griffin-Pierson (1990), Gefen and Straub (2004), Greene (2004), Hassanein and Head (2005), Rammstedt and John (2007), Byun and Mann (2011), Liu et al. (2013), and Vandercruysse et al. (2013). All questions were implemented with a Likert scale from 1 (not correct at all) to 7 (fully correct). Questions were randomized for the questionnaire.

- There is a sense of human sensitivity in this experiment.
- I see myself as someone who is reserved.
- I see myself as someone who is generally trusting.
- I see myself as someone who tends to be lazy.
- I see myself as someone who is relaxed, handles stress well.
- I see myself as someone who has few artistic interests.
- I see myself as someone who is outgoing, sociable.
- I see myself as someone who tends to find fault with others.
- I see myself as someone who does a thorough job.
- I see myself as someone who gets nervous easily.
- I see myself as someone who has an active imagination.
- I would want to get an A because that is the best grade a person can get.
- I perform better when I am competing against someone rather than when I am the only one striving for a goal.
- I do not care to be the best that I can be.
- When applying for an award I focus on my qualifications for the award and why I deserve it, not on how the other applicants compare to me.
- I do not feel that winning is important in both work and games.
- When I win an award or game it means that I am the best compared to everyone else that was playing. It is only fair that the best person wins the game.
- In school, I always liked to be the first one finished with a test.
- I am not disappointed if I do not reach a goal that I have set for myself.
- I have always wanted to be better than others.
- Achieving excellence is not important to me.

- When nominated for an award, I focus on how much better or worse the other candidates' qualifications are as compared to mine.
- I would want an A because that means that I did better than other people.
- I wish to excel in all that I do.
- Because it is important that a winner is decided, I do not like to leave a game unfinished.
- I would rather work in an area in which I can excel, even if there are other areas that would be easier or would pay more money.
- The experiment makes you feel active.
- The experiment makes you feel excited.
- The experiment makes you feel stimulated.
- The experiment makes you feel lively.
- The experiment makes you feel activated.
- During the experiment I felt competition with other participants.
- During the experiment I was conscious about other participants' behavior.
- During the experiment I felt like I am competing with other participants for points.
- During the experiment I felt like running a race.
- During the experiment, I experienced a high level of competition in general.
- I feel a sense of belonging to the group.
- I feel that I am a member of the group.
- I see myself as part of the group.
- I am enthusiastic about the group
- When I talk about this group, I usually say "we" rather than "they".
- This group's successes are my successes.
- I act like a person of this group to a great extent.

- I found my participation in this experiment interesting.
- I found my participation in this experiment entertaining.
- I found my participation in this experiment enjoyable.
- I found my participation in this experiment pleasant.

Your decision sheet shows ten decisions listed on the left[2]. Each decision is a paired choice between "Option A" and "Option B". You will make ten choices and record these in the final column, but only one of them will be used in the end to determine your earnings. Before you start making your ten choices, please let me explain how these choices will affect your earnings for this part of the experiment.

Here is a ten-sided die that will be used to determine payoffs; the faces are numbered from 1 to 10 (the "0" face of the die will serve as 10.) After you have made all of your choices, we will throw this die twice, once to select one of the ten decisions to be used, and a second time to determine what your payoff is for the option you chose, A or B, for the particular decision selected. Even though you will make ten decisions, only one of these will end up affecting your earnings, but you will not know in advance which decision will be used. Obviously, each decision has an equal chance of being used in the end.

Now, please look at Decision 1 at the top. Option A pays 2.00€ if the throw of the ten sided die is 1, and it pays 1.60€ if the throw is 2-10. Option B yields 3.85€ if the throw of the die is 1, and it pays 0.10€ if the throw is 2-10. The other Decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increases. In fact, for Decision 10 in the bottom row, the die will not be needed since each option pays the highest payoff for sure, so your choice here is between 2.00€ or 3.85€ .

To summarize, you will make ten choices: For each decision row you will have to choose between Option A and Option B. You may choose A for some decision rows and B for other rows, and you may change your decisions and make them in any order. When you are finished, we will come to your desk and throw the ten-sided die to select which of the ten Decisions will be used. Then we will throw the die again to determine your money earnings for the Option you chose for that Decision. Earnings (in €) for this choice will

[2]This short introduction and the following questionnaire are based on the risk-aversion test of Holt and Laury (2002).

be added to your previous earnings, and you will be paid all earnings in cash when we finish.

So now please look at the empty boxes on the right side of the record sheet. You will have to mark a decision, A or B in each of these boxes, and then the die throw will determine which one is going to count. We will look at the decision that you made for the choice that counts, and circle it, before throwing the die again to determine your earnings for this part.

Are there any questions? Now you may begin making your choices. Please do not talk with anyone while we are doing this; raise your hand if you have a question.

Option A	Option B	Choice
1/10 of 2.00€, 9/10 of 1.60€	1/10 of 3.85€, 9/10 of 0.10€	
2/10 of 2.00€, 8/10 of 1.60€	2/10 of 3.85€, 8/10 of 0.10€	
3/10 of 2.00€, 7/10 of 1.60€	3/10 of 3.85€, 7/10 of 0.10€	
4/10 of 2.00€, 6/10 of 1.60€	4/10 of 3.85€, 6/10 of 0.10€	
5/10 of 2.00€, 5/10 of 1.60€	5/10 of 3.85€, 5/10 of 0.10€	
6/10 of 2.00€, 4/10 of 1.60€	6/10 of 3.85€, 4/10 of 0.10€	
7/10 of 2.00€, 3/10 of 1.60€	7/10 of 3.85€, 3/10 of 0.10€	
8/10 of 2.00€, 2/10 of 1.60€	8/10 of 3.85€, 2/10 of 0.10€	
9/10 of 2.00€, 1/10 of 1.60€	9/10 of 3.85€, 1/10 of 0.10€	
10/10 of 2.00€, 0/10 of 1.60€	10/10 of 3.85€, 0/10 of 0.10€	

TABLE B.1: *The Ten Paired Lottery-Choice Decisions risk-aversion test based on Holt and Laury (2002)*.

Appendix C

Texts and Platform Screens of Chapter 6

This appendix lists the instructions, platform screens, and questionnaires from the experiments conducted in Chapter 6. Instructions and questionnaires from the experiments are translated to English from the original German version. The presented texts are from Session 2 (Experiments 3 and 4) of Chapter 6. All other texts were changed where appropriate.

Online Platform's Landing Page Text

Welcome to the online platform of the "Planspiel Flächenhandel".

Several changes were made to this platform in comparison to the first session. The most important is the following:

- Your actions are from now on under the supervision of several institutions, which will question your actions and when appropriate restrict your actions. In focus of this is your behavior when submitting orders as well as your possibility to loan money (credits). Information regarding this will be available via the header of the platform. When corresponding actions occur you will always be informed with pop-up windows.

We would like to highlight, that you participate in this experiment **as representative of a German local authority and you should act realistically**. Take all displayed building

land as given and ask yourself before taking an action, how the real local authority would decide, given the circumstances you face during the experiment.

Schedule, September 29, 2015:

- 10:00 am: start / log in / instructions part 1
- 10:30 am: playing a test-round
- 10:50 am: experiment 1
- 12:15 am: lunch break
- 01:15 pm: log in / instructions part 2
- 01:30 pm: experiment 2
- 03:00 pm: questionnaire / feedback
- 03:10 pm: end

At this session we will conduct 2 experiments. One in the morning and one in the afternoon. Please view the tutorial videos and instructions located in the help section of this platform until 10:30 am, to be prepared for the experiments.

At 10:30 am we will start a test round. In this round you will have 15 minutes time to get used to the experimental platform. Doing so you will be able to realize (hypothetical) "Monopoly-building-projects". You can test all functions of the platform during the test-round. This test-round will not be used in the analysis of the results, however we want to strongly encourage you to use this test-round to be more secure during the actual experiments.

At 10:50 am we will start the first experiment. There you will find all your (real) building projects which you entered in the survey census platform several weeks ago.

The experiments last until 12:15 am. Eventually we wil finish a bit early. After the first experiment a lunch break will take place.

Please come back to your desk until 01:15 pm, so that you have enough time to read the instructions for the second experiment. Here you can view again a short tutorial video or

read the text form of the instructions. These instructions will explain you the differences to the first experiment.

At 01:30 pm we will start the second experiment.

At 03:00 pm the second experiment is finished. Then you will have time to fill out a short questionnaire to give us feedback. After finishing this questionnaire you are done for today.

The team of the “Planspiel Flächenhandel” wants to thank you for your participation!

If you have additional questions feel free to contact us:

Tim Straub
“e-mail”
“phone”

Dr. Ralph Henger
“e-mail”
“phone”

Instruction

Dear participant of the “Planspiel Flächenhandel”,
this document describes the online platform of this planning game.

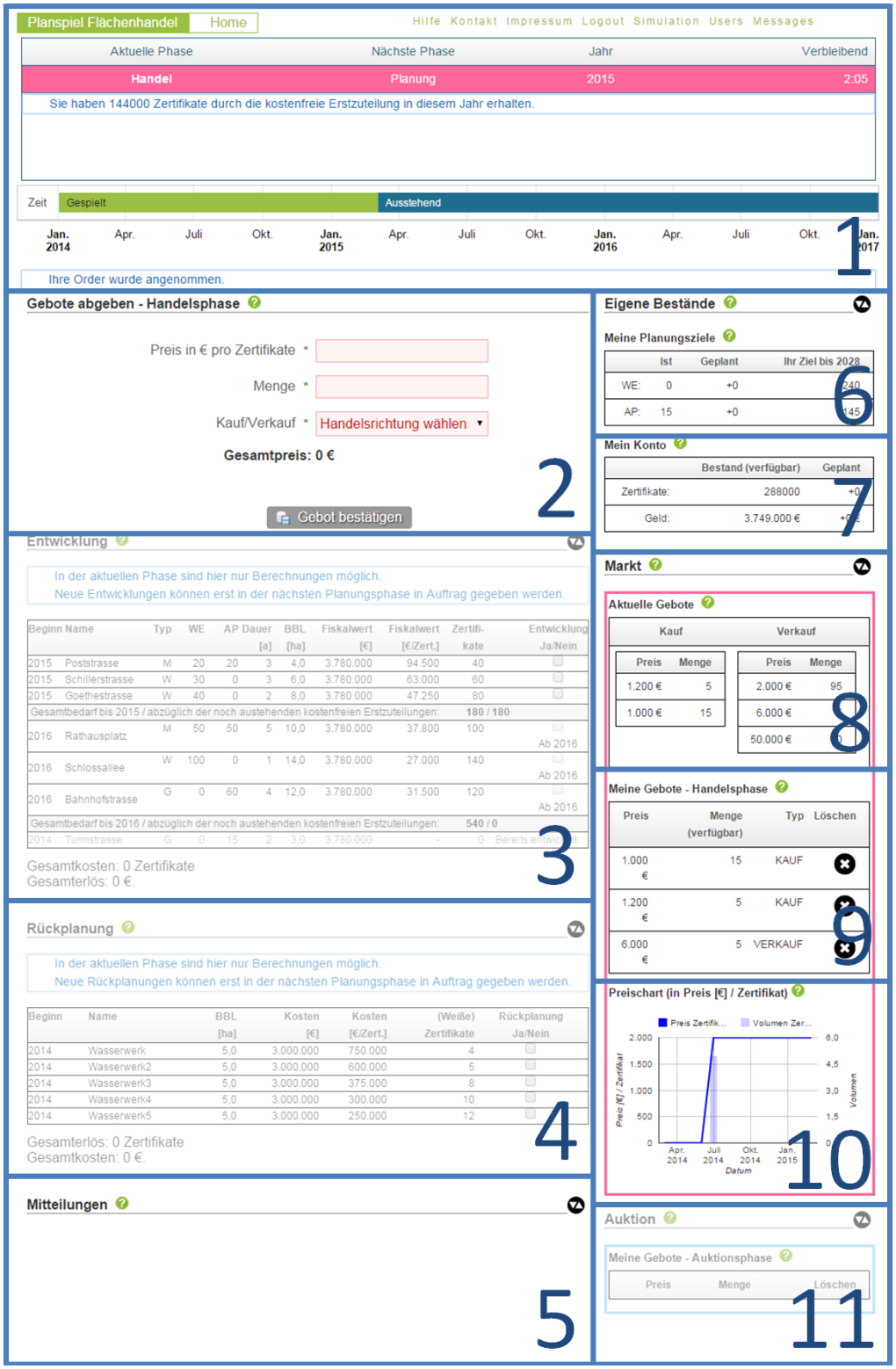
Please consider the following central points:

- You are taking the role of a planning department of a German local authority.
- Your actions during the simulation are supervised by different institutions. These institutions will question your actions and when the circumstances require it restrict your actions. This regards your bids as your possibility to take out credit loans. Information regarding this will be presented via the header of the platform as well as through pop-up windows.

