

Essays on Crowd-based Idea Evaluation - Empirical Evidence from an Anonymous Online Crowd

Der Fakultät für Wirtschaftswissenschaften der

Universität Paderborn

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

- Doctor rerum politicarum -

vorgelegte Dissertation

von

Thomas Görzen, M.Sc.

geboren am 23.06.1986 in Paderborn

Juli 2019

Acknowledgements

Throughout the writing of this dissertation and my time as a doctoral student, I was very fortunate in having people around me that supported me through this intensive time.

First, I would like to thank my supervisor Dennis for his feedback, guidance and the opportunities he gave me. Next, I would like to thank my wife Olga, who was always by my side and strengthened me at every step of the way. I would also like to thank my parents and my family who have always supported me. Without the warm love, continued patience, and endless support of my wife and my family, I would not have been able to reach my goals.

In addition, I would also like to thank my friends and (former and current) colleagues for their support. I will always remember the great time we had together - without you this time would have been less enriching and less fun. Finally, I would like to thank Gabi for your feedback and the time you spent, helping me to avoid the stumbling blocks of the English language.

Table of Contents

Synopsis

1. Introduction.....	1
2. Conceptual Basics.....	3
2.1 Crowdsourcing.....	3
2.2 Crowd Work.....	6
2.3 Crowd Ideation and Crowd Evaluation.....	7
3. Status Quo on Crowd-based Idea Evaluation.....	9
4. Overview and Contribution.....	14
4.1 Görzen and Kundisch (2016).....	16
4.2 Görzen and Laux (2019).....	17
4.3 Görzen and Kundisch (2019).....	18
4.4 Görzen (2019a).....	19
4.5 Görzen (2019b).....	20
5. Conclusion.....	21
5.1 Implications.....	22
5.2 Limitations and Future Research.....	23
Appendix.....	25
References.....	26

Submitted Research Papers as Part of the Thesis

1. Görzen, T. and Kundisch, D. 2016. "Can the Crowd Substitute Experts in Evaluation of Creative Ideas? An Experimental Study Using Business Models," *Proceedings of the 22nd Americas Conference on Information Systems, San Diego, USA*.
2. Görzen, T. and Laux, F. 2019 "Extracting the Wisdom from the Crowd: A Comparison of Approaches to Aggregating Collective Intelligence," *Working Paper*.
3. Görzen, T. and Kundisch, D. 2019. "When in Doubt Follow the Crowd: How Idea Quality Moderates the Effect of an Anchor on Idea Evaluation," *Working Paper*.
4. Görzen, T. 2019a "Can Experience be Trusted? Investigating the Effect of Experience on Decision Biases in Crowdfunding Platforms," *Working Paper*.
5. Görzen, T. 2019b. "What's the Point of the Task? Exploring the Influence of Task Meaning on Creativity in Crowdsourcing," *Working Paper*.

Synopsis

“To answer the most vexing innovation and research questions, crowds are becoming the partner of choice.”

(Boudreau and Lakhani 2013, p.3)

1. Introduction

Innovations are of central importance to the success of companies or, as Baumol (2002, p.1) argues, (...) “innovative activity becomes a mandatory, a life-and-death matter for a company.” Hence, companies continuously look for sources of ideas for new products, services or business models (Jaruzelski and Denhoff 2010). Traditionally, companies have mainly relied on internal resources to generate ideas, however, many companies have been disappointed with their innovation outcomes (Jaruzelski and Denhoff 2010). As a potential solution to this problem, organizations are increasingly using crowdsourcing to generate new and diverse ideas (Yoo et al. 2012; Palacios et al. 2016). Jeff Howe defined “crowdsourcing” as “the act of taking a task traditionally performed by a designated agent (such as an employee or a contractor) and outsourcing it by making an open call to an undefined but large group of people” (Howe 2008, p.1).

Over the last years, organizations like Nestle, Procter & Gamble, General Electric or the NASA have used crowdsourcing as a part of their research and development process to find new products or solutions (Lee et al. 2018; Palacios et al. 2016). However, most organizations are inexperienced in the use of crowdsourcing, since it represents a new way of ideation and idea selection (Palacios et al. 2016). Engaging the crowd to generate ideas, however, also comes with potential obstacles and challenges.

One of these challenges is the process of idea selection. As Jeff Howe (2008) defined it, crowdsourcing often represents an open call for ideas to an undefined large group of individuals. Depending on the size of this group, companies can easily receive hundreds or even thousands of different ideas which they then need to screen in order to identify the most promising ones (Toubia and Flores 2007). For example, when Starbucks, as part of a campaign called “MyStarbucksIdea” in 2008, asked their customers to submit new product ideas and general ideas to increase customers’ experience, the company received more than 115,000 ideas, as of 2011 (Rosen 2011). Further, Google received about 150,000 ideas to its Project “10 to the 100” (Blohm et al. 2013).

The “classical” way to evaluate ideas was mainly undertaken by internal resources, such as internal product development teams or experts (Mollick and Nanda 2016). However, facing these large numbers of ideas submitted by a crowd could lead to the situation where companies cannot rely on internal resources to evaluate such a large number of ideas, or if they do so, only with disproportionate effort (Poetz and Schreier 2012). Even large companies like Google face these challenges, since it took about 3 years and 3,000 of their employees to evaluate the 150,000 submitted ideas (Blohm et al. 2013).

Therefore, companies have to look for others ways to evaluate a large number of submitted ideas. As a result, companies increasingly not only engage the crowd to generate ideas but also include the crowd in the process of idea evaluation (Soukhoroukova et al. 2012).

Crowd-based idea evaluation is not only of great importance for practice but has also garnered attention in academic research. Previous literature in this context addresses several dimensions for further research, including a managerial (e.g., selection of evaluators), technological (e.g., the use of idea evaluation tools and appropriate evaluation methods) and behavioral dimensions (e.g., mechanisms to motivate individuals in a crowd or mechanisms to create trust in crowdsourcing) (Chiu et al. 2014). Further, existing meta-analyses (e.g., Wagenknecht et al. 2017) highlight the need for additional research to investigate the potential occurrence of decision making biases (e.g., information cascades) in crowd-based evaluation and the need to answer open research questions like “Which mechanisms allow a crowd to identify the best idea(s)?” (Merz 2018).

Since one of the key objectives of idea evaluation is to identify the best ideas out of a large number of ideas, previous literature already investigated the ability of crowds for this task. Results indicate that crowds often fail to do so (Blohm et al. 2016). In addition, other studies (e.g., Klein and Garcia 2015) indicate that crowds are better at eliminating the worst ideas rather than identifying the best ones. However, even if a crowd were able to eliminate the worst ideas, companies could still be left with too many ideas to evaluate. Hence, exploring ways to identify the best ideas using crowd-based idea evaluation is of importance to both practice and research.

The main goal of this cumulative dissertation is therefore to investigate different facets of crowd-based idea evaluation, aiming to find solutions on how an external anonymous crowd could identify the best ideas.

The first four articles in this dissertation investigate different aspects of crowd-based idea evaluation, focusing on engaging an external, anonymous crowd on an online platform. The first paper *Görzen and Kundisch (2016)* compares the evaluation of an external anonymous crowd with experts' evaluation. Based on this study, further studies address different aspects of how an anonymous online crowd can be used to identify good ideas. While the second paper *Görzen and Laux (2019)* compares different mechanisms (e.g., majority-voting, confidence-based mechanism) for aggregating multiple opinions by a crowd, in the third paper *Görzen and Kundisch (2019)*, we propose a potential solution on how to identify high quality ideas by developing an appropriate approach, using the established anchoring effect as a basis for a mechanism and the variance of ratings as a key metric. Next, the fourth paper *Görzen (2019a)* investigates the occurrence of a decision-making bias, namely the anchoring effect and the influence of experience on decision making.

The dissertation is rounded off with a fifth paper *Görzen (2019b)*, which contributes to the area of crowd ideation by investigating the influence of task meaning on output quantity as well as output creativity.

Together, these studies make an original contribution to knowledge by increasing the understanding of different facets of crowd-based idea evaluation. The following pages are structured as follows: Section two discusses the general concept of crowdsourcing and further describes the different approaches and applications in practice. The subsequent third section reviews the current state of the literature on crowd-based idea evaluation as well as the research gaps this dissertation seeks to address. Section four presents an overview of all five papers, including summary tables with detailed information regarding the contribution of co-authors and the scientific dissemination of the papers. Finally, section five summarizes the implications, discusses the limitations of the studies and offers directions for future research.

2. Conceptual Basics

2.1 Crowdsourcing

Due to increasing digitalization, the way businesses create value has changed. It is also worth noting that this technological development has blurred companies' boundaries, which has led to a closer connection between internal and external business processes (Durward et al. 2016). Moreover, companies increasingly engage the large number of Internet users into their own processes (Chen and Horton 2016; Boudreau et al. 2011). Crowdsourcing is an umbrella term that describes different approaches, all aiming to outsource companies' previously internal work to large, usually unknown groups of individuals (Chiu et al. 2014). Despite the potential advantages of using crowdsourcing, e.g. a significant time and cost reduction, there are several other reasons for companies to use crowdsourcing: finding solutions for different problems, limited internal resources or the need to find new ideas for products or services (Bayus 2013; Chiu et al. 2014). Crowdsourcing is typically applied via the Internet, via IT-based platforms that act as intermediaries and offer companies the opportunity to get access to a large pool of individuals that might work on the proposed task (Hammon and Hippner 2012; Durward and Blohm 2017). For example, Amazon MTurk offers access to more than 500,000 individuals (as of 2017) from more than 60 different countries (Peer et al. 2017). Hence, crowdsourcing platforms offer, as Amazon advertises, potential access to a "global, on demand, 24/7 workforce" (Amazon 2019). Apart from this example, the crowdsourcing market is recently growing, and the World Bank expects the market size for paid crowdsourcing to grow up to \$25 billion in 2020 (Kuek et al. 2015).

