UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

FACULTY OF BUSINESS ADMINISTRATION AND ECONOMICS

Working Paper Series

Working Paper No. 2019-10

“What’s the Point of the Task?” Exploring the Influence of Task Meaning on Creativity in Crowdsourcing

Thomas Görzen

June 2019

“What’s the Point of the Task?” Exploring the Influence of Task Meaning on Creativity in Crowdsourcing

Thomas Görzen

University of Paderborn, Germany
thomas.goerzen@upb.de

Abstract

In pursuit of product innovation, companies increasingly use crowdsourcing for idea generation. Apart from financial motives, previous studies on the motivation of crowd workers identified intrinsic motivation, such as task meaningfulness, to play a key role. These studies, however, focused on routine tasks such as image labelling, using output quantity as the variable of interest. Since the primary goal for a creative task such as idea generation is not abundance but high quality, we investigate the influence of task meaning on workers’ effort and on output creativity. In a field experiment involving idea generation by an online crowdsourcing platform, we vary the task meaning in two different conditions, and evaluate the quality of the ideas generated. We find that higher task meaning has no positive influence on either the quantity or the creativity of the output, which carries practical implications for both commissioners and crowdsourcing platform designers.

Keywords: Crowd working, Task meaning, Creativity, Crowdsourcing

Introduction

Man’s search for meaning is the primary motivation in his life. - Frankl, 1984

Aiming for product innovation to satisfy consumer demand, companies increasingly use crowdsourcing to generate new ideas for products or services. Crowdsourcing is defined as “the act of taking a task once performed by an employee and outsourcing it to a large, undefined group of people external to the company in the form of an open call” (Howe, 2008). With millions of workers spread across the globe, crowdsourcing offers huge potential to generate new ideas at minimal cost. The benefits of collective creativity include the speed with which ideas can be generated and the diversity of idea generators, all of which can increase idea quality at a low cost compared to more conventional approaches such as drawing on internal resources (Kittur, 2010; Füller et al., 2014).

Previous studies (e.g., Poetz and Schreier, 2012; Bayus, 2013; Huang et al., 2014) have shown the potential of crowds to generate valuable ideas. However, most of these studies examined settings in which companies recruited their own users or customers for idea generation, a process which is more costly than external crowdsourcing, both in terms of time and money. Not only does recruiting and motivating their own staff or customers involve setting up a dedicated online-forum and designing and implementing motivational strategies, but it also has more limited potential if a company lacks a well-known brand name or a sufficiently large customer base. External crowdsourcing platforms, however, overcome these hurdles by offering an almost unlimited number of potential contributors who can work on different tasks in return for a relatively small amount of money. The downside of this is that the impersonal nature of these platforms makes it harder for a company to build ties with these workers, who typically act anonymously, using artificial user names, or shifting from one job to another (Kittur et al., 2013). Another common feature of crowdsourcing platforms such as Amazon MTurk is that the tasks they offer are usually cognitively low demanding routine tasks like image labeling, audio transcription or basic internet research (Chandler and Kapelner, 2013) - which differs significantly from the cognitively more demanding task that is idea generation. Typically, on those platforms workers are paid on a “piece rate system” for these mundane tasks, with wages linked to output quantity. One might assume that the primary driver for motivation and effort invested by workers would be the level of financial incentive offered for a given task. However, previous

literature indicates that higher payment does not necessarily lead to an increase in effort (Mason and Watts, 2010). In fact, in the context of creative tasks, financial incentives may even reduce creativity (Erat and Gneezy, 2016). Results of prior research also highlight the influence of intrinsic motivation to participate in crowdsourcing and to solve tasks (Erat and Gneezy, 2016; Kaufman et al., 2011; Zheng et al., 2011). Compared with performing routine tasks, the creativity involved in idea generation is a cognitively much more demanding task (Mumford and Gustafson, 2007) which involves a divergent and a convergent thinking phase (Mumford et al., 1991). Idea creativity rather than quantity is, however, what matters most to companies when it comes to idea generation. Indeed, most companies would prefer to get one outstandingly good idea and 99 bad ones rather than 100 average ones (Girotra et al., 2010). They therefore face the challenge of how to motivate anonymous participants on a crowdsourcing platform to ensure that they deliver highly creative work (Kittur et al., 2013). This leads us to the question of how to positively enhance the quality of workers’ output on a crowdsourcing platform and, more specifically, how to motivate them to generate more creative ideas.

Apart from financial motives, previous studies on the motivation of crowd workers have emphasized the role played by intrinsic motivation, derived from - amongst others - the meaningfulness of a task. Task meaningfulness has been linked to the level of effort people invest in solving a task, whether in an offline (Ariely et al., 2008) or an online context (Chandler and Kapelner, 2013). However, previous studies mainly focus on quantity or accuracy rather than on the creativity level of the output. Accordingly, our study seeks to shed light on this issue by focusing on the following research question:

Does task meaningfulness improve crowd workers’ performance in terms of output creativity?

To address this question, we conduct an experiment on an online crowdsourcing platform in which we asked an anonymous online crowd to generate app names for a specific app. This experiment adopted two conditions, one low and one high meaning, with a task description pointing out the meaningfulness of each task. We then returned to the crowdsourcing platform to evaluate and compare the creativity of the ideas generated in both conditions. To measure our variables of interest – output quantity and creativity – we count the number of generated ideas per subject in each experimental condition. To compare the creativity of outputs we use a set of metrics, including idea creativity, the number of good ideas¹ and identify the overall best rated idea.