We would like to highlight, that you participate in this experiment **as representative of a German local authority and you should act realistically**. Take all displayed building land as given and ask yourself before taking an action, how the real local authority would decide, given the circumstances you face during the experiment.

Overview of the platform:

We would like to give you a short overview of the structure of the platform. This should make it easier for you to find the several elements of the platform. You will find detailed explanations of these elements below.


Planspiel Flächenhandel
Home
Hilfe Kontakt Impressum Logout Simulation Users Messages
Aktuelle Phase
Nächste Phase
Jahr
Verbleibend
Handel
Planung
2015
2:05
Sie haben 144000 Zertifikate durch die kostenfreie Erstzuteilung in diesem Jahr erhalten.
Zeit
Gespielt
Ausstehend
Jan. 2014
Apr.
Juli
Okt.
Jan. 2015
Apr.
Juli
Okt.
Jan. 2016
Apr.
Juli
Okt.
Jan. 2017
1
Ihre Order wurde angenommen.
Gebote abgeben - Handelsphase
Preis in € pro Zertifikate *
Menge *
Kauf/Verkauf *
Handelsrichtung wählen
Gesamtpreis: 0 €
Gebot bestätigen
2
Eigene Bestände
Meine Planungsziele
Ist
Geplant
Ihr Ziel bis 2028
WE: 0 +0 240
AP: 15 +0 145
6
Mein Konto
Bestand (verfügbar)
Geplant
Zertifikate: 288000 +0
Geld: 3.749.000 €
7
Entwicklung
In der aktuellen Phase sind hier nur Berechnungen möglich.
Neue Entwicklungen können erst in der nächsten Planungsphase in Auftrag gegeben werden.
Beginn Name Typ WE AP Dauer [a] BBL [ha] Fiskalwert [€] Fiskalwert [€/Zert.] Zertifikate Entwicklung Ja/Nein
2015 Poststrasse M 20 20 3 4,0 3.780.000 94.500 40
2015 Schillerstrasse W 30 0 3 6,0 3.780.000 63.000 60
2015 Goethestrasse W 40 0 2 8,0 3.780.000 47.250 80
Gesamtbedarf bis 2015 / abzüglich der noch austehenden kostenfreien Erstzuteilungen: 180 / 180
2016 Rathausplatz M 50 50 5 10,0 3.780.000 37.800 100 Ab 2016
2016 Schlossallee W 100 0 1 14,0 3.780.000 27.000 140 Ab 2016
2016 Bahnhofstrasse G 0 60 4 12,0 3.780.000 31.500 120 Ab 2016
Gesamtbedarf bis 2016 / abzüglich der noch austehenden kostenfreien Erstzuteilungen: 540 / 0
2014 Turmstrasse G 0 15 2 3,0 3.780.000 - 0 Bereits enthalten
Gesamtkosten: 0 Zertifikate
Gesamterlös: 0 €.
3
Markt
Aktuelle Gebote
Kauf
Verkauf
Preis Menge
1.200 € 5
1.000 € 15
Preis Menge
2.000 € 95
6.000 €
50.000 €
8
Meine Gebote - Handelsphase
Preis Menge (verfügbar) Typ Löschen
1.000 € 15 KAUF
1.200 € 5 KAUF
6.000 € 5 VERKAUF
9
Rückplanung
In der aktuellen Phase sind hier nur Berechnungen möglich.
Neue Rückplanungen können erst in der nächsten Planungsphase in Auftrag gegeben werden.
Beginn Name BBL [ha] Kosten [€] Kosten [€/Zert.] (Weiße) Zertifikate Rückplanung Ja/Nein
2014 Wasserwerk 5,0 3.000.000 750.000 4
2014 Wasserwerk2 5,0 3.000.000 600.000 5
2014 Wasserwerk3 5,0 3.000.000 375.000 8
2014 Wasserwerk4 5,0 3.000.000 300.000 10
2014 Wasserwerk5 5,0 3.000.000 250.000 12
Gesamterlös: 0 Zertifikate
Gesamtkosten: 0 €.
4
Preischart (in Preis [€] / Zertifikat)
Preis Zertifik...
Volumen Zer...
Preis [€] / Zertifikat
Volumen
2.000
1.500
1.000
500
0
6,0
4,5
3,0
1,5
0
Apr. 2014
Juli 2014
Okt. 2014
Jan. 2015
Datum
10
Mitteilungen
5
Auktion
Meine Gebote - Auktionsphase
Preis Menge Löschen
11

1. In the **header** of the page you find an overview of the course of the experiment – which phase is currently active, which will be the next phase, which year it is, and how much time is left in the current phase. Additionally you will get short update messages about the experiment.
2. In the **trading module** you can submit buy and sell orders for certificates. Therefore, you type in the price (in €) per certificate, the amount of certificates, as well as if you want to buy or sell.
3. The **building projects module** shows you all your potential building projects. Only during the "activation phase" you can realize them. You need enough certificates doing so. In all other phases the building projects are slightly grayed out.
4. The **depletion projects module** shows you all your depletion projects (white certificates projects). Only during the "activation phase" you can deplete them. You will gain white certificates doing so.
5. The **message module** will give detailed messages about the experiment. These are mostly similar to the messages of the header, but supplemented with more details.
6. The **planning goals module** informs you about your currently created working and living places and how many of them your local authority planned to realize until the end of the experiment (end of 2028).
7. The **depot module** informs you about your current certificates and money.
8. The **orderbook module** shows you the currently open orders at the certificate market. This shows you at any given point at which price you can buy or sell certificates.
9. The **own orders module** shows you your own orders, which you created during the "CDA trading phases". You can cancel orders which are not yet matched, which will remove them from the market.
10. The **price chart module** informs you about the course of certificate prices.
11. The **own auction orders module** shows you your orders from the "USBA auction phases". You can cancel orders which are not yet matched during an active "USBA auction phase", which will remove them from the market.

1. Experiment:

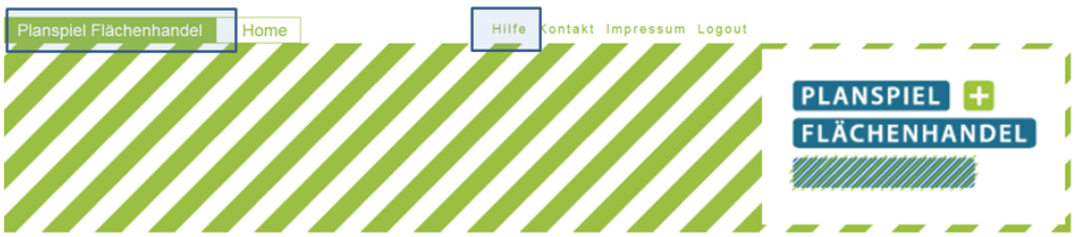
Video Tutorial:

The video tutorial is accessible here: `https://www.youtube.com/watch?v=VM6B4TG9u9c`[1]

.....

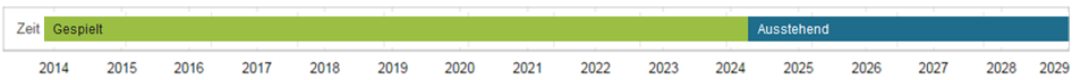
General (video 00:00 min):

After you logged in into the platform, you will first see the landing page. You can always come back to the landing page by clicking in the left upper corner on "Planspiel Flächenhandel". During the experiment, you can come back to this instruction page by clicking on "help".



At first the experiment will not be active. You will be automatically redirected to the experiment, once it starts.

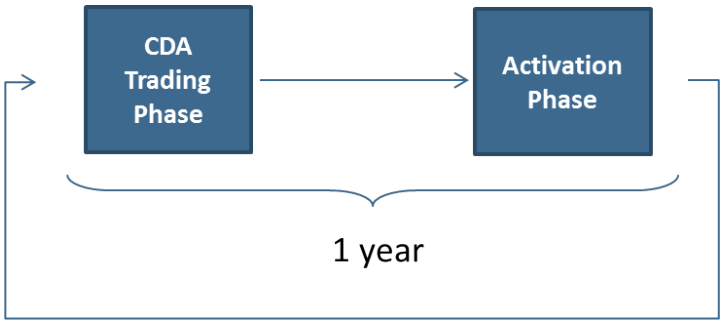
The experiment simulates the period of the years 2014 to 2028.



A year consists at first of 2 phases: The first phase comprises the trading of certificates; the second phase the realization and depletion of building projects. The phases repeat yearly until the end of the experiment. The phases last 3 minutes ("CDA treading phase")

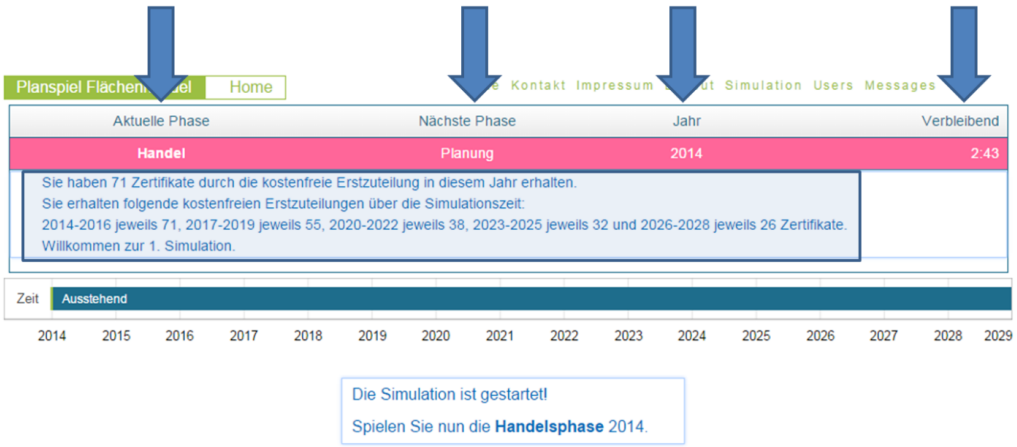
[1] Last accessed: September 6, 2016.

and 2 minutes ("activation phase"). A certificate accounts for 0.1 ha (1000 square meters) building land. To realize building projects during the "activation phase" you need enough certificates. You get certificates at the beginning of each year for free and you can buy them during the "CDA trading phase" from other participants.



Throughout the experiment you are informed in the upper part of the platform (header), which year and phase is currently active. Moreover, it shows how much time is left in the current phase and which will be the next phase.

The next line shows you, how many certificates you got for free in the current year. Each year you will get certificates for free dependent on the size of the local authority you represent. This yearly freely distributed amount will be shortened every 3 years. To give you an overview of this process we inform you about the outlook of your own certificate distribution in the first year as information message in the header. Moreover we inform you if any complications occur and we have to pause the experiment. The different phases are color coded to make it easier for you to identify which phase is currently active – "CDA trading phase" pink, "activation phase" white, and "USBA auction phase" blue.



CDA trading phase (video 05:55 min):

The experiment starts with the "CDA trading phase", which lasts 3 minutes. Here you can buy and sell certificates from or to other participants. The auction mechanism is a continuous double auction (CDA).

If you point your mouse over the small green question mark symbols you can access short explanations of the different modules, which should clarify most questions throughout the experiment.

During the "CDA trading phase" you see the "trading module" in the upper part of the screen. During the other phases this module is hidden. If you want to trade you have to submit buy or sell orders with this module. Once you have confirmed and submitted your order it will be available anonymized within the "orderbook module" for all other participants.