The term crowdsourcing is a neologism, combining the words crowd and outsourcing. As displayed in Figure 1, outsourcing (I) describes the process of outsourcing a specific corporate task to a designated third-party like another company or institution (see Figure 1). In crowdsourcing this task is assigned to an undefined large number of individuals ('the crowd', Zogaj et al. 2014). The outsourcing organization that crowdsources the task in this case is denoted as the *crowdsourcer*, while the other party, the group

of individuals who perform the outsourced task, is labeled *crowdsources* (Zogaj et al. 2014). Following Zogaj et al. (2014), crowdsourcing can be differentiated between (II) crowdsourcing without a mediation or (III) mediated crowdsourcing (Figure 1).

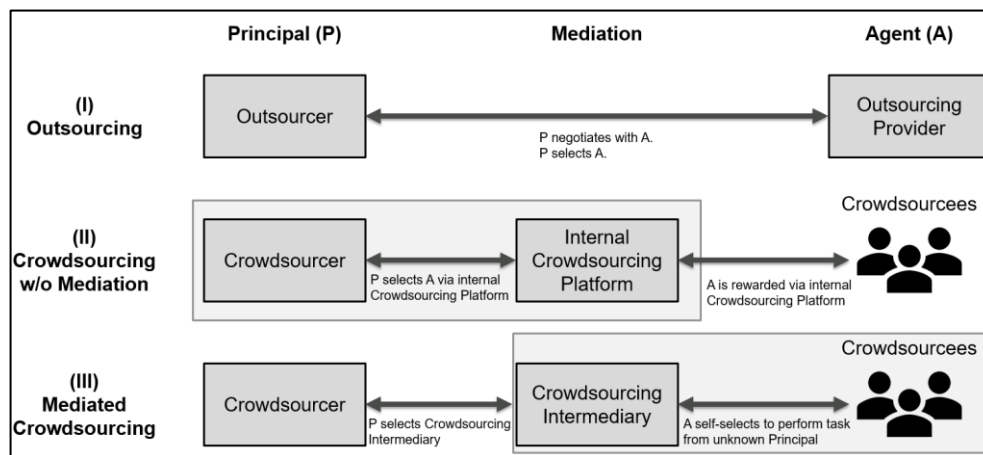


Figure 1. Roles and Mediation in Crowdsourcing Initiatives (Adopted from Zogaj et al. 2014)

Crowdsourcing without a mediation describes the case whereby the crowdsourcer establishes an internal crowdsourcing platform that is used to interact with the crowd. Crowdsources in this case can be employees working for the crowdsourcer or external individuals who, for example submit solutions or ideas to an internal crowdsourcing platform. This crowdsourcing platform is established and hosted by the crowdsourcer (Zogaj et al. 2014). As an example, Starbucks started a campaign called “MyStarbucksIdea” and asked their customers to submit ideas on a company’s own website for new product ideas and ideas to increase customers’ experience in Starbucks stores (Rosen 2011).

Mediated crowdsourcing describes another case of crowdsourcing whereby an external crowdsourcing platform acts as an intermediary between the crowdsourcer and the crowdsources (Zogaj et al. 2014). These crowdsourcing intermediaries are web platforms that manage the interaction between the two parties. While the crowdsourcer is responsible for creating the task for the crowdsources to solve, the crowdsourcing intermediary needs to attract and manage the crowdsources (Zogaj et al. 2014). In most cases companies represent the crowdsourcer, although not only non-profit organizations but potentially anyone with access to the Internet could act as an initiator for an open call (Hammon and Hippner 2012).

In recent years, several platforms acting as crowdsourcing intermediaries between crowdsourcer and crowdsources have emerged (see Table 1 for examples). While some platforms focus on microtasks (e.g. MTurk, Clickworker or Figure Eight), such as short sequences of audio description or image labeling, other platforms focus on design task (e.g., 99Designs), innovation (e.g., Innocentive), software

testing (Testbirds) or general access to freelancers (Freelancer) for designing logos or programming webpages for example.

Table 1. Examples of Crowdsourcing Platforms

<i>Platform</i>	<i>URL</i>	<i>Domain</i>	<i>Crowd (reference of data)</i>	<i>Size year</i>
99Designs	99designs.de	Design	1,4 Mio.	(2019)
Clickworker	clickworker.de	Microtasks	~1,5 Mio.	(2019)
Figure Eight	figure-eight.com	Microtasks	10,000+	(2017)
Freelancer	freelancer.com	Diverse projects (writing, programming, design etc.)	30 Mio.+	(2019)
Innocentive	innocentive.com	Innovation	400,000+	(2019)
MTurk	mturk.com	Microtasks	500,000+	(2017)
Prolific	prolific.ac	Online studies for academic research	45,000+	(2019)
Testbirds	testbirds.de	Software testing	300,000+	(2019)

Given the number of different domains and applications, crowdsourcing has garnered considerable interest from practitioners since it offers the potential of transforming permanent jobs into a flexible resource pool (Howe 2008). Presented on the basis of a companies' value chain, and divided into primary and secondary activities, many different approaches and applications have been established in practice. Hence, almost all primary and secondary activities (see Figure 2) can be performed completely or at least with some support from the crowd (Durward et al. 2016). For instance, Microsoft asked crowds to test their software (Crowd Testing), while the TV production firm Brainpool used the concept of Crowd Funding to partially fund the popular German Movie *Stromberg* (Durward et al. 2016).

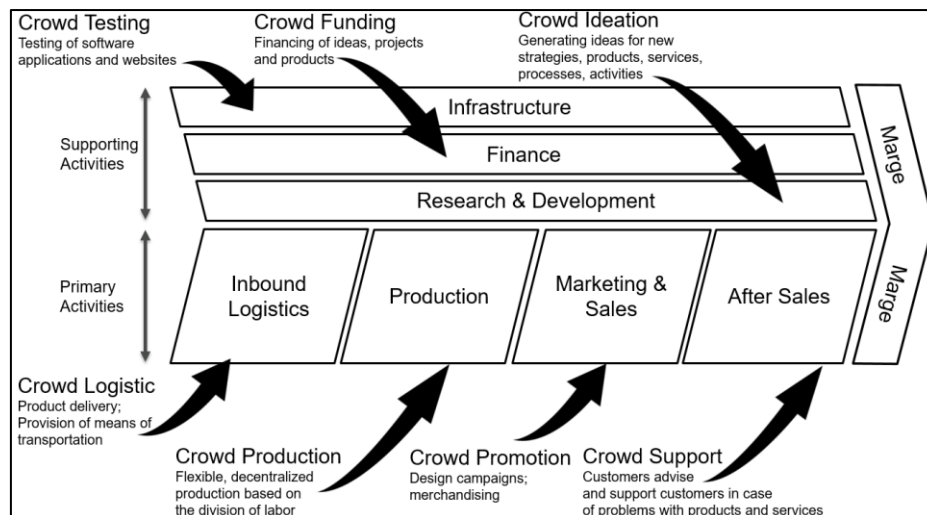
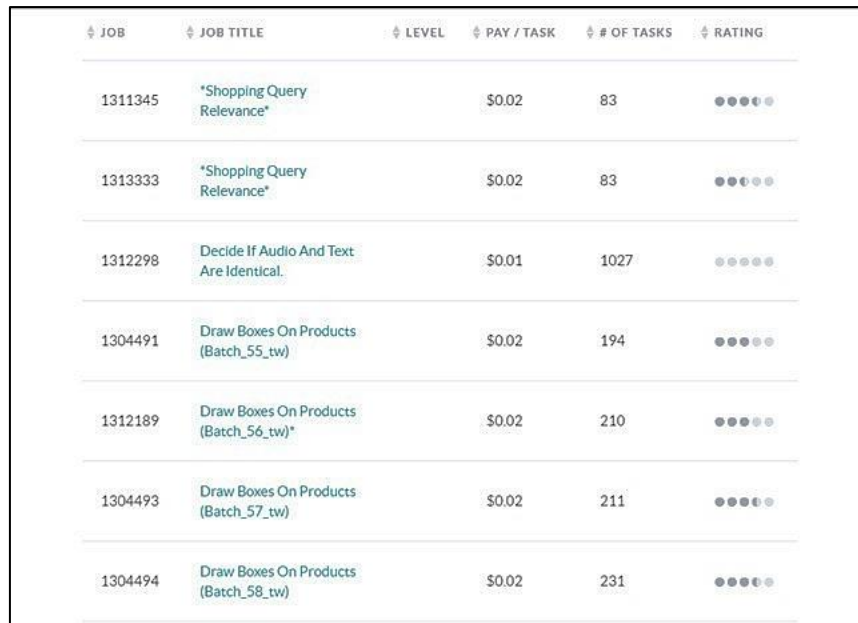


Figure 2. Applications of Crowdsourcing (Adopted from Durward et al. 2016)