During the phase you can submit as many orders as you want to. Submitting orders is free of charge. However, you have to pay the price for the certificates if you buy them or have to provide the certificates if you sell them, once your order is matched. A order gets matched if respective other order exists.

You should take into account that the price of your order is in relation to your building project's fiscally evaluation, so that you not lose money. The fiscally evaluation per certificate of your building projects' list helps you in this case. You find these evaluations in the "building projects module". We will later explain this module in more detail.

High spendings during the "CDA trading phase" can lead to the point, that your local authority has to loan to much money (credits). If your order is unrealistic, given your

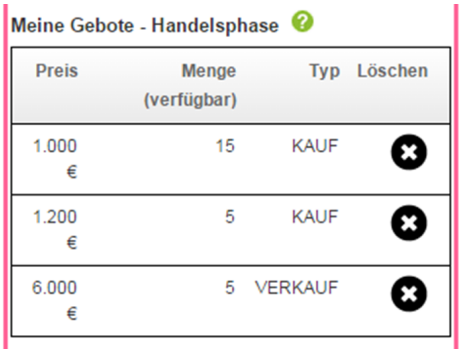
local authorities budget and overall market caps, your actions can be questioned by several institutions or even be prohibited. If you act to often against the intuitions of your local authority, you may even lose your right to take part in the experiment. As orientation you can see your yearly maximal budget in the header of the platform as short information message.

The "orderbook module" is structured as follows:

- On the left side you see the best 5 buy orders, sorted from the highest to the lowest price.
- On the right side you see the best 5 sell orders, sorted from the lowest to the highest price.

Orders are matched following the continuous double auction (CDA) mechanism. The execution price is based on the older order in the market.

Not yet matched orders can be canceled in the "own orders module".

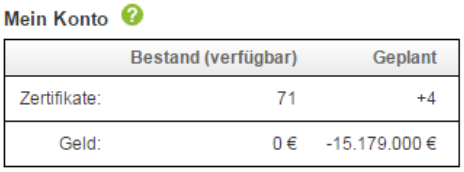


An overview of your current financial values and your current certificates is given in the "depot module". At the beginning of each year you get certificates for free on your depot.

If you have active orders, which are currently not yet matched, then the amount of money or certificates which you need for your order is blocked and not accessible for you. Your money depot starts with 0€.

You can overdraw your money depot yearly to a certain extend. The municipal supervisory will take action and might even forbid an action, if it would overdraw your money depot over your yearly limit or if you would need to borrow the money from other departments within your local authority. Please note, that your income during a year will counteract to this, so that your financial scope will get bigger again. As orientation we will tell you your yearly budget as short information message in the header.

Please note, that you can already select building projects during the “CDA trading phase”. Doing so, you can see a preview of the impact on your depot in the “depot module”. This can help you with the decision if you should buy or sell certificates.


Mein Konto

	Bestand (verfügbar)	Geplant
Zertifikate:	71	+4
Geld:	0 €	-15.179.000 €

Activation phase (video 15:30 min):

After the “CDA trading phase” the “activation phase” is active. During this phase you can realize or deplete building projects. This phase lasts 2 minutes.

For the “activation phase” the modules “building projects module” and “depletion projects module” are important. During the phase all other modules are slightly grayed out. In these modules you will find a list of your building projects or your depletion projects (white certificates). Depletion projects cost money to deplete but create certificates (white certificates) over the following years.

The “building project module” shows the following informations in a aggregated manner: the name of the project, the type (commercial or living), the amount of created working and living places, as well as the expected duration until the projects is fully used, the space of the project in ha, the fiscally evaluation of the project, the fiscally evaluation per certificate needed to realize the project, and the amount of certificates needed to realize

the project. You can only realize a project if you have enough certificates in your depot. By doing so you will get the fiscally evaluation booked on your depot.

Entwicklung

Beginn	Name	Typ	WE	AP	Dauer [a]	BBL [ha]	Fiskalwert [€]	Fiskalwert [€/Zert.]	Zertifikate	Entwicklung Ja/Nein
2014	Turmstrasse	G	0	15	2	3,0	3.780.000	-	0	☑
Gesamtbedarf bis 2014 / abzüglich der noch austehenden kostenfreien Erstzuteilungen:									**0 / 0**	
2015	Schillerstrasse	W	30	0	3	6,0	3.780.000	63.000	60	☑
2015	Goethestrasse	W	40	0	2	8,0	3.780.000	47.250	80	☐
2015	Poststrasse	M	20	20	3	4,0	3.780.000	94.500	40	☐
Gesamtbedarf bis 2015 / abzüglich der noch austehenden kostenfreien Erstzuteilungen:									**180 / 180**	
2016	Schlossallee	W	100	0	1	14,0	3.780.000	27.000	140	☐ Ab 2016
2016	Rathausplatz	M	50	50	5	10,0	3.780.000	37.800	100	☐ Ab 2016
2016	Bahnhofstrasse	G	0	60	4	12,0	3.780.000	31.500	120	☐ Ab 2016
Gesamtbedarf bis 2016 / abzüglich der noch austehenden kostenfreien Erstzuteilungen:									**540 / 0**	

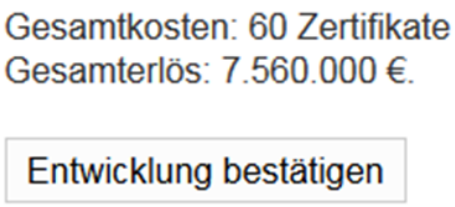
Gesamtkosten: 60 Zertifikate
Gesamterlös: 7.560.000 €.

Entwicklung bestätigen

The building projects are sorted following their potential first realization year. Below the list of all projects of one year, you can see a summary row, showing the total certificates needed to realize all projects. The projects from the current year are included in this summary calculation. Projects first realizable in later years are not included in this calculation. The second number of this summary row subtracts all freely distributed certificates you would get until the respective year. E.g., we currently have the year 2020. The building projects we want to realize cannot be realized before 2022. Then the second number shows the amount of certificates you need additionally for the realization, given your current certificates depot and the freely distributed certificates from the years 2021 and 2022. Hence, you can easily identify how many additional certificates you need to buy to realize all building projects.

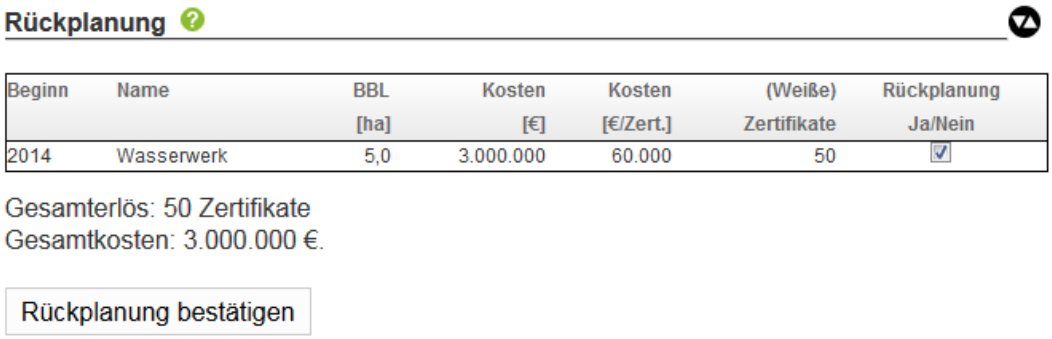
To realize a building project you have to check mark the respective box in the row of the project. In each “activation phase” you can realize as many projects as you want to. You can see the impact on your depot and your overall goals in the “depot module” and the “planning goals module”. For the realization you need to confirm your selection by clicking

on the “confirm” button below the table. If you have enough certificates the action will be accepted. These certificates will be subtracted from your depot. The fiscally evaluation of the realized building projects will be booked on your money depot as well. Note that not all building projects have a positive fiscally evaluation.



After clicking on the confirm button the selected building projects will appear at the end of the list with the comment “realized”. Furthermore, your “planning goals module” and your “depot module” will be updated.

Besides the realization of building projects you can as well deplete projects (“renaturation”) during an “activation phase”. Thereby you generate so called white certificates. You will find the respective module below the “building projects module”.



If you decide to deplete projects, you will get the displayed certificates distributed over several years on your “depot module”. Depletion projects do have costs, which will be subtracted from your “depot module”. Like for the case of the realization of building projects you have to confirm your action by clicking on the confirm button below the list. The maximal amount of white certificates you will get each year is aligned to the size of the local authority you represent, similar to the free distribution of certificates each year. You can see the maximal yearly amount of white certificates by moving your mouse over the small green question mark of the “depletion projects module”. Hence, you will not get all certificates instantly, but stretched over several years.

2. Experiment:

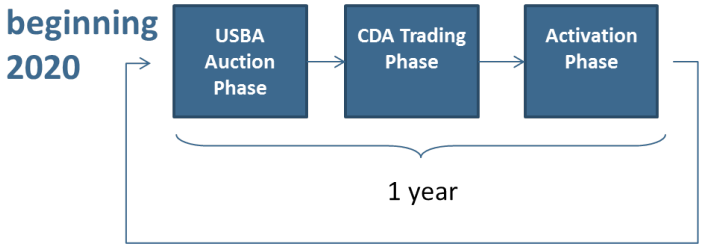
Dear participant of the “Planspiel Flächenhandel”,
this document describes the extensions and changes from the second experiment compared to the first experiment you played this morning.

Video Tutorial:

The video tutorial is accessible here: `https://www.youtube.com/watch?v=X39TZae25Wc`[2]

USBA Auction Phase (video 00:00 min):

The second experiment is overall very similar to the first experiment. Like this morning we will again simulate the period from 2014 to 2028. The big difference is, that beginning with the year 2020 we will introduce the “USBA auction phase”. During this phase the yearly freely distributed certificates will be partially auctioned by the public administrations.



For you that means, that besides the “activation phase” and the “CDA trading phase” there will be a third phase, the so called “USBA auction phase”. Each year this will be the first phase, taking place before the other two phases. To understand this new phase we will first look at the changes to your freely distributed certificates each year. Different than before, beginning with the year 2020, you will get stepwise less certificates for free. In a

[2]Last accessed: April 6, 2016.

first period from 2020 to 2022 you will receive 20 % less certificates for free than in the first experiment. During the period from 2023-2025 you will get 40 % less and during the period from 2026-2028 you will get 60 % less. Now you will probably ask yourself what we will do with these not distributed certificates.

Thats where the USBA auction comes into play. You will get the chance to auction those not-distributed certificates (certificates from all participants). In reality such an auction would be conducted by the public administration, similar to the emission trading. The "USBA auction phase" lasts 1:30 minutes and all participants take part. To make the process more clear, we will guide you exemplarily through it.

We will tell you that an auction will happen through a short information message in the header. The phase is color coded with blue, so that you can easily distinguish it from the other phases.

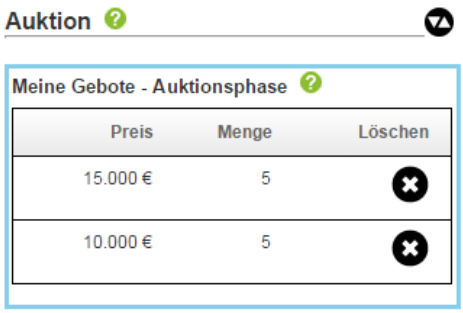
During the "USBA auction phase" the "auction module" will be active, which looks similar to the "trading module".

Similar to the "CDA trading phase" you will submit price per certificate and amount of certificates as an order. In our example the participant is willing to pay 15,000€ per certificate. He wants to buy 5 certificates for this price. Hence, this results in a buying volume of 75,000€. You can submit as many orders during this phase as you want to.

To not lose money, you should take into account that the price of your order is in relation to your building project's fiscally evaluation. If you would lose money several institutions could prevent your actions as well. The fiscally evaluation per certificate of your building projects lists helps you in this case. You find these evaluations in the "building projects module". The spendings during this phase will as well affect your "depot module" and hence your credit limit. As orientation we will tell you your maximal budget as short information message in the header.

Your order will be submitted by clicking on the confirmation button. The auction will be sealed, hence you will not see the orders from the other participants.

During the "USBA auction phase" you can cancel your orders in the blue highlighted "own auction orders module".

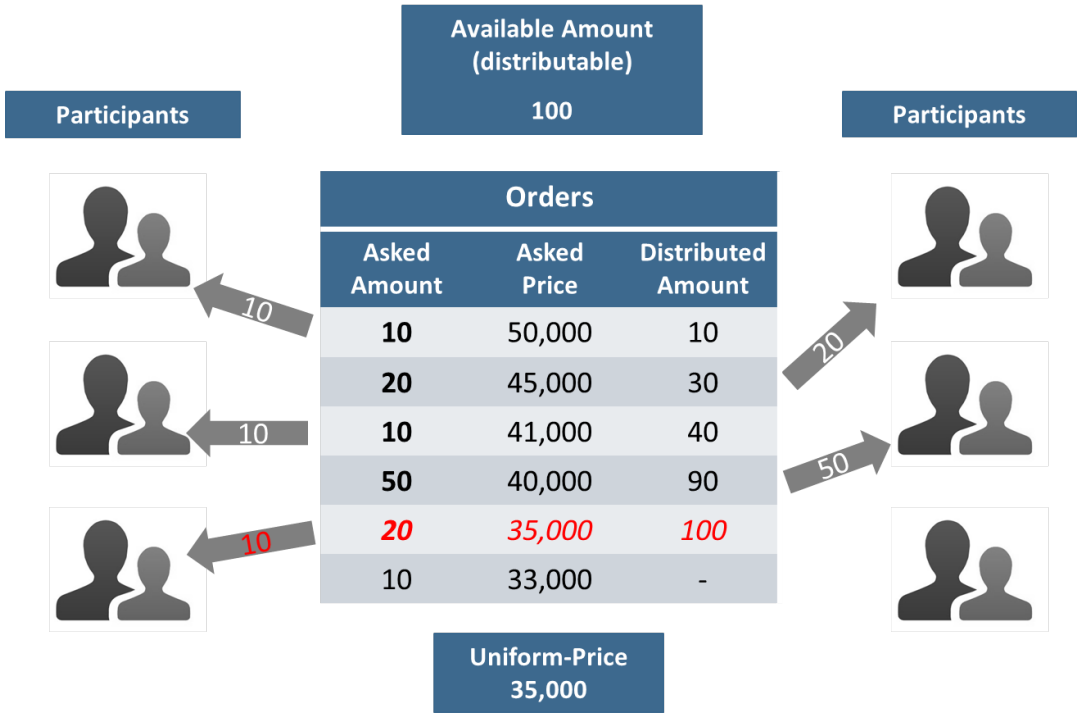


Explanation of a Uniform-Price Auction (video 05:35 min):

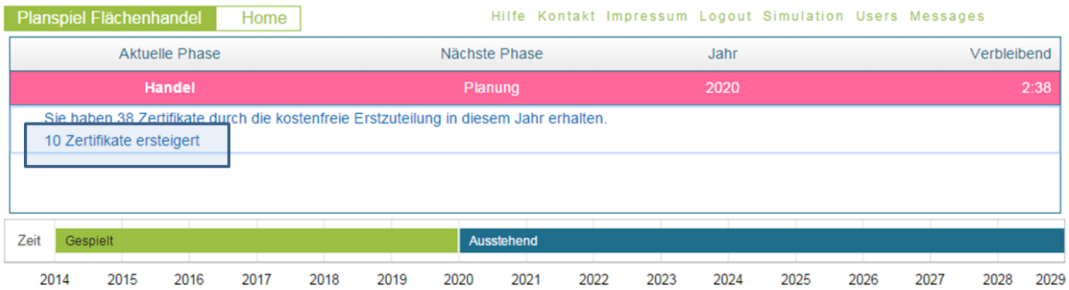
Different than the "CDA trading phase" the auction follows a uniform-price scheme. E.g. overall 100 certificates will be auctioned. During the "USBA auction phase" all participants submit their orders. After the phase, when all orders are submitted, they are sorted from the highest to the lowest price. Then the certificates will be distributed from top to bottom (order with highest price to order with lowest price). This happens as long as the orders can be executed, meaning as long as certificates are available for distribution. It is possible that the last executed order does not get all certificates he initially wanted to auction. In that case he gets the rest of the still available certificates. The last executed order (the

order who got certificates with the lowest price), defines the price for all participants – the so called uniform-price.

In our example, where 100 certificates are auctioned, the participant who wants to auction 20 certificates and at a price of 35,000€ only gets 10 certificates, since all other certificates are already distributed to the other participants with orders including higher prices. However, since his order is the last one executed, it defines the uniform-price. Hence, all participants pay per certificate 35,000€, even when they were willing to pay more per certificate initially.



At the end of the "USBA auction phase" you can see a short information message in the header, how many certificates you auctioned and at which price.



All other yearly following phases will proceed analogous to experiment 1.

If you have any questions during the experiments, you can always come back to the instructions by clicking on help at the top of the platform.

We wish you good luck during the two experiments!

Information Papers from “Planspiel Flächenhandel

All papers are in German language. They are accessible from the project's homepage[3] as well.

- Information paper 1, Information:
  http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Fl%C3%A4chenhandel-InfoPapier-Nr01_Fl%C3%A4chenhandel_Informationspapier_Nutzen_und_Anforderungen_f%C3%BCr_Kommunen.pdf
- Information paper 2, certificate distribution:
  http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Fl%C3%A4chenhandel-InfoPapier-Nr02_NEU.pdf
- Information paper 3, inner city development:
  http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Fl%C3%A4chenhandel-InfoPapier-Nr03-NEU.pdf
- Information paper 4, survey census platform:
  http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Planspiel_Fl%C3%A4chenhandel_Erhebungsplattform_Anleitung.pdf
- Information paper 5, white certificates:
  http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Fl%C3%A4chenhandel-InfoPapier-Nr05.pdf
- Information paper 6, field experiment:
  http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Fl%C3%A4chenhandel-InfoPapier-Nr06.pdf
- Information paper 7, fiscally evaluation:
  http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Fl%C3%A4chenhandel_Infopapier_Nr._07_-_Fiskalische_Wirkungsanalyse.pdf

[3] http://www.flaechenhandel.de/ last accessed: September 6, 2016.

- Information paper 8, results of case studies:
  `http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Fl%C3%A4chenhandel-Info-Papier-Nr.08.pdf`

Platform Screens

FIGURE C.1: *Homepage of the project "Planspiel Flächenhandel" of Chapter 6.*

FIGURE C.2: *Participating local authorities of the project "Planspiel Flächenhandel" of Chapter 6.*

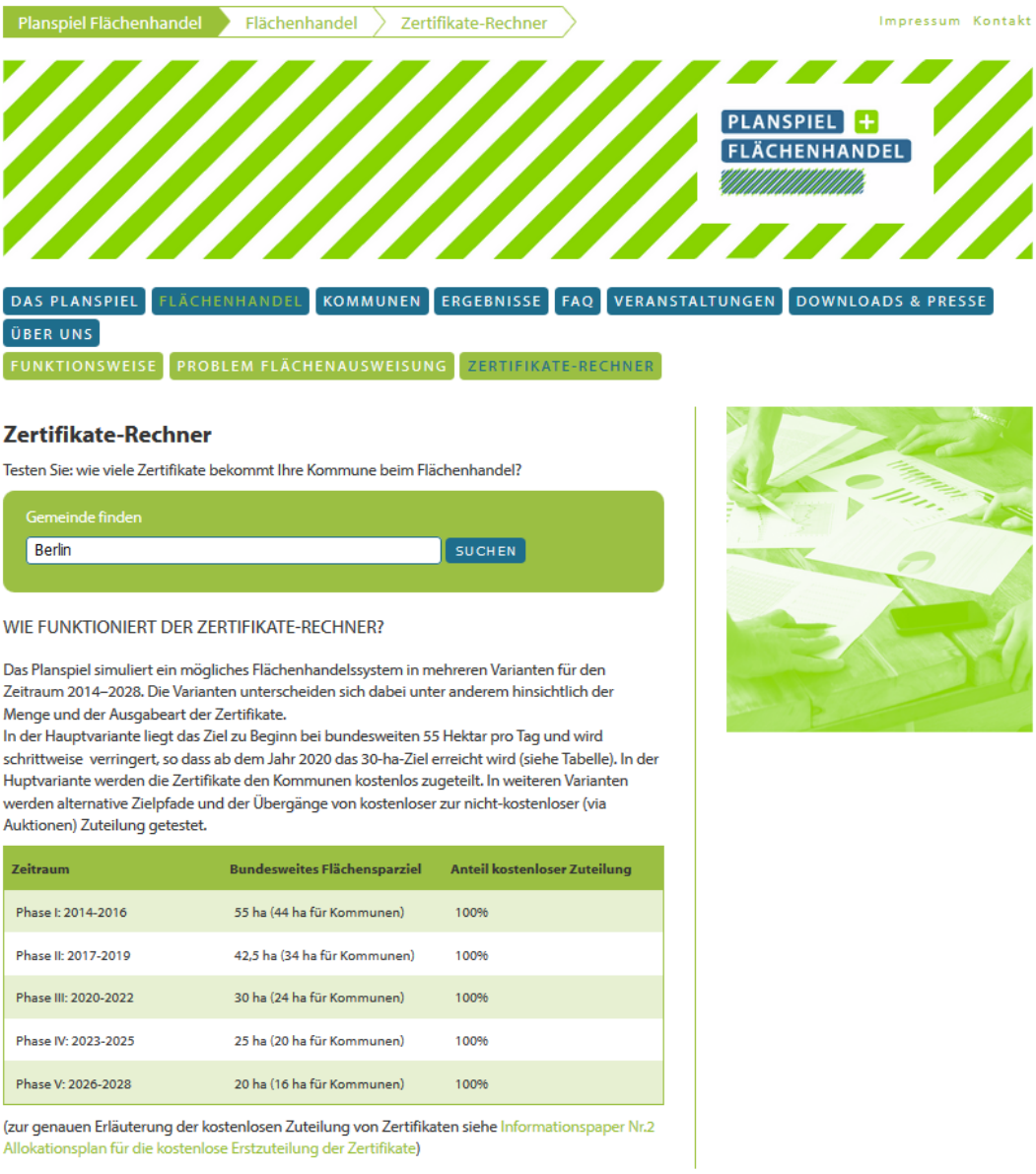


FIGURE C.3: *Certificate calculator (1) of the project "Planspiel Flächenhandel" of Chapter 6 – searching for a local authority (here Berlin, city).*

Zertifikate-Rechner

Allgemeine Informationen

Gemeindename	Berlin, Stadt
PLZ	10178
Einwohner	3.375.222

Phasen I - V

Menge kostenlos zugeteilter Flächen-Zertifikate pro Jahr (ein Zertifikat entspricht 1.000 m²):				
Phase I: 2014-2016: **1811** Zertifikat(e)	Phase II: 2017-2019: **1397** Zertifikat(e)	Phase III: 2020-2022: **985** Zertifikat(e)	Phase IV: 2023-2025: **817** Zertifikat(e)	Phase V: 2026-2028: **652** Zertifikat(e)

Jährliche Zunahme der Siedlungs- und Verkehrsfläche laut amtlicher Statistik

zwischen 2004 und 2008	zwischen 2008 und 2012
98,5 ha/a	52 ha/a

Information: Die Berechnung erfolgte auf Basis der Kommunalstatistik 31.12.2012 (mit 11.328 Städten und Gemeinden) und dem Bevölkerungstand vom 31.12.2012. Bei der Angabe einer Bruchzahl 1/3 bzw. 2/3 erhält eine Kommune in einer Phase (Dauer 3 Jahre), entweder ein bzw. zwei Zertifikate in der gesamten Phase.
Negative Siedlungs- und Verkehrsflächenzunahme: In einigen Fällen werden negative Zuwächse, also Rückgänge an Siedlungs- und Verkehrsflächen Flächen ausgewiesen. Diese sind in der Regel nicht auf tatsächlich stattgefundene Rückbaumaßnahmen, sondern auf Änderungen der Datenerhebung zur Statistik zurückzuführen. Für eine detaillierte Stellungnahme hat das Statistischen Bundesamt eine Beschreibung der Datenerhebung veröffentlicht. Vor allem Unterpunkt 6 ist hier hilfreich (Link):
"In [...] Bereichen von Verkehrsinfrastrukturprojekten und Flurbereinigungsmaßnahmen sowie im Zusammenhang mit Klassifikationsänderungen [...] und Defiziten im Hinblick auf die Vollständigkeit der Daten [...] ist die zeitliche und/oder räumliche Vergleichbarkeit der Daten eingeschränkt."