2.2 Crowd Work

All mentioned crowdsourcing platforms above (see Table 1) have in common that they offer crowdsourcers monetary rewards in exchange for solving the offered tasks. Despite other motivations to participate in crowdsourcing, like the opportunity to develop one's skills (e.g., Kosonen et al. 2014) or obtaining reputation and recognition (Zheng et al. 2011), the opportunity to earn money represents a key motivation to participate in crowdsourcing (e.g., Brabham 2010; Leimeister et al. 2009). While Wikipedia, MyStarbucksIdea or Dell IdeaStorm represent popular examples of unpaid crowdsourcing without offering monetary rewards, most crowdsourcers offer payments in exchange for the work of crowdsourcers on large crowdsourcing intermediaries like MTurk, Clickworker or Figure Eight that are available world-wide. Hence, crowdsourcing should be differentiated according to whether it is paid or unpaid. In line with previous literature (Durward et al. 2016), I consider paid crowdsourcing as *crowd work*, since many people work in full-time on these platforms and the monetary rewards from crowd work represent the only or at least an additional income for a large number of individuals (Strube 2015). Further, contest-based platforms (e.g., 99Designs or Innocentive) only offer monetary rewards for the winners, so crowdsourcers only get paid if their suggested design or solution is selected by the crowdsourcer in a contest. In contrast, crowdsourcers on microtask platforms (e.g. MTurk, Clickworker or Figure Eight) mostly receive monetary rewards in form of a fixed payment per completed task. Figure 3 displays an example of a dashboard from the perspective of a crowdsourcer on the platform Figure Eight. As displayed, the crowdsourcer could freely choose between different tasks with different content (e.g., "Decide if Audio and Text are Identical"). The offered amount of money of tasks on this screenshot varies between \$0.01 - \$0.02 and represents the typical form of micropayment on these platforms (Mason and Watts 2009).



JOB	JOB TITLE	LEVEL	PAY / TASK	# OF TASKS	RATING
1311345	*Shopping Query Relevance*		\$0.02	83	●●●●●
1313333	*Shopping Query Relevance*		\$0.02	83	●●●●●
1312298	Decide If Audio And Text Are Identical.		\$0.01	1027	●●●●●
1304491	Draw Boxes On Products (Batch_55_tw)		\$0.02	194	●●●●●
1312189	Draw Boxes On Products (Batch_56_tw)*		\$0.02	210	●●●●●
1304493	Draw Boxes On Products (Batch_57_tw)		\$0.02	211	●●●●●
1304494	Draw Boxes On Products (Batch_58_tw)		\$0.02	231	●●●●●

Figure 3. Screenshot from a Task Dashboard from Figure Eight (as of October 20, 2018)

This dissertation focuses on crowd work since this form represents one of the most popular forms of crowdsourcing (Feyisetan et al. 2015). All studies in this cumulative dissertation used the platform Figure Eight, where microtasks are offered to the crowdsourcees via an open call and participants get paid after completing a task.

2.3 Crowd Ideation and Crowd Evaluation

Although different applications of crowdsourcing have evolved, crowd ideation - asking a crowd to generate ideas for new products, services or business models - has received particular attention from both practice and research (e.g., Wagenknecht et al. 2017; Mack and Landau 2015; Zhu et al. 2017). Previous studies (e.g., Poetz and Schreier 2012; Ebel et al. 2016) show that crowds are able to generate valuable ideas for products, services or business models. Hence, companies increasingly integrate the crowd into their own ideation process. However, studies in this context have found that using crowdsourcing to generate new ideas could lead to a large quantity of ideas generated, regardless of specific participation architectures or approaches (Boudreau 2012; Schemmann et al. 2016), and needs to be evaluated in a next step. The process of evaluating ideas is of great importance to organizations (e.g., Girotra et al. 2010; Riedl et al. 2013) since the quality of the ideas can lead to either market success or market failure (Goldenberg et al. 2001).

Idea evaluation was previously undertaken by internal resources like experts or product development teams (Mollick and Nanda 2016). However, experts are often much sought-after persons with very limited time availability and are therefore a rare resource (Galati 2015). Further, given the situation of evaluating thousands of ideas, identifying the best ideas becomes a time-consuming and resource

intensive challenge, even for large companies (Toubia 2006; Blohm et al. 2013). Therefore, companies increasingly also include the crowd in the idea evaluation process (Soukhoroukova et al. 2012; Mollick and Nanda 2016; Blohm et al. 2016).

The typical process of crowd ideation and crowd-based idea evaluation looks as follows (Figure 4): First, a crowdsourcer seeks to resolve a situation, for example the need to develop a new product. Hence, the crowdsourcer designs an open call for ideas which is broadcasted on a crowdsourcing platform. Next, the members of the crowd can assign themselves to this task. After the crowd has generated the ideas, the crowdsourcer collects these ideas and evaluates them. Zooming into the evaluation (indicated by the circle), the idea evaluation represents a sub-process itself. Idea evaluation can be described as follows: the crowdsourcer broadcasts the open call to evaluate the ideas on a crowdsourcing platform. After the crowd has evaluated the ideas, the crowdsourcer aggregates the evaluation results. Next, the crowdsourcer selects the most promising idea(s) based on the evaluation results. Finally, the crowdsourcer could implement the best idea(s), resulting for example in new products, services, or business models. This dissertation examines the idea evaluation using external crowdsourcees on an external crowdsourcing platform (Figure Eight), aiming to identify the best idea(s).

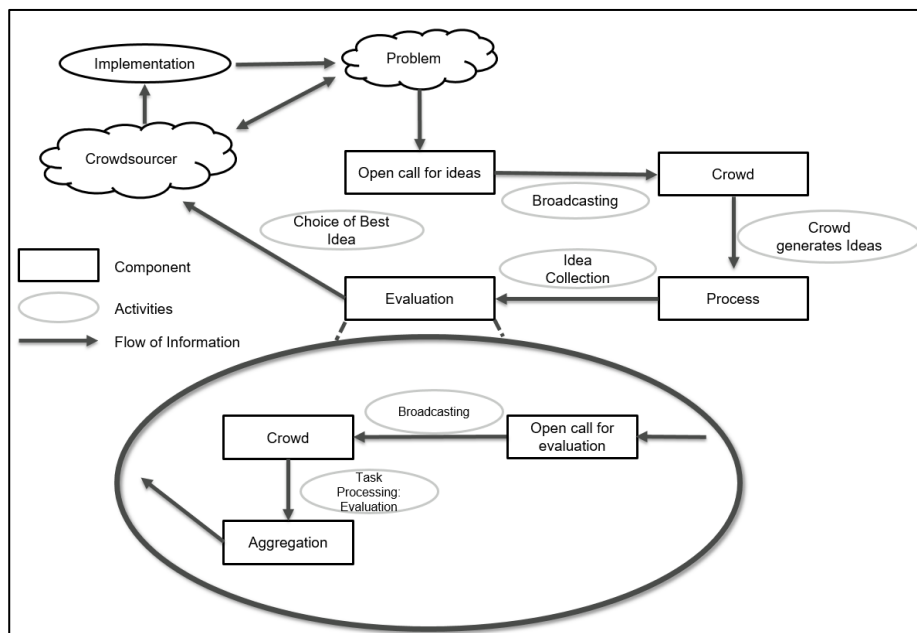


Figure 4. Typical Process of Crowd Ideation (Modified from Chiu et al. 2014)

3. Status Quo on Crowd-based Idea Evaluation

To evaluate the current body of empirical research, I adopted a systematic review strategy (vom Brocke et al. 2015; Webster and Watson 2002). Before evaluating the body of the existing literature myself, I consulted existing reviews focusing on crowd evaluation and open evaluation, which also highlight open research questions which this cumulative dissertation seeks to address (see Table 2) (Merz 2018; Wagenknecht et al. 2017)¹.

First, I screened the literature cited in two reviews, both the concept matrix from Wagenknecht et al. (2017) and the overview table in Merz (2018) and extracted the cited articles. While Wagenknecht et al. (2017) offer a concept matrix that provides an overview of prior studies focusing on idea evaluation, Merz (2018) provides a table that summarizes research designs and methods of previous work on idea evaluation. In addition, I screened the titles of the remaining articles quoted in each of these literature reviews with respect to their relevance to topics like of idea evaluation, idea selection, idea shortlisting and other synonyms². To avoid duplicates, I cross-checked citations from both reviews and consulted potentially relevant articles. The basis for this cross-check represents the review by Wagenknecht et al. (2017), since this review provides a concept matrix that categorizes the identified articles according to specific subjects, for example whether an article focuses on idea generation, collaboration or idea evaluation. Hence, the number of potentially relevant papers from the reviews was higher for Wagenknecht et al. (2017) (31) than for Merz (2018) (13).

Table 2. Consulted Literature Reviews

<i>Source</i>	<i>Focus</i>	<i>Selected research gap(s)*</i>
Merz (2018)	Systematic review of existing literature on selection mechanism in crowdsourcing and open innovation	<ul style="list-style-type: none"> ▪ Which mechanisms allow a crowd to identify the best idea(s)? ▪ Which mechanisms serve the crowd best to shortlist ideas? ▪ How to compensate flaws of existing mechanisms?
Wagenknecht et al. (2017)	Classification and reflection of the debate idea generation, collaboration and evaluation	<ul style="list-style-type: none"> ▪ How robust are different rating scales against decision biases (e.g., information cascades) and related effects? ▪ How to motivate individuals to participate in crowdsourcing?

Note: *Selected research gaps related to the topic of this cumulative dissertation, since the reviews provided more research gaps addressing several additional aspects.

¹ Other reviews on open evaluation exist (e.g., Adamczyk et al. 2012; Hrastinski et al. 2010; West and Bogers 2013), however these reviews are considerably older and therefore do not include some of the relevant work published since then. The reviews by Merz (2018) and Wagenknecht et al. (2017) are more up-to-date and thematically more relevant to this cumulative dissertation.

² For example: “voting”, “rating”, “ranking”, “screening”, “filtering”.

In addition to the paper identification based on the cited papers in the reviews, I conducted a keyword based search via Google Scholar using a variety of terms most commonly used in the context of open evaluation and crowd-based idea evaluation (see Appendix A1). Again, I screened the titles in respect of their relevance to topics like idea evaluation, idea selection, idea shortlisting and other synonyms (see Footnote 2). To avoid duplicates, I did not consider articles already been cited in the literature reviews. The keyword-based search using Google Scholar revealed ten additional potentially relevant articles, which in sum led to 54 articles for my analysis.

Next, I screened the abstracts of all potentially relevant papers and further screened these papers in their full text by searching for terms like ‘experts’, ‘experts’ evaluation’, ‘replacing experts’ etc. The aim of this step was to identify studies where evaluations of both groups (experts and crowds) were compared with each other, since experts’ quality assessment is commonly accepted as the ‘gold-standard’ in idea evaluation (Bayus 2013; Galati 2015). Therefore, crowds’ and experts’ evaluations should be compared with each other to investigate whether crowds could act as potential substitutes for experts in idea evaluation. This screening resulted in 19 potentially relevant articles which I examined in their full text. This final screening identified 11 highly pertinent articles that compared crowds’ and experts’ evaluations, however these articles focused on different facets of idea evaluation. Results of an in-depth analysis of these articles are presented in Table 3.

Previous literature can be differentiated into three streams, according to the type of crowd used for idea evaluation: (i) internal crowds, for example companies’ employees, (ii) external crowds, e.g. users or customers, and (iii) external, i.e. anonymous crowds on an external platform (e.g., Figure Eight or MTurk).

First, several studies used companies’ own employees as a crowd to evaluate ideas. For example, Chen et al. (2009) and Soukhoroukova et al. (2012) compare the results of companies’ internal preference/prediction markets with evaluations from experts. Results are contradictory, since Soukhoroukova et al. (2012) show that the agreement between results from prediction markets and an expert committee is far from high, while Chen et al. (2009) found a high correlation (0.899) between a preference market and the evaluation by an expert panel. Another study compared the evaluation results of experienced and inexperienced employees with those of experts (Onarheim and Christensen 2012). Results indicate the highest share of matches (7 out of 12 top picks) between experienced employees acting as a crowd and experts (Onarheim and Christensen 2012). Further, another study (Klein and Garcia 2015) asked current and former employees of an R&D lab to evaluate ideas for petroleum

exploration. The results of this study (Klein and Garcia 2015) investigating multi-voting mechanisms³ indicate that crowds are better at eliminating the worst ideas instead of identifying the best ones.

The second stream of literature used either students, customers or users as external crowds for idea evaluation. Two studies (Riedl et al. 2013; Blohm et al. 2016) using students from information system courses asked them to evaluate ideas with the aim of improving ERP software. The former study (Riedl et al. 2013) showed that multi-criteria ratings lead to higher decision quality in terms of conformance with experts' rating, than single-criterion ratings. However, this result only holds for relatively short ideas, since additional analyses reveal that idea elaboration (i.e., idea length) negatively moderates this effect, indicating that the single-criterion rating scale outperforms the multi-criteria scale for long ideas (Riedl et al. 2013). Next, the latter study showed that the use of rating scales leads to higher decision quality compared to preference markets. In line with Klein and Garcia (2015) these results indicate that crowds are better at eliminating the worst ideas instead of identifying the best ones (Blohm et al. 2016). Another two studies investigated data from projects funded by the European Union, focusing on ideas for smart cities (Schuurman et al. 2012) and best practice aimed at collecting and translating local projects into 'good practice case studies' (Jarke 2017, p. 373) that can be shared with others in the 'European eGovernment community of practice' (Jarke 2017). Both studies used user from dedicated online-platforms in connection to the respective projects as crowds. However, using a qualitative approach (Jarke 2017) as well as multi-criteria scales (Schuurman et al. 2012) show only low agreements between evaluations by experts and by a crowd. Another study compared evaluation by experts and technically naïve as well as technically skilled users of the Android operation system with each other (Magnusson et al. 2016). Results of this study show good conformance based on ranking of ideas only between technically skilled users and experts (Magnusson et al. 2016).

³ Mechanisms whereby raters are given a limited number of votes which they can freely allocate to ideas. Raters can allocate multiple votes to one idea, for example two or more positive/negative votes, if they believe the idea under consideration to be very good/very bad.

Table 3. Current State of the Literature

Source	Aim of Study	Methodology	Context of Evaluated Ideas	Crowd Used for Evaluation		Could Crowd Identify the Best Ideas?
				Internal (e.g. Employees)	External (e.g. Consumers) Anonymous Crowd on External Platform	
Chen et al. (2009)	Compare two preference markets with ratings by expert panels	Field experiment	Emerging technologies	Employees		✓
Klein and Garcia (2015)	Compare accuracy of different rating mechanisms (Likert scale, 'bag of lemons' and 'bag of stars')	Field experiment	Productivity enhancements for petroleum exploration	Current and former members of R&D lab		x
Onarheim and Christensen (2012)	Compare employees' with executives' ratings	Field experiment	Disposable medical equipment	Employees (inexperienced / experienced)		Partly (experienced shared 7 out of 12 picks with experts)
Soukhoroukova et al. (2012)	Comparison between internal prediction market and senior managers	Field study	Emerging technologies	Employees		x
Blohm et al. (2016)	Compare rating-scale and prediction market for idea evaluation	Web experiment	Ideas to improve the software of SAP		Students from SAP-related MIS courses	x
Jarke (2017)	To examine how 'best practice' may be identified (produced) through a community-based evaluation process as opposed to traditional expert-based evaluation frameworks	Case study (Qualitative approach)	Ideas for the project ("ePractice") from the European Government		User from the online community initiative run by the European Commission "ePractice"	x
Magnumsson et al. (2016)	Investigate conformance of users' assessment to that of experts	Field experiment	Telco services		Users	Partly (ranking of ideas)
Riedl et al. (2013)	Comparison of single-criterion scale and multicriteria rating scale	Web experiment	Ideas to improve an ERP software		Students from four Information System courses	Partly (binary variable)
Schuurman et al. (2012)	Investigating crowdsourcing for generating and selecting ideas for smart cities	Field Experiment	Ideas for smart cities		User from UserVoice.com, mainly citizens of Ghent	x
Kornish and Ulrich (2014)	Explore whether online consumer panelists or experts are better at identifying the potential of raw ideas	Field experiment	Household consumer products			✓
Mollick and Nanda (2016)	Examine whether experts agree with a crowd's funding decisions	Comparative study	Theater projects		Paid online panelists from Qualtrics Funder on crowdfunding platform	✓

A third research stream used anonymous crowds on an external platform. I identified two studies that are of particular importance for this analysis of the current state of literature on crowd-based evaluation, since both studies engaged an anonymous crowd from an external crowdsourcing platform and compared evaluations by experts and by crowds. Mollick and Nanda (2016) compared funding decisions of theater projects with experts' evaluation, while Kornish and Ulrich (2014) asked online panelists to indicate their purchase intention for household products. Results of both studies indicate that online consumer panelists as well as individuals on a crowdfunding platform were able to evaluate ideas on a same level as experts, leading to comparable results between both groups. However both studies represent special evaluation cases: first, Kornish and Ulrich (2014) asked online consumer panelists to evaluate household products, hence a product category with a direct connection to the consumers, with the raters representing the main target group or end users of the given product. As the authors mention in the limitations of their study, they would not expect the results from this specific product domain to apply to other domains (Kornish and Ulrich 2014). Further, results of the other study (Mollick and Nanda 2016) indicate that the crowd systematically misses projects liked by experts because they are not presented in a 'crowdfunding-friendly' format (with videos and pictures).

Next, both evaluations represent specific cases, since companies do not always use crowdsourcing to evaluate relatively simple products, like household goods or theater projects, where nearly every member in the crowd represents a potential consumer. Further, Mollick and Nanda (2016) highlight that theater projects represent evaluation subjects with a highly subjective artistic component, where taste may play an important role. Hence the authors limit the generalizability of their findings to other forms of crowd judgment.

Summarizing the current state of literature, the majority of the studies engaged companies' employees or consumers/users as crowds to evaluate ideas. Further, studies that engaged crowds from external crowdsourcing platforms used relatively simple products (e.g., household products) or products where the crowd represent potential end users (theater projects). Hence, studies using an external crowd from a crowdsourcing platform asking them to evaluate products with a higher distance to the crowd (compared to everyday products like household products) are missing. Further, I did not find a single paper that analyzes the potential occurrence or the potential influence of decision making biases in crowd-based idea evaluation which, however, Wagenknecht et al. (2017) identified as a relevant open research question in this context.

4. Overview and Contribution

This dissertation comprises five research papers. In the following, I describe how this dissertation contributes to the existing literature. The questions that each paper in this context explores empirically are presented in Table 4. A more detailed summary of each paper can be found from chapter 4.1 onwards below.