ZURÜCK ZUM ZERTIFIKATE-RECHNER

FIGURE C.4: *Certificate calculator (2) of the project "Planspiel Flächenhandel" of Chapter 6 – certificates distributed to a local authority (here Berlin, city).*

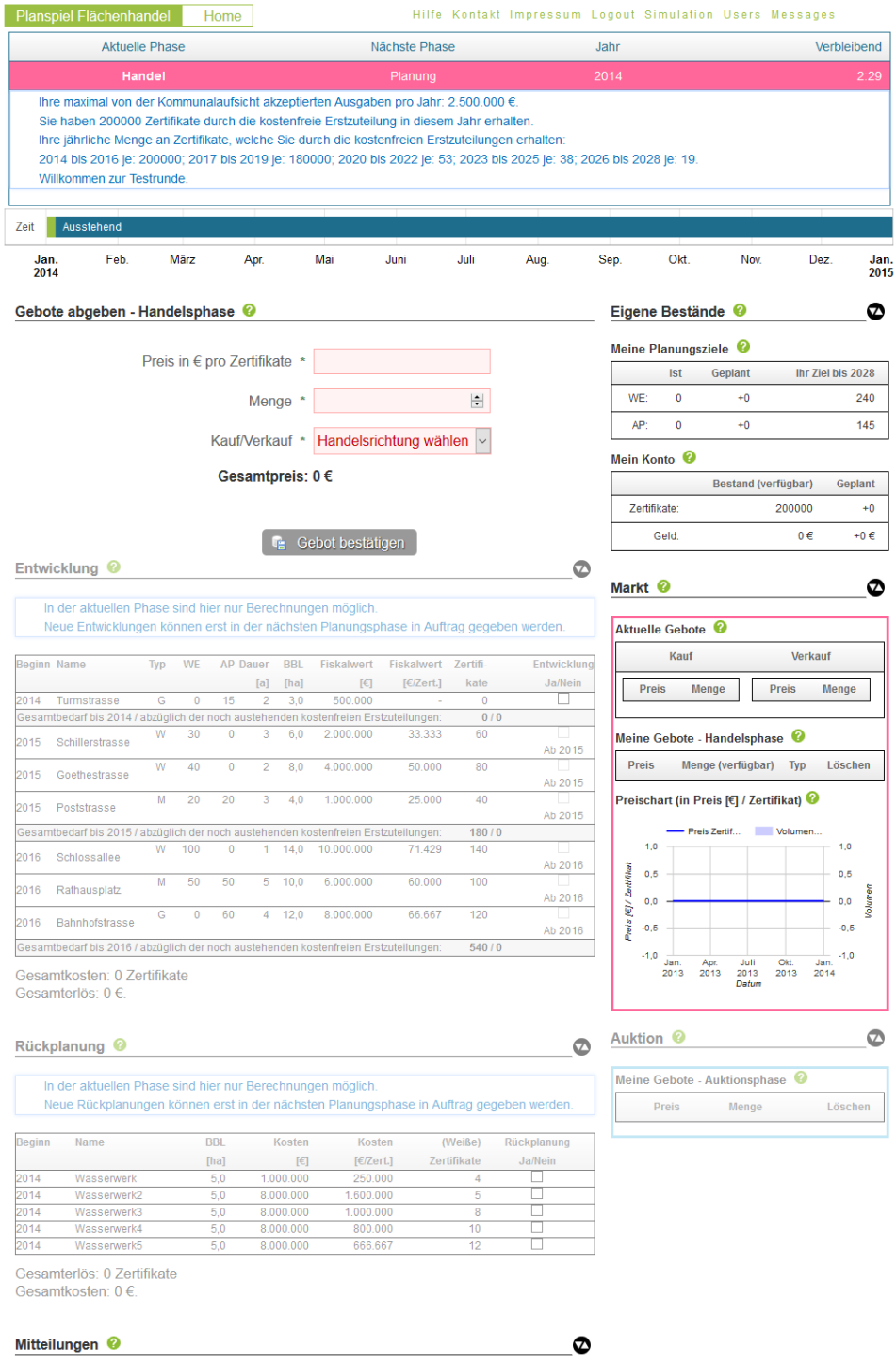


FIGURE C.5: *Experiment platform of the project "Planspiel Flächenhandel" of Chapter 6 – active CDA trading phase.*

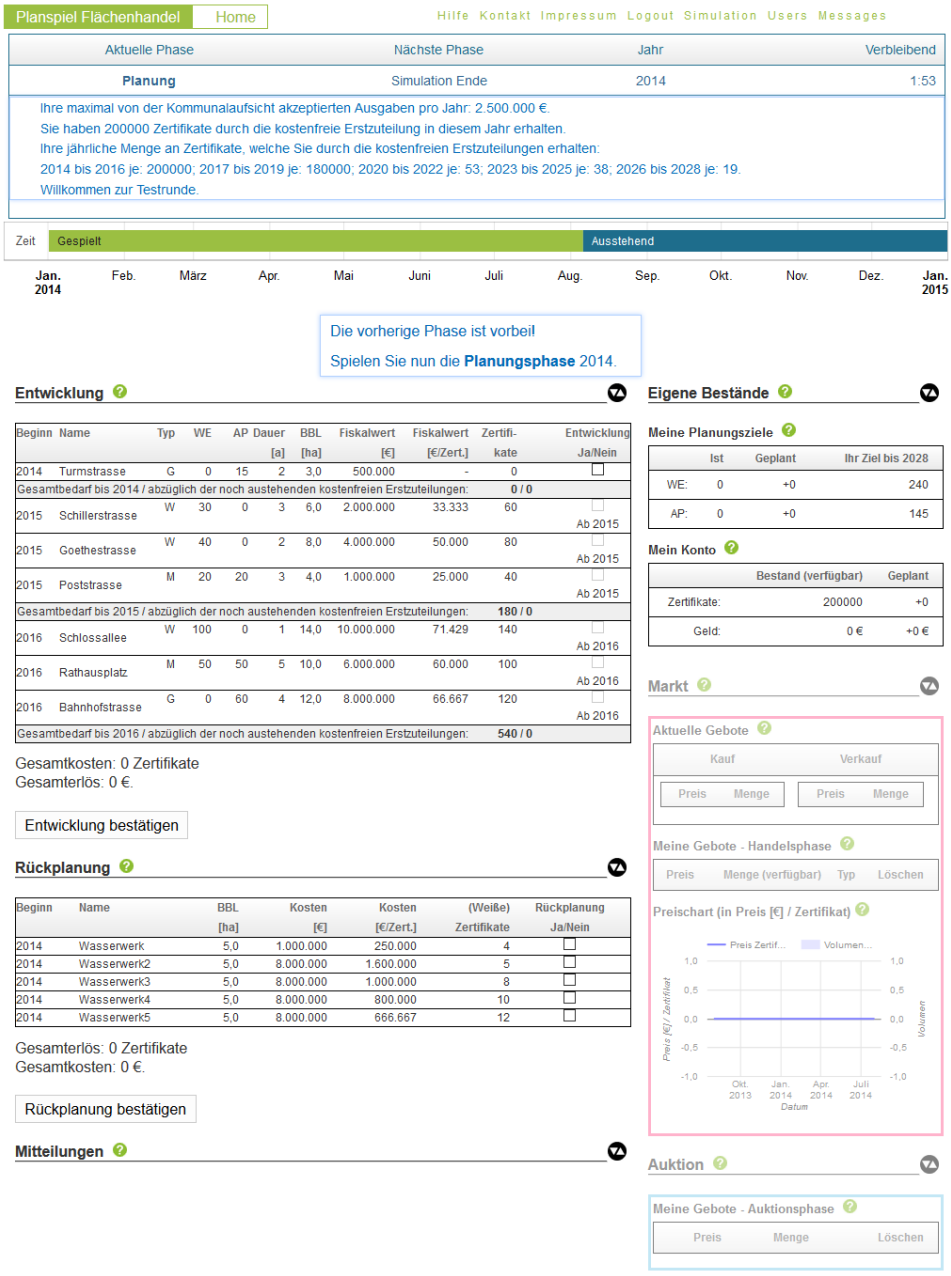


FIGURE C.6: *Experiment platform of the project "Planspiel Flächenhandel" of Chapter 6 – active activation phase.*

Questionnaire

1. How strongly did the following factors impact your buy orders?[4]
    - Price on the market
    - Political importance
    - Fiscally evaluation
    - Demand (long-term)
    - Demand (mid-term)
    - Demand (short-term)
2. How strongly did the following factors impact your sell orders?
    - Price on the market
    - Political importance
    - Fiscally evaluation
    - Demand (long-term)
    - Demand (mid-term)
    - Demand (short-term)
3. To which extend do you think your actions correspond to real decisions from your local authority?
4. Here you can give us feedback:

Please insert your feedback in this textfield.

[4]Questions are constructed mainly by Dr. Ralph Henger and Michael Schier from the "Institut der deutschen Wirtschaft Köln". All questions were only answered from planners from local authorities (session 1 and 2). Students did not answer questions (session 3 and 4). The following questions were implemented with a scale ranging from 0 % to a 100 %.

Bonus Payment Structure

Rank	Bonus-Payment
1	21.00€
2	21.00€
3	21.00€
4	20.00€
5	20.00€
6	20.00€
7	19.00€
8	19.00€
9	19.00€
10	18.00€
11	18.00€
12	18.00€
13	17.00€
14	17.00€
15	17.00€
16	16.50€
17	16.50€
18	16.50€
19	15.50€
20	15.50€
21	15.50€
22	14.50€
23	14.50€
24	14.50€
25	13.50€
26	13.50€
27	13.50€
28	12.50€
29	12.50€
30	12.50€
31	12.00€
32	12.00€
33	12.00€
34	11.00€
35	11.00€
36	11.00€
37	10.00€
38	10.00€

TABLE C.1: *Bonus payment of student participants in the Mo setting.*

References

Adamic, L. A., J. Zhang, E. Bakshy, and M. S. Ackerman (2008). Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proceedings of the 17th International Conference on World Wide Web*, pp. 665–674. ACM.

Allahbakhsh, M., B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar (2013). Quality Control in Crowdsourcing Systems: Issues and Directions. *IEEE Internet Computing 17*(2), 76–81.

Amabile, T. M., B. A. Hennessey, and B. S. Grossman (1986). Social Influences on Creativity: The Effects of Contracted-for Reward. *Journal of Personality and Social Psychology 50*(1), 14–23.

Amabile, T. M., K. G. Hill, B. A. Hennessey, and E. M. Tighe (1994). The Work Preference Inventory: Assessing Intrinsic and Extrinsic Motivational Orientations. *Journal of Personality and Social Psychology 66*(5), 950–967.

Amir, O., D. G. Rand, and Y. K. Gal (2012). Economic Games on the Internet: The Effect of $1 Stakes. *PLoS ONE 7*(2), 1–4.

Araujo, R. M. (2013). 99designs: An Analysis of Creative Competition in Crowdsourced Design. In *Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing*, pp. 17–24.

Archak, N. and A. Sundararajan (2009). Optimal Design of Crowdsourcing Contests. In *Proceedings of the 2009 International Conference on Information Systems (ICIS 2009)*, pp. 1–16.

Ariely, D. and J. Levav (2000). Sequential Choice in Group Settings: Taking the Road Less Traveled and Less Enjoyed. *Journal of Consumer Research 27*(3), 279–290.

Barankay, I. (2011). Rankings and Social Tournaments: Evidence from a Crowd-Sourcing Experiment. In *Wharton School of Business, University of Pennsylvania Working Paper*, pp. 1–38.