The first paper *Görzen and Kundisch (2016)* addresses the research gap of comparing both experts' and crowds' ratings, using an anonymous crowd on an external crowdsourcing platform. Further, the ideas under consideration consisted of business model ideas represented in form of a Business Model Canvas, comprising different elements (e.g., key activities, cost structure etc.) with a number of interdependencies. Hence, these ideas represent relatively complex objects to evaluate (Galati 2015). The second paper *Görzen and Laux (2019)* compares different mechanisms to aggregate multiple opinions from a crowd, aiming to identify potential alternatives to simply aggregating judgements since this mechanism is suspected of leading to judgements that are either too extreme or biased, or both (e.g., Bettman et al. 1998; Budescu et al. 2015). The third paper *Görzen and Kundisch (2019)* offers a new approach on how to evaluate a large number of ideas aiming to identify high quality ideas, using an anonymous external crowd from a crowdsourcing platform. Additionally, this paper contributes to the call of research decision-making biases in crowd-based evaluation (Wagenknecht et al. 2017) by including the anchoring effect as a key element of the new evaluation approach. Next, *Görzen (2019a)* addresses the question on the influence of experience on the occurrence of decision biases. Since previous literature provides contradictory results on the influence of experience on the occurrence of the anchoring effect, this paper addresses the question of whether more experienced individuals in the crowd might be less prone to the anchoring effect. Finally, to the best of our knowledge, *Görzen (2019b)* is the first paper to investigate the influence of task meaning on both the number as well as the creativity of generated ideas, using an external crowdsourcing platform.

Table 4. Own Contribution - Overview

<i>Source</i>	<i>Gap Addressed</i>	<i>Methodology</i>	<i>Context of Evaluated Ideas</i>	<i>Research Question(s)</i>	<i>Results and Contribution(s)</i>
Görzen and Kundisch (2016)	Comparison of crowd and experts' assessments using an external anonymous crowd	Field Study	Business model ideas for perfume	1) Can an external anonymous crowd evaluate ideas to a comparable level as experts? 2) Do experienced members in the crowd evaluate ideas as well as experts, compared with inexperienced individuals?	1) First evidence that experts' and external, anonymous crowd ratings are neither comparable on an absolute (average rating) nor on a relative rating. 2) Experience does not decrease the difference between crowds' and experts' assessments.
Görzen and Laux (2019)	Comparison of different approaches to aggregate the opinion of crowds	Field Study	Factual questions of general knowledge & Business model ideas for perfume	Which approach for extracting wisdom from the crowd yields the best and most reliable results?	1) Evidence that the most popular aggregation mechanism (simple aggregating) is outperformed by confidence-based approaches. 2) First comparison of aggregation approaches for general knowledge questions as well as idea evaluation.
Görzen and Kundisch (2019)	Development of a mechanism enabling the use of an anonymous external crowd to identify high quality ideas	Field Experiment & Qualitative Study (Think Aloud)	Business model ideas for perfume	How can an anonymous crowd on an online-platform be used to identify ideas of high quality?	1) Development of a new approach to identify high quality ideas. 2) Evidence of individuals adopting different cognitive evaluation processes depending on idea quality.
Görzen (2019a)	Investigating the influence of experience on the occurrence of the anchoring effect	Field Experiment	Business model ideas for perfume	Are more experienced people in the crowd less prone to the anchoring effect in their decision-making?	1) Experience does not decrease the probability to fall victim to the anchoring effect. 2) Experienced members in the crowd feel more confident about their decision than less experienced members, even when they are wrong (i.e. deviate from the experts' evaluation of idea quality).
Görzen (2019b)	Influence of task meaning on output quantity and creativity in crowd ideation	Field Experiment	Ideas for an app name	Does task meaningfulness improve crowd workers' performance in terms of output quantity and creativity?	Results indicate that higher task meaning neither positively affects output quantity of idea generation nor the creativity of generated ideas.

Recalling the process of crowd ideation presented in section 2.3, the four studies on crowd-based idea evaluation in this dissertation contribute to two different steps in the sub-process of idea evaluation, namely the task of idea evaluation itself as well as the aggregation. Finally, *Görzen (2019b)* contributes to the open call for ideas, since this paper investigates the influence of increased task meaning in the task description on the outcome of idea generation (Figure 5).

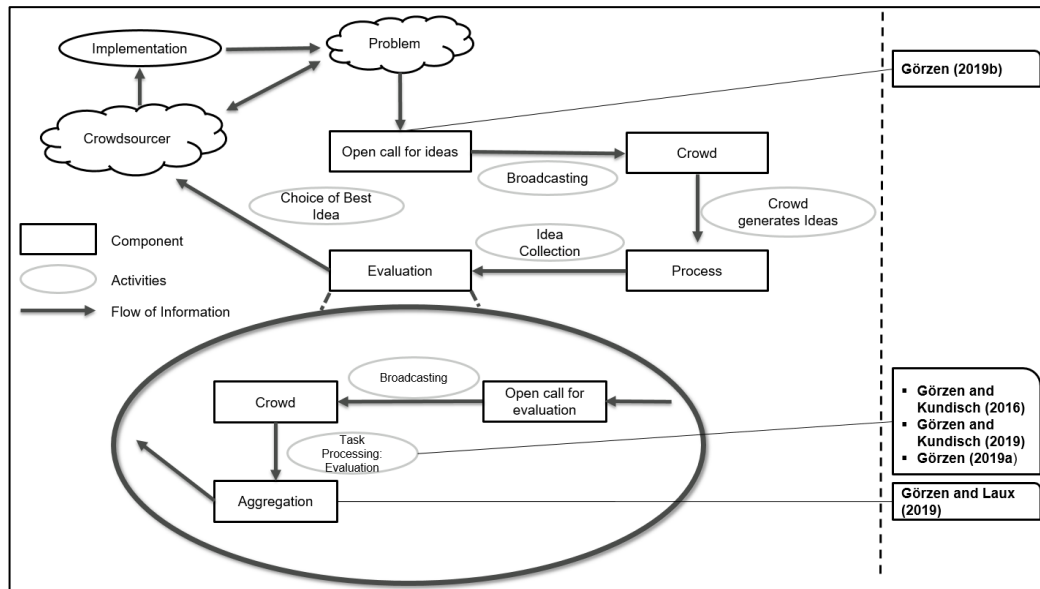


Figure 5. Contributions to Idea Evaluation Process

The following sub-sections provide a summary of each study of this cumulative dissertation as well as details of the scientific contributions and the dissemination of research findings.

4.1 Görzen and Kundisch (2016)

The first, co-authored paper, *Görzen and Kundisch (2016)* investigates whether the evaluation of a large number of business model ideas can match that of experts and whether the crowd has the potential to substitute experts. In contrast to previous literature, we use an anonymous online crowd on an external platform and ask them to evaluate ideas. Further, in contrast to relatively simple ideas, we asked the crowd to evaluate business model ideas, represented in form of the Business Model Canvas. Hence, the ideas are a relatively complex construct to evaluate since the elements of a business model are connected with each other and their interdependencies should be considered. We compared both the average rating for each idea as well as the relative rating by experts and by the crowd. Results indicate that an anonymous online crowd cannot evaluate business models to the same level as experts. Further, even more experienced crowd members, taking into account several dimensions of experience, were unable to evaluate ideas on a par with the experts.

Table 5. Görzen and Kundisch (2016) - Joint Work, Presentations, Scientific Dissemination

	Co-authorship with D. Kundisch (60 % T. Görzen, 40% D. Kundisch)
Joint Work	<ul style="list-style-type: none"> ▪ Concretization of the research question joint with D. Kundisch ▪ Literature review jointly with D. Kundisch ▪ Field study designed jointly with D. Kundisch ▪ Field study conducted by T. Görzen ▪ Statistical analysis performed by T. Görzen ▪ Write-up of paper jointly with D. Kundisch ▪ Write-up of the response to the ECIS and AMCIS reviewers and revision of the paper by T. Görzen ▪ Feedback, comments and corrections by D. Kundisch
Presentations	<ul style="list-style-type: none"> ▪ 06/2016: European Conference on Information Systems (ECIS), Istanbul, Turkey (Poster presentation) ▪ 08/2016: Americas Conference on Information Systems (AMCIS), San Diego, USA
Scientific Dissemination	<ul style="list-style-type: none"> ▪ Work on this paper started in 06/2015 ▪ A short version (Research-in-Progress) of this paper was published in the proceedings of the European Conference on Information Systems (ECIS) 2016 (VHB Jourqual 3 ranking: B) ▪ The full paper was accepted for publication in the proceedings of the Americas Conference on Information Systems (AMCIS) 2016 (VHB Jourqual 3 ranking: D)

4.2 Görzen and Laux (2019)

In the second, co-authored paper, *Görzen and Laux (2019)*, we compare different mechanisms for aggregating multiple opinions by a crowd. Since research has raised several concerns regarding the currently most popular mechanism - simple aggregating - we investigate other approaches for aggregating multiple opinions and compare them to each other. We conduct two different application cases, first asking the crowd to answer factual questions of general knowledge and second, asking the crowd to evaluate business model ideas. Our results confirm the concerns raised by previous literature since we identify other, confidence-based aggregation approaches, that provide better results.