Baron, R. M. and D. A. Kenny (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Psychology 51*(6), 1173–1182.

Becker, B. E. and M. A. Huselid (1992). The Incentive Effects of Tournament Compensation Systems. *Administrative Science Quarterly 37*(2), 336–350.

Bénabou, R. and J. Tirole (2003). Intrinsic and Extrinsic Motivation. *Review of Economic Studies 70*(3), 489–520.

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Berinsky, A. J., G. A. Huber, and G. S. Lenz (2012). Evaluating Online Labor Markets for Experimental Research: Amazon's Mechanical Turk. *Political Analysis 20*(3), 351–368.

Bizer, K., R. Henger, L. Meub, and T. Proeger (2014). The Political Economy of Certificates for Land Use in Germany – Experimental Evidence. *Discussion Papers, Center for European Governance and Economic Development Research*.

Boecking, B., M. Hall, and J. Schneider (2015). Event Prediction With Learning Algorithms – A Study of Events Surrounding the Egyptian Revolution of 2011 on the Basis of Micro Blog Data. *Policy & Internet 7*(2), 159–184.

Bollen, K. A. and R. H. Hoyle (1990). Perceived Cohesion: A Conceptual and Empirical Examination. *Social Forces 69*(2), 479–504.

Bothner, M. S., J.-h. Kang, and T. E. Stuart (2007). Competitive Crowding and Risk Taking in a Tournament: Evidence from NASCAR Racing. *Administrative Science Quarterly 52*(2), 208–247.

Boudreau, M.-C., D. Gefen, and D. W. Straub (2001). Validation in Information Systems Research: A State-of-the-Art Assessment. *MIS Quarterly 25*(1), 1–16.

Brabham, D. C. (2008). Moving the Crowd at iStockphoto: The Composition of the Crowd and Motivations for Participation in a Crowdsourcing Application. *First Monday 13*(6).

Brabham, D. C. (2010). Moving the Crowd at Threadless: Motivations for Participation in a Crowdsourcing Application. *Information, Communication & Society 13*(8), 1122–1145.

Brabham, D. C. (2012). Crowdsourcing: A Model for Leveraging Online Communities. *The Participatory Cultures Handbook*, 120–129.

Bracha, A. and C. Fershtman (2013). Competitive Incentives: Working Harder or Working Smarter? *Management Science 59*(4), 771–781.

Briggs, R. O. and G. Schwabe (2011). On Expanding the Scope of Design Science in IS Research. In *Service-Oriented Perspectives in Design Science Research (DESRIST)*, pp. 92–106. Springer.

Buhrmeister, M., T. Kwang, and S. D. Gosling (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science 6*(1), 3–5.

Bull, C., A. Schotter, and K. Weigelt (1987). Tournaments and Piece Rates: An Experimental Study. *Journal of Political Economy 95*(1), 1–33.

Byun, S.-E. and M. Mann (2011). The Influence of Others The Impact of Perceived Human Crowding on Perceived Competition, Emotions, and Hedonic Shopping Value. *Clothing and Textiles Research Journal 29*(4), 284–297.

Camerer, C. F. (2004). Prospect Theory in the Wild: Evidence from the Field. *Advances in Behavioral Economics*, 148–161.

Cavallo, R. and S. Jain (2012). Efficient Crowdsourcing Contests. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, pp. 677–686. International Foundation for Autonomous Agents and Multiagent Systems.

Chandler, D. and A. Kapelner (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization 90*, 123–133.

Chandler, J., P. Mueller, and G. Paolacci (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavioral Research Methods 46*(1), 112–130.

Chesbrough, H. (2006). Open Innovation: A New Paradigm for Understanding Industrial Innovation. *Open Innovation: Researching a New Paradigm*, 1–12.

Chilton, L. B., J. J. Horton, R. C. Miller, and S. Azenkot (2010). Task Search in a Human Computation Market. In *Proceedings of the 2010 Workshop on Human Computation (HCOMP 2010)*, pp. 1–9. ACM.

Coalition Treaty (2013). Between the CDU/CSU and the SPD. `http://www.kas.de/wf/doc/kas_36853-544-2-30.pdf?`

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.

Cooper, J. (2011). Cognitive Dissonance Theory. *Handbook of Theories of Social Psychology 1*, 377–398.

Corney, J. R., C. Torres-Sánchez, A. P. Jagadeesan, and W. C. Regli (2009). Outsourcing labour to the cloud. *International Journal of Innovation and Sustainable Development 4*(4), 294–313.

Cragg, J. G. and R. S. Uhler (1970). The Demand for Automobiles. *The Canadian Journal of Economics 3*(3), 386–406.

Crump, M. J. C., J. V. McDonnell, and T. M. Gureckis (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE 8*(3), 1–18.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row.

Deci, E. L., G. Betley, J. Kahle, L. Abrams, and J. Porac (1981). When Trying to Win: Competition and Intrinsic Motivation. *Personality and Social Psychology Bulletin 7*(1), 79–83.

Deci, E. L. and R. M. Ryan (1985). Intrinsic Motivation and Self-Determination in Human Behaviour.

Descartes, R. (1637). *Discourse on the Method of Rightly Conducting One's Reason and Seeking Truth in the Sciences*.

Dewett, T. (2007). Linking intrinsic motivation, risk taking, and employee creativity in an R&D environment. *R&D Management 37*(3), 197–208.

Doan, A., R. Ramakrishnan, and A. Y. Halevy (2011). Crowdsourcing Systems on the World-Wide Web. *Communications of the ACM 54*(4), 86–96.

Donabedian, A. (1980). *Explorations in Quality Assessment and Monitoring: The Definition of Quality and Approaches to Its Assessment*, Volume 1. Ann Arbor: Health Administration Press.

Donabedian, A. (2003). *An Introduction to Quality Assurance in Health Care*. New York: Oxford University Press.

Eccles, J. S. and A. Wigfield (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology 53*(1), 109–132.

Effron, D. A., C. J. Bryan, and J. K. Murnighan (2015). Cheating at the End to Avoid Regret. *Journal of Personality and Social Psychology 109*(3), 395–414.

Ehrenberg, R. G. and M. L. Bognanno (1990). Do Tournaments Have Incentive Effects? *The Journal of Political Economy 98*(6), 1307–1324.

Eriksson, T., A. Poulsen, and M.-C. Villeval (2009). Feedback and incentives: Experimental evidence. *Labour Economics 16*, 679–688.

Eriksson, T., T. Teyssier, and M.-C. Villeval (2009). Self-Selection and the Efficiency of Tournaments. *Economic Inquiry 47*(3), 530–548.

Estellés, E. and F. González (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science 38*(2), 189–200.

Federal Government (2002). Perspectives for Germany: Our Strategy for Sustainable Development. `http://www.nachhaltigkeitsrat.de/fileadmin/user_upload/English/pdf/Perspectives_for_Germany.pdf`.

Feldmann, N. and H. Gimpel (2016). Financing Projects through Enterprise Crowdfunding: Understanding the Impact of Proposal Characteristics on Funding Success. In *Proceedings of the 24th European Conference on Information Systems (ECIS 2016)*.

Fershtman, C. and U. Gneezy (2011). The Tradeoff between Performance and Quitting in High Power Tournaments. *Journal of the European Economic Association 9*(2), 318–336.

Festinger, L. (1957). A Theory of Cognitive Dissonance.

Fischbacher, U. (2007). Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics 10*(2), 171–178.

Fraser, C., C. Gouge, and M. Billig (1971). Risky shifts, cautious shifts, and group polarization. *European Journal of Social Psychology 1*(1), 7–30.

Frey, B. S. (2012). Crowding effects on intrinsic motivation. *Renewal: A Journal of Labour Politics 20*(2/3), 91–98.

Frey, B. S. and R. Jegen (2001). Motivation Crowding Theory. *Journal of Economic Surveys 15*(5), 589–611.

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

Gefen, D. and D. W. Straub (2004). Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega 32*(6), 407–424.

Geiger, D., M. Rosemann, E. Fielt, and M. Schader (2012). Crowdsourcing Information Systems – Definition, Typology, and Design. In *Proceedings of the 2012 International Conference on Information Systems (ICIS 2012)*.

Gill, D. and V. Prowse (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review 102*(1), 469–503.

Gill, D. and V. L. Prowse (2011). A Novel Computerized Real Effort Task Based on Sliders. *Working Paper*.

Gimpel, H., N. R. Jennings, G. E. Kersten, A. Ockenfels, and C. Weinhardt (2008). Market Engineering: A Research Agenda. In *Negotiation, Auctions, and Market Engineering*, pp. 1–15. Springer.

Golder, S. A. and B. A. Huberman (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science 32*(2), 198–208.

Goodman, J. K., C. E. Cryder, and A. Cheema (2012). Data Collection in a Flat World: The Strenghts and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making 26*(3), 213–224.

Greene, S. (2004). Social Identity Theory and Party Identification. *Social Science Quarterly 85*(1), 136–153.

Greiner, B. (2004). The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics. *University of Cologne, Working Paper Series in Economics 10*(23), 63–104.

Griffin-Pierson, S. (1990). The Competitiveness Questionnaire: A Measure of Two Components of Competitiveness. *Measurement and Evaluation in Counseling and Development 23*(3), 108–15.

Grund, C. and O. Gürtler (2005). An empirical study on risk-taking in tournaments. *Applied Economics Letters 12*(8), 457–461.

Hackman, J. R. and G. R. Oldham (1976). Motivation through the Design of Work: Test of a Theory. *Organizational Behavior and Human Performance 16*(2), 250–279.

Hall, M. and S. Caton (2014). A Crowdsourcing Approach to Identify Common Method Bias and Self-Representation. In *IPP2014: Crowdsourcing for Politics and Policy*.

Hammon, L. and H. Hippner (2012). Crowdsourcing. *Business & Information Systems Engineering 4*(3), 1–4.

Harbring, C. and B. Irlenbusch (2003). An experimental study on tournament design. *Labour Economics 10*(4), 443–464.

Harrison, G. W. and J. A. List (2004). Field Experiments. *Journal of Economic Literature 42*(4), 1009–1055.

Hars, A. and S. Ou (2001). Working for Free? Motivations of Participating in Open Source Projects. In *Proceedings of the 34th Hawaii International Conference on System Sciences (HICSS 2001)*. IEEE.

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology 17*(3), 300–312.

Hassanein, K. and M. Head (2005). The Impact of Infusing Social Presence in the Web Interface: An Investigation Across Product Types. *International Journal of Electronic Commerce 10*(2), 31–55.

Hawlitschek, F., T. Teubner, and C. Weinhardt (2016). Trust in the Sharing Economy. *Die Unternehmung 70*(1), 26–44.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millenium. *Communication Monographs 76*(4), 408–420.

Henger, R., K. Bizer, L. Blecken, K. Fahrenkrug, U. Ferber, J.-M. Gutsche, T. Kranz, M. Melzer, L. Meub, T. Proeger, S. Siedentop, T. Schmidt, T. Straub, A. Tack, and C. Weinhardt (2017). Ergebnisse und Implikationen aus dem Modellversuch zum Handel mit Flächenzertifikaten. In G. Meinel, D. Förtsch, S. Schwarz, and T. Krüger (Eds.), *IÖR Schriften Nr. 69, Flächennutzungsmonitoring VIII: Flächensparen - Ökosystemleistungen - Handlungsstrategien*, pp. 11–22. Rhombos-Verlag.

Henger, R. and M. Schier (2014). Flächenhandel-Informationspapier Nr. 02: Allokationsplan für die kostenlose Erstzuteilung der Zertifikate. `http://www.flaechenhandel.de/fileadmin/std_site/content/Downloads/Fl%C3%A4chenhandel-InfoPapier-Nr02_NEU.pdf`.

Herzberg, F. (1968). One more time: How do you motivate employees. *Harvard Business Review*, 53–62.

Herzberg, F., B. Mausner, and B. B. Snyderman (2011). *The Motivation to Work*, Volume 1. Transaction Publishers.

Holt, C. A. and S. K. Laury (2002). Risk Aversion and Incentive Effects. *American Economic Review 92*(5), 1644–1655.

Hoover, H. (1930). State of the Union Address.

Horton, J. J. (2010). Online Labor Markets. In *Proceedings of the 2010 International Workshop on Internet and Network Economics*, pp. 515–522. Springer.

Horton, J. J. and L. B. Chilton (2010). The Labor Economics of Paid Crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, pp. 209–218. ACM.

Horton, J. J., D. G. Rand, and R. J. Zeckhauser (2011). The Online Laboratory: Conducting Experiments in a Real Labor Market. *Experimental Economics 14*(3), 399–425.

Hossain, M. (2012). Users' Motivation to Participate in Online Crowdsourcing Platforms. In *Proceedings of the 2012 International Conference on Innovation Management and Technology Research (ICIMTR 2012)*, pp. 310–315. IEEE.

Howe, J. (2006a). Crowdsourcing: A Definition. `http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html`.

Howe, J. (2006b). The Rise of Crowdsourcing. *Wired Magazine 14*(6), 1–4.

Huang, E., H. Zhang, D. C. Parkes, K. Z. Gajos, and Y. Chen (2010). Toward Automatic Task Design: A Progress Report. In *Proceedings of the 2010 Workshop on Human Computation (HCOMP 2010)*, pp. 77–85. ACM.

Huang, Y., P. Singh, and T. Mukhopadhyay (2012). How to Design Crowdsourcing Contest: A Structural Empirical Analysis. In *Proceedings of the 2012 Workshop of Information Systems and Economics (WISE 2012)*.

Huberman, B. A., D. M. Romero, and F. Wu (2009). Crowdsourcing, attention and productivity. *Journal of Information Science 35*(6), 758–765.

Hvide, H. K. (2002). Tournament Rewards and Risk Taking. *Journal of Labor Economics 20*(4), 877–898.

Ipeirotis, P. G. (2010a). Analyzing the Amazon Mechanical Turk Marketplace. *XRDS 17*(2), 16–21.

Ipeirotis, P. G. (2010b). Demographics of Mechanical Turk. *Technical Report, Working Paper CeDER-10-01*.

Ipeirotis, P. G., F. Provost, and J. Wang (2010). Quality Management on Amazon Mechanical Turk. In *Proceedings of the 2010 Workshop on Human Computation (HCOMP 2010)*, pp. 64–67.

Jain, S., Y. Chen, and D. C. Parkes (2009). Designing Incentives for Online Question and Answer Forums. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, pp. 129–138. ACM.

Kanfer, R. (1990). Motivation Theory and Industrial and Organizational Psychology. *Handbook of Industrial and Organizational Psychology 1*(2), 75–130.

Kaufmann, N., T. Schulze, and D. Veit (2011). More than fun and money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk. In *Proceedings of the 17th Americas Conference on Information Systems (AMCIS 2011)*, Volume 11, pp. 1–11.

Kittur, A. (2010). Crowdsourcing, Collaboration and Creativity. *ACM Crossroads 17*(2), 22–26.

Kittur, A., S. Khamkar, P. André, and R. E. Kraut (2012). CrowdWeaver: Visually Managing Complex Crowd Work. In *Proceedings of the 2012 Conference on Computer Supported Cooperative Work (CSCW 2012)*, pp. 1033–1036.

Kittur, A., J. V. Nickerson, M. S. Bernstein, E. M. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. J. Horton (2013). The Future of Crowd Work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pp. 1301–1318.

Kloker, S., T. Kranz, T. Straub, and C. Weinhardt (2016). Shouldn't Collaboration be social? - Proposal of a social Real Time Delphi. In *Proceedings of the 2nd Karlsruhe Service Summit Research Workshop (KSS 2016)*.

Knorr, E. (2003). The Year of Web Services. *CIO*, 90.

Kokkodis, M. and P. G. Ipeirotis (2013). Have you Done Anything Like That? Predicting Performance Using Intercategory Reputation. In *Proceedings of the 6th International Conference on Web Search and Data Mining (WSDM 2013)*, pp. 435–444.

Kosonen, M., C. Gan, M. Vanhala, and K. Blomqvist (2014). User Motivation and Knowledge Sharing in Idea Crowdsourcing. *International Journal of Innovation Management 18*(05), 1–23.

Kräkel, M. (2008). Optimal risk taking in an uneven tournament game with risk averse players. *Journal of Mathematical Economics 44*(11), 1219–1231.

Kranz, T. T., F. Teschner, and C. Weinhardt (2014). Combining Prediction Markets and Surveys: An Experimental Study. In *Proceedings of the 22nd European Conference on Information Systems (ECIS 2014)*.

Krause, M., M. Hall, and S. Caton (2016). Is Quality Control Pointless? In *The Internet, Politics, and Policy (IPP)*.

Kuhnen, C. M. and A. Tymula (2012). Feedback, Self-Esteem, and Performance in Organizations. *Management Science 58*(1), 94–113.

Kunze von Bischhoffshausen, J., P. Hottum, and T. Straub (2015). Service Co-creation. In *Fundamentals of Service Systems*, pp. 261–296. Springer.

Lacity, M. C. and R. Hirschheim (1993). The Information Systems Outsourcing Bandwagon. *Sloan Management Review 35*(1), 73.

Latane, B., K. Williams, and S. Harkins (1979). Many Hands Make Light the Work: The Causes and Consequences of Social Loafing. *Journal of Personality and Social Psychology 37*(6), 822.

Lawler, E. (1971). *Pay and Organizational Effectiveness: A Psychological View*. McGraw-Hill.

Layman, L. and G. Sigurdsson (2013). Using Amazon's Mechanical Turk for User Studies: Eight Things You Need to Know. In *Proceedings of the 2013 ACM/IEEE International Symposium on IEEE Empirical Software Engineering and Measurement*, pp. 275–278.

Lazear, E. P. and R. Sherwin (1981). Rank-Order Tournaments as Optimum Labor Contracts. *The Journal of Political Economy 89*(5), 841–864.

Lee, S. M., D. L. Olson, and S. Trimi (2012). Innovative Collaboration for Value Creation. *Organizational Dynamics 41*(1), 7–12.

Leimeister, J. M. (2010). Collective Intelligence. *Business & Information Systems Engineering 2*(4), 245–248.

Leimeister, J. M., M. Huber, U. Bretschneider, and H. Krcmar (2009). Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition. *Journal of Management Information Systems 26*(1), 197–224.

Liu, D., X. Li, and R. Santhanam (2013). Digital Games and Beyond: What Happens When Players Compete? *MIS Quarterly 37*(1), 111–124.

Locke, E. A. (2001). Motivation by Goal Setting. In *Handbook of Organizational Behavior*, pp. 43–56. Dekker.

Locke, E. A. and J. F. Bryan (1969). The Directing Function of Goals in Task Performance. *Organizational Behavior and Human Performance 4*(1), 35–42.

Locke, E. A. and G. P. Latham (2002). Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. *American Psychologist 57*(9), 705–717.

Lorenz, J., H. Rauhut, F. Schweitzer, and D. Helbing (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences 108*(22), 9020–9025.

Malone, T. W., R. Laubacher, and C. Dellarocas (2010). The Collective Intelligence Genome. *IEEE Engineering Management Review 38*(3), 38.

Mao, A., Y. Chen, K. Z. Gajos, D. Parkes, and A. D. Procaccia (2012). TurkServer: Enabling Synchronous and Longitudinal Online Experiments. In *Proceedings of the 2012 Workshop on Human Computation (HCOMP 2012)*.

Marden, O. S. (1908). *He Can who Thinks He Can, and Other Papers on Success in Life*. TY Crowell & Company.

Marjanovic, S., C. Fry, and J. Chataway (2012). Crowdsourcing based business models: In search of evidence for innovation 2.0. *Science and Public Policy 39*(3), 1–15.

Maslow, A. H. (1943). A Theory of Human Motivation. *Psychological Review 50*(4), 370–396.

Mason, W. and S. Suri (2012). Conducting Behavioral Research on Amazon's Mechanical Turk. *Behavioral Research Methods 44*(1), 1–23.

Mason, W. and D. J. Watts (2010). Financial Incentives and the Performance of Crowds. *ACM SigKDD Explorations Newsletter 11*(2), 100–108.

Mazarakis, A. (2013). *Feedback und Anreize für die Nutzung von Web 2.0 Diensten*. Ph. D. thesis, Karlsruhe, Karlsruher Institut für Technologie (KIT).

Mazarakis, A., S. Braun, and V. Zacharias (2011). Feedback in social semantic applications. *International Journal of Knowledge Engineering and Data Mining 1*(4), 291–302.

Mazarakis, A. and J. Krämer (2010). Feedbackmechanismen zur Steigerung der Beteiligung im Web 2.0 – Ergebnisse einer Feldstudie zu Vorlesungswikis. In *Mensch & Computer Workshopband*, pp. 186–191.

Mazarakis, A. and C. Van Dinther (2011). Feedback Mechanisms and their Impact on Motivation to Contribute to Wikis in Higher Education. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pp. 215–216. ACM.

Meub, L., T. Proeger, K. Bizer, and R. Henger (2015). Die Effizienz von Zuteilungsmechanismen bei Flächenzertifikaten zwischen Versteigerung und Grandfathering: Experimentelle Evidenz. *Discussion Papers, Center for European Governance and Economic Development Research*.

Meub, L., T. Proeger, K. Bizer, and R. Henger (2016). Experimental evidence on the resilience of a cap & trade system for land consumption in Germany. *Land Use Policy 51*, 95–108.

Muchnik, L., S. Aral, and S. J. Taylor (2013). Social Influence Bias: A Randomized Experiment. *Science 341*(6146), 647–651.

Nakagawa, S. and H. Schielzeth (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution 4*(2), 133–142.

Nardi, B. A., D. J. Schiano, M. Gumbrecht, and L. Swartz (2004). Why We Blog. *Communications of the ACM 47*(12), 41–46.

Nickerson, J. V., Y. Sakamoto, and L. Yu (2011). Structures for Creativity: The crowdsourcing of design. In *Proceedings of the 2011 CHI Workshop on Crowdsourcing and Human Computation*, pp. 1–4. Citeseer.

Nieken, P. and D. Sliwka (2010). Risk-taking tournaments – Theory and experimental evidence. *Journal of Economic Psychology 31*(3), 254–268.

Niemeyer, C., T. Wagenknecht, T. Teubner, and C. Weinhardt (2016). Participatory Crowdfunding: An Approach towards Engaging Employees and Citizens in Institutional Budgeting Decisions. In *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS 2016)*, pp. 2800–2808.

Oh, J. and G. Wang (2012). Evaluating Crowdsourcing through Amazon Mechanical Turk as a Technique for Conducting Music Perception Experiments. In *Proceedings of the 12th International Conference on Music Perception and Cognition*, pp. 1–6.

Oppenheimer, D. M., T. Meyvis, and N. Davidenko (2009). Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power. *Journal of Experimental Social Psychology 45*(4), 867–872.

Oreg, S. and O. Nov (2008). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior 24*(5), 2055–2073.

O'Reilly, T. (2006). Web 2.0 Compact Definition: Trying Again. `http://radar.oreilly.com/2006/12/web-20-compact-definition-tryi.html`.

Orne, M. T. (1962). On the social psychology experiment: With particular reference to demand characteristics and their impications. *American Psychologist 17*(11), 776.

Paolacci, G. and J. Chandler (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Curent Directions in Psychological Science 23*(3), 184–188.

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010). Running experiments on Amazon Mechanical Turk. *Judgement and Decision Making 5*(5), 411–419.

Pedersen, W. C., T. F. Denson, R. J. Goss, E. A. Vasquez, N. J. Kelley, and N. Miller (2011). The impact of rumination on aggressive thoughts, feelings, arousal, and behaviour. *British Journal of Social Psychology 50*(2), 281–301.

Peer, E., J. Vosgrau, and A. Acquisti (2013). Reputation as a Sufficient Condition for Data Quality on Amazon Mechanical Turk. *Behavioral Research Methods 46*(4), 1023–1031.

Petty, R. E., S. G. Harkins, K. D. Williams, and B. Latane (1977). The Effects of Group Size on Cognitive Effort and Evaluation. *Personality and Social Psychology Bulletin 3*(4), 579–582.