Table 6. Görzen and Laux (2019) - Joint Work, Presentations, Scientific Dissemination

	Co-authorship with F. Laux (50 % T. Görzen, 50% F. Laux)
Joint Work	<ul style="list-style-type: none"> ▪ Concretization of the research question by T. Görzen ▪ Literature review by F. Laux ▪ Field study designed jointly with F. Laux ▪ Empirical analysis jointly with F. Laux ▪ Write-up of paper jointly with F. Laux ▪ Write up of the responses to the reviewers by T. Görzen ▪ Feedback, comments and corrections by T. Görzen and F. Laux
Presentations	<ul style="list-style-type: none"> ▪ 03/2018: Multikonferenz Wirtschaftsinformatik 2018, Lüneburg, Germany
Scientific Dissemination	<ul style="list-style-type: none"> ▪ Work on this paper started 08/2017 ▪ The paper was accepted for publication in the proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI) 2018 (VHB Jourqual 3 ranking: D)

4.3 Görzen and Kundisch (2019)

In the third, co-authored paper, *Görzen and Kundisch (2019)* we develop an approach that enables the use of crowd evaluation for the identification of high quality ideas. We first derive a general mechanism from previous literature using idea quality as the main characteristic influencing task complexity. We operationalize the need of individuals to look for additional information to reduce task complexity by displaying an anchor about previous ratings made by other individuals. Using the established anchoring effect and the decreasing effect on the variance of ratings when displaying an anchor enables us to use crowd evaluation for the identification of high quality ideas, since individuals incorporate the displayed anchor stronger into their own decision when evaluating ideas of high quality. We evaluate our new approach in different experimental conditions and analyze the results when displaying different values as anchors. Further, following a mixed method approach, we conduct an additional qualitative study, using the thinking aloud method. Results of this study are in line with our empirical study, indicating that idea quality influences task complexity, resulting in different evaluation processes applied by the participants of this study.

Table 7. Görzen and Kundisch (2019) - Joint Work, Presentations, Scientific Dissemination

Joint Work	<p>Co-authorship with D. Kundisch (70 % T. Görzen, 30% D. Kundisch)</p> <ul style="list-style-type: none"> ▪ Concretization of the research question jointly with D. Kundisch ▪ Literature review by T. Görzen ▪ Hypotheses development jointly with D. Kundisch ▪ Field experiment designed jointly with D. Kundisch ▪ Field experiment conducted by T. Görzen ▪ Statistical analysis performed by T. Görzen ▪ Write-up of paper jointly with D. Kundisch ▪ Qualitative study planned and conducted by T. Görzen and F. Rose (student assistant) ▪ Feedback, comments and corrections by D. Kundisch
Presentations	<ul style="list-style-type: none"> ▪ 02/2017: Workshop on IS Design and Economic Behavior, St. Gallen, Switzerland (presented by the co-author) ▪ 12/2017: International Conference on Information Systems (ICIS), Seoul, South Korea
Scientific Dissemination	<ul style="list-style-type: none"> ▪ Work on this paper started 09/2016 ▪ The paper was accepted for publication in the proceedings of the International Conference on Information Systems (ICIS) 2017 (VHB Jourqual 3 ranking: A) ▪ The paper was extended by the qualitative study and is currently under review at Business & Information Systems Engineering (VHB Jourqual 3 ranking: B)

4.4 Görzen (2019a)

In the fourth, single-authored paper *Görzen (2019a)*, I investigate whether more experienced judges in the crowd are less prone to a decision bias, namely the anchoring effect. Since online platforms, like crowdsourcing platforms, are characterized by increasing information richness and often provide information such as ratings given by other workers (Duan et al. 2009), individuals might fall victim to the anchoring effect. This could be the case especially if the evaluation process is structured sequentially, where results from previous evaluation rounds are displayed, thus potentially acting as anchors. Previous literature (e.g., Aspinall 2013) in this context proposed weighted models that favor more experienced judges in the crowd. The assumption behind this approach is that more experienced judges are less likely to be affected by the anchoring effect or, if they were affected, that their estimation

will still be more valid than that of a less experienced person. Results of this study indicate that not only does greater experience not protect crowd workers from the anchoring effect but it increases their own confidence in their decision, compared to less experienced individuals, even if they are wrong.

Table 8. Görzen (2019a) - Joint Work, Presentations, Scientific Dissemination

Joint Work	Single-authored paper
Presentations	<ul style="list-style-type: none">03/2018: Multikonferenz Wirtschaftsinformatik 2018, Lüneburg, Germany01/2019: 52nd Hawaii International Conference on System Sciences (HICSS), Hawaii, USA
Scientific Dissemination	<ul style="list-style-type: none">Work on this paper started 08/2017A short version (Research-in-Progress) of this paper was published in the proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI) 2018 (VHB Jourqual 3 ranking: D)The full paper was accepted for publication in the proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS) 2019 (VHB Jourqual 3 ranking: C)

4.5 Görzen (2019b)

The dissertation is rounded off with a fifth, single-authored paper, *Görzen (2019b)* which contributes to knowledge about crowd-based ideation. Previous literature in the offline context has shown a positive influence of high task meaning on task performance (Ariely et al. 2008; Kosfeld et al. 2014). High task meaning has also been found to have positively influenced crowd workers solving a number of tasks on MTurk (Chandler and Kapelner 2013). However, previous studies in the offline and the online context have mainly focused task performance in respect of output quantity (e.g., solved tasks). Hence, little is known about the influence of high task meaning on output quality, for example in terms of output creativity. In this paper, I investigate the influence of task meaning in crowd-based ideation by varying task meaning in two different experimental conditions, and analyze the influence of induced higher task meaning on output quantity as well as output creativity. Results indicate that higher task meaning has no positive influence on either the quantity or the creativity of the output, which carries practical implications for both crowdsourcer and crowdsourcing platform designers.

Table 9. Görzen (2019b) - Joint Work, Presentations, Scientific Dissemination

Joint Work	Single-authored paper
Presentations	<ul style="list-style-type: none"> ▪ 12/2017: International Conference on Information Systems (ICIS), Seoul, South Korea (Poster presentation)
Scientific Dissemination	<ul style="list-style-type: none"> ▪ Work on this paper started 02/2017 ▪ A short version (Research-in-Progress) of this paper was published in the proceedings of the International Conference on Information Systems (ICIS) 2017 (VHB Jourqual 3 ranking: A) ▪ The paper was extended and is currently under review at the International Journal of Innovation Management (VHB Jourqual 3 ranking: B)

5. Conclusion

This dissertation provides novel insights on the currently highly debated opportunities and limits of crowd ideation and more specifically, on crowd-based idea evaluation. The results show that engaging an anonymous external crowd on an online platform, using a simple average rating does not lead to rating results that are as good as those of experts (*Görzen and Kundisch 2016*). On that basis, I investigated several other facets of crowd-based idea evaluation aiming to find potential solutions on how to use an external crowd on a crowdsourcing platform to identify high quality ideas. Next, our results suggest that the currently most popular evaluation mechanism, the simple aggregating mechanism, was outperformed in two studies and therefore does not represent the best alternative to aggregating multiple opinions (*Görzen and Laux 2019*). The third study (*Görzen and Kundisch 2019*) provides a new approach on how to identify high quality ideas using an anonymous external online crowd. Results of an additional qualitative study indicates that the quality of an idea under consideration influences the evaluation process. Next, although previous literature suggests weighted evaluation mechanism that favor more experienced judges in the crowd, results (*Görzen 2019a*) indicate that, not only do more experienced judges not evaluate ideas as well as experts but their experience does not protect against the occurrence of decision making biases, more specifically the anchoring effect. Finally, the last study (*Görzen 2019b*) investigates the potential influence of task meaning as an alternative factor for motivating individuals in the crowd and increases task performance in terms of quantity and creativity. Results show that higher task meaning does neither positively influence output quantity nor creativity, however, higher task meaning seems to positively influence the perceived effort of working on a task.

5.1 Implications

Theoretical Implications

The results presented in this dissertation carry valuable theoretical and practical implications. First, results indicate that a crowd cannot simply be used to replace expert ratings (*Görzen and Kundisch 2016*) but, moreover, other ways to engage crowds in idea evaluation aimed at substituting experts should be investigated. Compared to previous studies which engaged companies' own employees, users or customers as crowds, individuals on an external anonymous crowdsourcing platform seem to evaluate ideas differently. Further, not only complex ideas with different interdependencies of several elements but also ideas that do not represent the final product, where the crowd represent potential end-users, seem to be more difficult to evaluate for an external crowd on a crowdsourcing platform than, say, household products or theater projects.

Results from *Görzen and Kundisch (2019)* offer several valuable insights from a theoretical perspective: first, this paper provides a new underlying mechanism of differences in individuals' idea evaluation, influenced by the quality of the idea under consideration. Second, results from an empirical as well as a qualitative study provide evidence for the existence of these cognitive differences in idea evaluation. We therefore provide new insights not only on crowd-based idea evaluation but also extend the literature on idea evaluation in general. Finally, previous literature has mainly regarded decision making biases, like the anchoring effect, as a negative phenomenon in decision making which should be avoided since it may lead to distorted decisions. However, the approach we developed is unique in integrating a decision-making bias, the anchoring effect, for a 'good' purpose since we use this bias by manipulating the decision making environment to nudge people's behavior, which enables us to identify high quality ideas. Lastly, results from *Görzen (2019b)* highlight that creative tasks on an external crowdsourcing platform represent a different type of task with different motivational factors influencing task performance, compared to typical routine tasks (e.g., image labeling) since task meaning neither positively influenced output quantity nor output creativity.

Practical Implications

Despite theoretical implications, the results of this dissertation also carry practical implications, for example for companies using crowdsourcing for idea generation and idea evaluation as well as for crowdsourcing platform operators.

First, results from *Görzen and Laux (2019)* indicate that when companies ask an external crowd to evaluate ideas they should also ask them to indicate how confident they are about their own evaluation. This enables companies to use different confidence-based evaluation. Next, companies could design an iterative evaluation mechanism using the evaluation mechanism developed in *Görzen and Kundisch (2019)*. Instead of using the absolute rating of an idea, companies could use an iterative approach using our mechanism to identify good ideas based on the reduction of variance. For example, after using the

crowd for an initial rating, the results of this rating could be used as an anchor in the next rating stage to analyze the reduction of variance as an indication of the (high) quality of an idea.

Finally, platform operators should be aware of the potential occurrence of decision making biases. Our results (*Görzen and Kundisch 2019; Görzen 2019a*) indicate that individuals on crowdsourcing platforms incorporate anchors, in the form of, for example the displayed rating from previous others, into their own decision-making process. Hence, platform designers should carefully consider which information should be displayed in the decision-making environment of the crowd.

5.2 Limitations and Future Research

Naturally, there are also limitations underlying the papers of my dissertation, which conversely also offer potential avenues for future research. First, I used the same research environment, more specifically, the same crowd-working platform (Figure Eight, previously known as Crowdfunder) for the five studies of this dissertation. This platform is characterized by a high diversity of individuals offering their work (Peer et al. 2017), however the crowd on this platform only represents one particular population. Next, we used business model ideas for perfume as the object for evaluation in four studies (apart from *Görzen 2019b*). These business model ideas represent a specific and complex set of ideas. However, in contrast to more simple ideas (e.g., product names or T-shirt designs), this type enables us to investigate potential boundaries of the crowd's ability to evaluate a complex idea with different elements (cost structure, key resources etc.) and different interdependencies inherent in the business model. Lastly, although this dissertation provides valuable results, the underlying theoretical mechanism remained unclear at certain points. For example, it is still unclear why task meaning has positively influenced output quantity for a routine task such as image labeling (Chandler and Kapelner 2013) but not the output quantity of a creative task like generating app names (*Görzen 2019b*).

The limitations outlined above also represent avenues for future research. Despite a large number of other potential avenues for future research, I would like to focus on three specific avenues relating to my dissertation:

- 1) **Applications to identify the best ideas:** Although *Görzen and Kundisch (2019)* already offers a potential approach on how to identify high quality ideas out of a large number of ideas, other ways to further support an external crowd to identify good ideas would be valuable for research and practice alike. While an online crowdsourcing/crowd working platform represents a specific environment with no fixed ties between the crowd and the crowdsourcing companies, brief training sessions on how to evaluate specific ideas to enable specific cognitive processes of idea evaluation could help crowd workers to identify good ideas. Additionally, comparisons of different, newly developed mechanisms to aggregate multiple opinions from a crowd could be carried out to further

increase the robustness of our results (*Görzen and Laux 2019*). For example, additional types of ideas could be used to investigate potential boundaries of specific aggregation approaches.

- 2) Differences in evaluation processes dependent on idea quality or the type of idea:** The results of our study (*Görzen and Kundisch 2019*) indicate that individuals apply different cognitive processes when evaluating ideas of different quality. Based on these initial results, a further investigation on cognitive processes, for example investigating differences in activated brain regions during the evaluation of ideas of varying levels of quality could be gainful, potentially extending existing theories of idea evaluation. Further insights on the ability of an external, anonymous crowd could be gained from comparing the evaluation of different types of ideas (e.g., T-shirt designs and business model ideas).

- 3) Investigating mechanisms to enhance creativity in crowd ideation:** Since previous results indicate that task meaning does not increase idea creativity in crowd evaluation and further, that financial incentives may even reduce creativity (*Erat and Gneezy 2016*), other ways to increase output creativity should be investigated. For example, stimulating the divergent ideation phase via specific task designs to open out the potential solution space for ideation could help individuals in the crowd to think about more innovative and creative solutions.

Appendix

A1. Keywords and Search Strings (adopted from Merz 2018)

crowd idea evaluation
open evaluation
crowd open evaluation
crowdsourcing evaluation
community innovation
crowd idea selection
crowd idea voting
crowd idea rating
quality AND idea AND crowd* AND (select* OR screen* OR shortlist* OR filter* OR review* OR evaluat* OR con-verg* OR eliminat* OR reduc* OR vot* OR jud* OR as-sess*)
idea AND crowd* AND (select* OR screen* OR shortlist* OR filter* OR review* OR evaluat* OR converg* OR elimi-nat* OR reduc* OR vot* OR jud* OR assess*) [in Anlehnung an Merz (2018)]
best idea AND crowd* AND (select* OR screen* OR shortlist* OR filter* OR review* OR evaluat* OR converg* OR eliminat* OR reduc* OR vot* OR jud* OR assess*)
crowd idea evaluation

A2. Further Publications not Part of this Thesis

Neumann, J., Gutt, D., Görzen, T. and Kundisch, D. 2019. "When does Local Status Matter? - The Relationship between Reviewer Location and Perceived Usefulness of Online Reviews." *Proceedings of the Twenty-fifth Americas Conference on Information Systems (AMCIS)*, Cancun, Mexico.

Poniatowski, M., Neumann, J., Görzen, T. and Kundisch, D. 2019. "Organizing Their Thoughts - How Online Review Templates Affect the Review Text." *Proceedings of the Twenty-Seventh European Conference on Information Systems (ECIS)*, Stockholm, Sweden.

Poniatowski, M., Neumann, J., Görzen, T. and Kundisch, D. 2019. "A Semi-Automated Approach for Generating Online Review Templates." *Proceedings of the 14th International Conference on Wirtschaftsinformatik*, Research-in-Progress, Siegen, Germany.

Görzen, T., Kundisch, D., Meier, C., Ottensmann, A. and Haubrock, J. 2014. "Research on Electric Mobility in Germany: Monitoring a Commercial Battery Electric Vehicle Fleet." *Proceedings of the European Electric Vehicle Congress (EEVC)*, Brussels, Belgium.

Görzen, T., Meier, C. and Kundisch, D. 2014. "Ein TVO-Modell zur Bewertung elektrifizierter Fahrzeugflotten." *Proceedings der INFORMATIK, Workshop "Elektromobilität, Agenten und Smart Grids" (EASG)*, Stuttgart, *Lecture Notes in Informatics (LNI)*, GI.

References

- Adamczyk, S., Bullinger, A. C. and Möslin, K. M. 2012. "Innovation Contests: A Review, Classification and Outlook." *Creativity and Innovation Management* (21), pp. 335-360.
- Amazon MTurk. 2019. <https://www.mturk.com/>. Accessed 05 June 2019.
- Ariely, D., Kamenica, E. and Prelec, D. 2008. "Man's Search for Meaning: The Case of Legos," *Journal of Economic Behavior & Organization* (67), pp. 671-677.
- Aspinall, W. 2013. "A Route to more Tractable Expert Advice," *Nature* (463:7279), pp. 294-295.
- Baumol W. J. 2002. "The Free-Market Innovation Machine," Princeton University Press, Princeton, NJ.
- Bayus, B. L. 2013. "Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community," *Management Science* (59:1), pp. 226-244.
- Bettman, J. R., Luce, M. F. and Payne, J. W. 1998. "Constructive Consumer Choice Processes," *Journal of Consumer Research* (25:2), pp. 187-217.
- Blohm, I., Leimeister, J. M. and Krcmar, H. 2013. "Crowdsourcing: How to Benefit from (Too) Many Great Ideas," *MISQ Executive* (12:4), pp. 199-211.
- Blohm, I., Riedl, C., Füller, J. and Leimeister, J. M. 2016. "Rate or Trade? Identifying Winning Ideas in Open Idea Sourcing," *Information Systems Research* (27:1), pp. 27-48.
- Boudreau, K., Lacatera, N. and Lakhani, K. 2011. "Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis," *Management Science* (57:5), pp. 843-863.
- Boudreau, K. J. 2012. "Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software App Developers and Patterns of Innovation," *Organization Science* (23:5), pp. 1409-1427.
- Boudreau, K. J. and Lakhani, K. R. 2013. "Using the Crowd as an Innovation Partner," *Harvard Business Review* April 2013, pp. 1-11.
- Brabham, D. C. 2010. "Moving the Crowd at Threadless: Motivations for Participation in a Crowdsourcing Application. Information," *Communication & Society* (13:8), pp. 1122-1145.
- Budescu, D. V. and Chen, E. 2015. "Identifying Expertise to Extract the Wisdom of Crowds." *Management Science* (61:2), pp. 267-280.