Pilz, D. and H. Gewald (2013). Does Money Matter? Motivational Factors for Participation in Paid-and Non-Profit-Crowdsourcing Communities. In *Proceedings of the 11th International Conference on Wirtschaftsinformatik*, pp. 37.

Preacher, K. J. and A. F. Hayes (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers 36*(4), 717–731.

Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature 37*(1), 7–63.

Proeger, T., L. Meub, and K. Bizer (2015). Tradable development rights under uncertainty: An experimental approach. *Discussion Papers, Center for European Governance and Economic Development Research*.

Proeger, T., L. Meub, and K. Bizer (2016). The role of communication on an experimental market for tradable development rights. *Discussion Papers, Center for European Governance and Economic Development Research*.

Prpić, J. (2016). Next Generation Crowdsourcing for Collective Intelligence. *Collective Intelligence 2016: Proceedings*.

Prpić, J., P. P. Shukla, J. H. Kietzmann, and I. P. McCarthy (2015). How to work a crowd: Developing crowd capital through crowdsourcing. *Business Horizons 58*(1), 77–85.

Pull, K., H. Bäker, and A. Bäker (2013). The Ambivalent Role of Idiosyncratic Risk in Asymmetric Tournaments. *Theoretical Economics Letters 3*(3), 16–22.

Qiu, J. (2003). Termination Risk, Multiple Managers and Mutual Fund Tournaments. *European Finance Review 7*(2), 161–190.

Rafaeli, S. and Y. Ariel (2008). Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia. In A. Barak (Ed.), *Psychological Aspects of Cyberspace*, pp. 243–267. New York, NY: Cambridge University Press.

Rammstedt, B. and O. P. John (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality 41*(1), 203–212.

Rand, D. G. (2012). The Promise of Mechanical Turk: How Online Labor Markets Can Help Theorists Run Behavioral Experiments. *Journal of Theoretical Biology 299*, 172–179.

Reeve, J. and E. L. Deci (1996). Elements of the Competitive Situation That Affect Intrinsic Motivation. *Personality and Social Psychology Bulletin 22*(1), 24–33.

Reiss, S. (2012). Intrinsic and Extrinsic Motivation. *Teaching of Psychology 39*(2), 152–156.

Rheinberg, F. (2008). *Motivation, 7. ed.* Grundriss der Psychologie. Stuttgart: Kohlhammer.

Rheinberg, F. (2011). Vom Motiv zur Motivation. *Forschung & Lehre 11*(12), 929–931.

Rogstadius, J., V. Kostakos, A. Kittur, B. Smus, J. Laredo, and M. Vukovic (2011). An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*.

Röll, F. J. (2010). Social Network Sites. In *Digitale Jugendkulturen*, pp. 209–224. Springer.

Roth, A. E. (1986). Laboratory experimentation in economics. *Economics and Philosophy 2*(2), 245–273.

Roth, A. E. (1987). *Laboratory experimentation in economics: Six points of view*. Cambridge University Press.

Rouse, A. C. (2010). A Preliminary Taxonomy of Crowdsourcing. In *Proceedings of the 21st Australasian Conference on Information Systems (ACIS 2010)*.

Ryan, R. M. and E. L. Deci (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology 25*(1), 54–67.

Ryckman, R. M., M. Hammer, L. M. Kaczor, and J. A. Gold (1996). Construction of a Personal Development Competitive Attitude Scale. *Journal of Personality Assessment 66*(2), 374–385.

Saxton, G. D., O. Oh, and R. Kishore (2013). Rules of Crowdsourcing: Models, Issues, and Systems of Control. *Information Systems Management 30*(1), 2–20.

Schattke, K. P. (2011). *Flow Experience as Consequence and Self-Determination as Antecedent of Congruence between Implicit and Explicit Motives*. Ph. D. thesis, München, Technische Universität München.

Schlagwein, D. and D. Farhad (2014). User Requirements of a Crowdsourcing Platform for Researchers: Findings from a Series of Focus Groups. In *PACIS 2014 Proceedings*.

Schultheiss, O. C., A. Kordik, J. S. Kullmann, M. Rawolle, and A. G. Rösch (2009). Motivation as a natural linchpin between person and situation. *Journal of Research in Personality 43*(2), 268–269.

Shaw, A. D., J. J. Horton, and D. L. Chen (2011). Designing Incentives for Inexpert Human Raters. In *Proceedings of the 2011 Conference on Computer Supported Cooperative Work (CSCW 2011)*, pp. 275–284.

Smith, V. L. (1976). Experimental Economics: Induced Value Theory. *American Economic Review 66*(2), 274–279.

Snijders, T. A. and R. J. Bosker (1994). Modeled Variance in Two-Level Models. *Sociological Methods & Research 22*(3), 342–363.

Sorokin, A. and D. Forsyth (2008). Utility data annotation with Amazon Mechanical Turk. In *Proocedings of the 1st IEEE Workshop on Internet Vision*. IEEE.

Spann, M. and B. Skiera (2003). Internet-Based Virtual Stock Markets for Business Forecasting. *Management Science 49*(10), 1310–1326.

Stallen, M., C. K. De Dreu, S. Shalvi, A. Smidts, and A. G. Sanfey (2012). The Herding Hormone: Oxytocin Stimulates In-Group Conformity. *Psychological Science 23*(11), 1288–1292.

Stebbins, R. A. (2001). *Exploratory Research in the Social Sciences*, Volume 48. Sage.

Stoner, J. A. (1968). Risky and Cautious Shifts in Group Decisions: The Influence of Widely Held Values. *Journal of Experimental Social Psychology 4*(4), 442–459.

Straub, T., H. Gimpel, and F. Teschner (2014). The negative effect of feedback on performance in crowd labor tournaments. In *Collective Intelligence 2014: Proceedings*.

Straub, T., H. Gimpel, F. Teschner, and C. Weinhardt (2014). Feedback and Performance in Crowd Work: A Real Effort Experiment. In *Proceedings of the 22nd European Conference on Information Systems (ECIS 2014)*.

Straub, T., H. Gimpel, F. Teschner, and C. Weinhardt (2015). How (not) to Incent Crowd Workers. *Business & Information Systems Engineering 57*(3), 167–179.

Straub, T., F. Hawlitschek, and C. Weinhardt (2016). Towards a Guideline for Conducting Economic Experiments on Amazon's Mechanical Turk. In *Proceedings of the 2nd Karlsruhe Service Summit Research Workshop (KSS 2016)*.

Straub, T., M. Kohler, P. Hottum, V. Arrass, and D. Welter (2013). Customer Integration in Service Innovation: An Exploratory Study. *Journal of Technology Management & Innovation 8*(3), 25–33.

Straub, T., T. Teubner, and C. Weinhardt (2016). Risk Taking in Online Crowdsourcing Tournaments. In *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS 2016)*, pp. 1851–1860.

Subramani, M. R. and N. Peddibhotla (2003). Contributing to Document Repositories – An Examination of Prosocial Behavior. *Working Paper*.

Suri, S. and D. J. Watts (2011). Cooperation and Contagion in Web-Based, Networked Public Goods Experiments. *PLoS ONE 6*(3), 1–18.

Tapscott, D. (2012). Don Tapscott: Four Principles for the Open World (Ted Talk). `https://www.ted.com/talks/don_tapscott_four_principles_for_the_open_world_1`, `https://www.ted.com/talks/don_tapscott_four_principles_for_the_open_world_1/transcript`.

Taylor, J. (2003). Risk-taking behavior in mutual fund tournaments. *Journal of Economic Behavior & Organization 50*(3), 373–383.

Tedjamulia, S. J. J., D. L. Dean, D. R. Olsen, and C. C. Albrecht (2005). Motivating Content Contributions to Online Communities: Toward a More Comprehensive Theory. In *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS 2005)*. IEEE.

Terranova, T. (2004). *Network Culture: Politics for the Information Age*. Pluto Press.

Terwiesch, C. and Y. Xu (2008). Innovation Contests, Open Innovation, and Multiagent Problem Solving. *Management Science 54*(9), 1529–1543.

Teschner, F. and H. Gimpel (2013). Crowd Labor Markets as Platform for IS Research: First Evidence from Electronic Markets. In *Proceedings of the 2013 International Conference on Information Systems (ICIS 2013)*.

Teschner, F., A. Mazarakis, R. Riordan, and C. Weinhardt (2011). Participation, Feedback & Incentives in a Competitive Forecasting Community. In *Proceedings of the 2011 International Conference on Information Systems (ICIS 2011)*.

Teubner, T., M. T. Adam, S. Camacho, and K. Hassanein (2014). Understanding Resource Sharing in C2C Platforms: The Role of Picture Humanization. In *Proceedings of the 25th Australasian Conference on Information Systems (ACIS 2014)*. ACIS.

Thorsnes, P. and G. P. Simons (1999). Letting the Market Preserve Land: The Case for a Market-Driven Transfer of Development Rights Program. *Contemporary Economic Policy 17*(2), 256–266.

Tranquillini, S., F. Daniel, P. Kucherbaev, and F. Casati (2015). Modeling, Enacting, and Integrating Custom Crowdsourcing Processes. *ACM Transactions on the Web (TWEB) 9*(2), 7.

United Nations (1998). Kyoto Protocol to the United Nations Framework Convention on Climate Change. `http://unfccc.int/resource/docs/convkp/kpeng.pdf`.

Vallerand, R. J., L. G. Pelletier, and R. Koestner (2008). Reflections on Self-Determination Theory. *Canadian Psychology 49*(3), 257.

van Dijk, F., J. Sonnemans, and F. van Winden (2001). Incentive systems in a real effort experiment. *European Economic Review 45*(2), 187–214.

Vandercruysse, S., M. Vandewaetere, F. Cornillie, and G. Clarebout (2013). Competition and students' perceptions in a game-based language learning environment. *Educational Technology Research and Development 61*(6), 927–950.

Vansteenkiste, M., C. P. Niemiec, and B. Soenens (2010). The Development of the Five Mini-Theories of Self-Determination Theory: An Historical Overview, Emerging Trends, and Future Directions. *Advances in Motivation and Achievement 16*, 105–166.

Von Ahn, L. (2006). Games with a Purpose. *Computer 39*(6), 92–94.

Vroom, V. H. (1964). *Work and Motivation*. Wiley.

Wallach, M. A., N. Kogan, and D. J. Bem (1964). Diffusion of Responsibility and Level of Risk Taking in Groups. *The Journal of Abnormal and Social Psychology 68*(3), 263.

Wang, J., P. G. Ipeirotis, and F. Provost (2013). Quality-Based Pricing for Crowdsourced Workers. *NYU Stern Research Working Paper*.

Wang, T., D. Wang, and F. Wang (2014). Quantifying Herding Effects in Crowd Wisdom. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1087–1096. ACM.

Weinhardt, C. and H. Gimpel (2007). Market Engineering: An Interdisciplinary Research Challenge. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Weinhardt, C., C. Holtmann, and D. Neumann (2003). Market-Engineering. *Wirtschaftsinformatik 45*(6), 635–640.

Wilson, R. E., S. D. Gosling, and L. T. Graham (2012). A Review of Facebook Research in the Social Sciences. *Perspectives on Psychological Science 7*(3), 203–220.

Wolfers, J. and E. Zitzewitz (2004). Prediction Markets. Technical report, National Bureau of Economic Research.

Xu, B. and D. Li (2015). An empirical study of the motivations for content contribution and community participation in Wikipedia. *Information & Management 52*(3), 275–286.

Yu, L. and J. V. Nickerson (2011). Cooks or Cobblers? Crowd Creativity through Combination. In *Proceedings of the 29th SIGCHI Conference on Human Factors in Computing Systems*, pp. 1393–1402. ACM.

Zajonc, R. B., R. J. Wolosin, M. A. Wolosin, and S. J. Sherman (1968). Individual and Group Risk-Taking in a Two-Choice Situation. *Journal of Experimental Social Psychology 4*(1), 89–106.

Zwass, V. (2010). Co-Creation: Toward a Taxonomy and an Integrated Research Perspective. *International Journal of Electronic Commerce 15*(1), 11–48.