- Chandler, D. and Kapelner, A. 2013. "Breaking Monotony with Meaning: Motivation in Crowdsourcing Markets," *Journal of Economic Behavior & Organization* (90), pp. 123-133.
- Chiu C.-M., Liang, T.-P. and Turban, E. 2014. "What can Crowdsourcing do for Decision Support?," *Decision Support Systems* (65), pp. 40-49.
- Chen, L., Goes, P., Marsden, J. R. and Zhang, Z. 2009. "Design and Use of Preference Markets for Evaluation of Early Stage Technologies," *Journal of Management Information Systems* (26:3), pp. 45-70.
- Chen, D. L. and Horton, J. J. 2016. "Research Note - Are Online Labor Markets Spot Markets for Tasks? A Field Experiment on the Behavioral Response to Wage Cuts," *Information Systems Research* (27:2), pp. 403-423.
- Duan, W., Gu, B. and Whinston, A. B. 2009. "Informational Cascades and Software Adoption on the Internet: An Empirical Investigation," *MIS Quarterly* (33:1), pp. 23-48
- Durward, D., Blohm, I. and Leimeister, J. M. 2016. "Crowd Work," *Business & Information Systems Engineering* (58:4), pp. 281-286.
- Durward, D. and Blohm, I. 2017. "The Rise of Crowd Aggregators - How Individual Workers Restructure Their Own Crowd." *Proceedings of the 13th International Conference on Wirtschaftsinformatik*, St. Gallen, Switzerland.
- Ebel, P. A., Bretschneider, U. and Leimeister, J. M. 2016. "Can the Crowd Do the Job? Exploring the Effects of Integrating Customers into a Company's Business Model Innovation," *International Journal of Innovation Management* (20:7), 1650071-1-1650071-24.
- Erat, S. and Gneezy, U. 2016. "Incentives for Creativity," *Experimental Economics* (19), pp. 269-280.
- Feyisetan, O., Simperl, E., Van Kleek, M. and Shadbolt, N. 2015. "Improving Paid Microtasks through Gamification and Adaptive Furtherance Incentives," *Proceedings of the 24th International Conference on World Wide Web*, Florence, Italy.
- Galati, F. 2015. "Complexity of Judgement: What Makes Possible the Convergence of Expert and Nonexpert Ratings in Assessing Creativity," *Creativity Research Journal* (27:1), pp. 24-30.
- Girotra, K., Terwiesch, C. and Ulrich, K. T. 2010. "Idea Generation and the Quality of the Best Idea," *Management Science* (56:4), pp. 591-605
- Goldenberg, J., Lehmann, D. R. and Mazursky, D. 2001. "The Idea Itself and the Circumstances of its Emergence as Predictors of New Product Success," *Management Science* (47:1), pp. 69-84.

- Görzen, T. 2019a. "Can Experience be Trusted? Investigating the Effect of Experience on Decision Biases in Crowdfunding Platforms," *Working Paper*.
- Görzen, T. 2019b. "What's the Point of the Task? Exploring the Influence of Task Meaning on Creativity in Crowdsourcing," *Working Paper*.
- Görzen, T. and Kundisch, D. 2016. "Can the Crowd Substitute Experts in Evaluation of Creative Ideas? An Experimental Study Using Business Models," *Proceedings of the 22nd Americas Conference on Information Systems*, San Diego, USA.
- Görzen, T. and Kundisch, D. 2019. "When in Doubt Follow the Crowd: How Idea Quality Moderates the Effect of an Anchor on Idea Evaluation," *Working Paper*.
- Görzen, T. and Laux, F. 2019 "Extracting the Wisdom from the Crowd: A Comparison of Approaches to Aggregating Collective Intelligence," *Working Paper*.
- Hammon, L. and Hippner, H. 2012. "Crowdsourcing," *Business and Information Systems Engineering* (4:3), pp. 163-166.
- Howe, J. 2008. "Crowdsourcing. Why the Power of the Crowd is Driving Future of Business," *Three Rivers Press*, New York.
- Hrastinski, S., Kviselius, N. Z., Ozan, H., and Edenius, M. 2010. "A Review of Technologies for Open Innovation: Characteristics and Future Trends," *Proceedings of the Annual Hawaii International Conference on System Sciences*, Hawaii, USA.
- Jarke, J. 2017. "Community-Based Evaluation in Online Communities," *Information Technology & People* (30:2), pp. 371-395.
- Jaruzelski, B. and Dehoff, K. 2010. "The Global 1000: How the Top Innovators Keep Winning," *Strategy and Business* (61), pp. 1-14.
- Klein, M. and Garcia, A. C. B. 2015. "High-Speed Idea Filtering with Bag of Lemons," *Decision Support Systems* (78), pp. 39-50.
- Kornish, L. J. and Ulrich, K. T. 2014. "The Importance of the Raw Idea in Innovation: Testing the Sow's Ear Hypothesis," *Journal of Marketing Research* (51:1), pp. 14-26.
- Kosfeld, M., Neckermann, S. and Yang, X. 2014. "Knowing that You Matter, Matters! The Interplay of Meaning, Monetary Incentives, and Worker Recognition," *ZEW - Centre for European Economic Research Discussion Paper* 14-097.
- Kosonen, M. G. C., Vanhala, M. and Blomqvist, K. 2014. "User Motivation and Knowledge Sharing in Idea Crowdsourcing," *International Journal of Innovation Management* (18:5), pp. 1-23.

- Kuek, S. C., Paradi-Guilford, C., Fayomi, T., Imaizumi, S., Ipeirotis, P., Pina, P. and Singh, M. 2015. "The Global Opportunity in Online Outsourcing," Washington: World Bank.
- Lee, H. C. B., Ba, S., Li, X., and Stallaert, J. 2018. "Salience Bias in Crowdsourcing Contests," *Information Systems Research* (29:2), pp. 401-418.
- Leimeister, J. M., Huber, M., Bretschneider, U. and Krcmar, H. 2009. "Leveraging Crowdsourcing: Activation-Supporting Components for IT-based Ideas Competition," *Journal of Management Information Systems* (26:1), pp. 197-224.
- Mack, T., and Landau, C. 2015. "Winners, Losers, and Deniers: Self-Selection in Crowd Innovation Contests and the Roles of Motivation, Creativity, and Skills," *Journal of Engineering and Technology Management* (37), pp. 52-64.
- Magnusson, P. R., Wästlund, E., and Netz, J. 2016. "Exploring Users' Appropriateness as a Proxy for Experts When Screening New Product/Service Ideas," *Journal of Product Innovation Management* (33:1), pp. 4-18.
- Mason, W. A. and Watts, D. J. 2009. "Financial Incentives and the Performance of Crowds," *Association for Computing Machinery Explorations Newsletter* (11:2), pp. 100-108.
- Merz, A. B. 2018. "Mechanism to Select Ideas in Crowdsourced Innovation Contests - A Systematic Literature Review and Research Agenda," *Proceedings of the 26th European Conference on Information Systems*, Portsmouth, UK.
- Mollick, E. and Nanda, R. 2016. "Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts," *Management Science* (62:6), pp. 1533-1553
- Onarheim, B. and Christensen, B. T. 2012. "Distributed Idea Screening in Stage-Gate Development Processes," *Journal of Engineering Design* (23:9), pp. 660-673.
- Palacios, M., Martinez-Corral, A., Nisar, A. and Grijalvo, M. 2016. "Crowdsourcing and Organizational Forms: Emerging Trends and Research Implications," *Journal of Business Research* (69:5), pp. 1834-1839.
- Peer, E., Brandimarte, L., Samat, S. and Acquisti, A. 2017. "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research," *Journal of Experimental Social Psychology* (70), pp. 153-163.
- Poetz, M. K., and Schreier, M. 2012. "The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?," *Journal of Product Innovation Management* (2:29), pp. 245-256.

- Riedl, C., Blohm, I., Leimeister, J. M. and Krcmar, H. 2013. "The Effect of Rating Scales on Decision Quality and User Attitudes in Online Innovation Communities," *International Journal of Electronic Commerce* (17:3), pp. 7-36
- Rosen, P. A. 2011. "Crowdsourcing Lessons for Organizations," *Journal of Decision Systems* (20:3), pp. 309-324.
- Schemmann, B., Herrmann, A. M., Chappin, M. M. H. and Heimeriks, G. J. 2016. "Crowdsourcing Ideas. Involving Ordinary Users in the Ideation Phase of New Product Development," *Research Policy* (45:6), pp. 1145-1154.
- Schuurman, D., Baccarne, B., De Marez, L. and Mechant, P. 2012. "Smart Ideas for Smart Cities: Investigating Crowdsourcing for Generating and Selecting Ideas for ICT Innovation in a City Context," *Journal of Theoretical and Applied Electronic Commerce Research* (7:3), pp. 49-62.
- Soukhoroukova, A., Spann, M. and Skiera, B. 2012. "Sourcing, Filtering, and Evaluating New Product Ideas: An Empirical Exploration of the Performance of Idea Markets," *Journal of Product Innovation Management* (29:1), pp. 100-112.
- Toubia, O. 2006. "Idea Generation, Creativity, and Incentives," *Marketing Science* (25:5), pp. 411-425.
- Toubia, O. and Flores, L. 2007. "Adaptive Idea Screening Using Consumers," *Marketing Science* (26:3), pp. 342-360.
- vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R. and Cleven, A. 2015. "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Synopsis Search in Information Systems Research," *Communications of the Association for Information Systems* (37:9), pp. 205-224.
- Wagenknecht, T., Crommelinck, J., Teubner, T. and Weinhardt, C. 2017. "Ideate. Collaborate. Repeat. A Research Agenda for Idea Generation, Collaboration and Evaluation in Open Innovation," *Proceedings 13th International Conference on Wirtschaftsinformatik*, St. Gallen, Switzerland.
- Webster, J. and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:2), pp. xiii-xxiii.
- West, J. and Bogers, M. 2013. "Leveraging External Sources of Innovation: A Review of Research on Open Innovation," *Journal of Product Innovation Management* (31), pp. 814-831.
- Yoo, Y., Boland, R. J., Lyytinen, K. and Majchrzak, A. 2012. "Organizing for Innovation in the Digitized World," *Organization Science* (23:5), pp. 1398-1408.

Zhu, J. J., Li, S. Y. and Andrews, M. 2017. "Ideator Expertise and Cocreator Inputs in Crowdsourcing-Based New Product Development," *Journal of Product Innovation Management* (34:5), pp. 598-616.

Zogaj, S., Bretschneider, U., and Leimeister, J. M. 2014. "Managing Crowdsourced Software Testing: A Case Study based Insight on the Challenges of a Crowdsourcing Intermediary," *Journal of Business Economics* (84), pp. 375-405.