Our work makes several contributions to the literature and carries important practical implications. First, we contribute to an enhanced understanding of the motivation of crowd workers since their motivation is not only based on monetary reward but incorporates a broader set of motivations such as task meaning, recognition, or the feeling of contributing towards a greater good (Kittur et al., 2013). Next, to the best of our knowledge, we are the first to investigate the effect of task meaningfulness in the context of crowdsourcing not only in terms of quantity but, crucially, in terms of its effect on the creativity level of outputs. On the practical level, our results offer useful insights for commissioners of crowdsourcing platforms as well as for platform designers on how to motivate an anonymous crowd on an online-platform to generate creative ideas. Instead of simply increasing the monetary incentive, which for many companies would be the most obvious approach, we explore the role of task meaning as a way of enhancing idea creativity.

Related Literature

Factors Influencing Creativity

Researchers interested in creativity commonly focus on idea development, but there are other aspects worthy of investigation that are relevant to our question, including how to measure creativity, what drives creativity, and how to stimulate creativity in individuals. For example, some researchers focus on the creative process, including the components of problem finding, of development and evaluation (Davis, 2009), while others focus on a person’s creativity as measured by tests such as the Torrance Test of Creative Thinking (Torrance, 1974), for example. In addition, some studies (e.g., Silva, 2008) investigate the

¹Following Jung et al., 2010; Poetz and Schreier, 2012: An idea rated above five on a seven-point Likert scale

relationship between intelligence and creativity, indicating a modest relation between the two constructs. Despite other definitions, since our study’s main focus is output creativity, we use the established definition of creativity in terms of a novel and useful outcome, whether it be products, services or, more generally, ideas (Amabile, 1996; Baer et al., 2010). Prior research identified a number of forces which foster creativity. These forces can be seen as a system operating at different levels, including the neurological level (e.g., where creative capability is located in the human brain), the cognitive level (e.g., abstracting from specific brain locations, how creativity is created subconsciously), the individual level (e.g., how creativity can be promoted consciously, for example, through certain creativity techniques), the group level (e.g., what group composition makes a group creative) and higher levels of analysis such as organizations and societies (Amabile and Hennessey, 2010). These forces and their influence on creativity have been investigated from various perspectives. One stream of literature in this field focused on different creativity techniques, with different underlying mechanisms such as the division of the problem into basic parts, or the combination and integration of established solutions (Meinel and Voigt, 2017). For example, one literature review encompassing 77 empirical articles published between 1966 and 2015 revealed that brainstorming, including different specifications of it, is the most commonly used creativity technique (Meinel and Voigt, 2017). However, while on average brainstorming leads to a relatively large number of ideas compared to techniques such as the six thinking hats method, brainsketching led to better idea quality (van der Lugt, 2002). Another stream of literature focused on personal characteristics associated with creative achievement (Martindale, 1989). These studies demonstrate that personal characteristics such as broad interests, attraction to complexity, toleration of ambiguity and self-confidence are positively associated with creative performance in a variety of domains (Martindale, 1989; Gough, 1979). Another study suggests that lower levels of self-confidence may predict higher levels of creativity (Kaufman, 2002), thus calling in question the relationship between self-confidence and creativity.

Another stream of literature examined the effects of contextual factors such as goals and motivation on individuals’ creative performance (Oldham and Cummings, 1996), highlighting the importance of intrinsic motivation (Amabile, 1996). Here it is argued that individuals are expected to be at their most creative when experiencing a high level of intrinsic motivation for a given task, since in these circumstances they are excited by the activity and feel free to explore new cognitive pathways (Amabile et al., 1990).

Worker Motivation in Crowdsourcing

In general, to be motivated to do something could be defined as *to be moved* by it (Deci and Ryan, 2000). Most theories of work motivation differentiate between two types of motivation, intrinsic and extrinsic. Intrinsic motivation is defined as undertaking an activity for the inherent satisfaction it brings rather than because of external incentives (Deci and Ryan, 2000). In contrast, extrinsic motivation occurs when a reward, such as money, is associated with a specific task (Deci and Ryan, 2000). While neither of these types of motivation play an important role in an offline-labor context, they both do in online labor markets such as online crowdsourcing platforms. While a traditional, rational economic approach would suggest that higher extrinsic motivation such as a higher payment for the completion of a task leads to higher work quality (Prendergast, 1999), previous results indicate that this connection is not quite as obvious in the context of crowdsourcing. For example, an experiment on a crowdsourcing platform where the crowd had to solve two different tasks – sorting car traffic images and finding words in a puzzle – led to the result that whilst their efforts increased with an increase in payment, the wage rate had no effect on improving their accuracy in solving the tasks (Mason and Watts, 2010). Moreover, a survey study with 431 participants on a crowdsourcing platform, designed to investigate the motivation of the crowd, also led to the result that while extrinsic motivation such as payoffs had an effect on time spent on the platform, intrinsic motivational factors such as enjoyment of the tasks was more important (Kaufman et al., 2011). Similarly, in the context of idea generation contests, two independent studies on crowdsourcing platforms in China (Zheng et al., 2011) and Germany (Leimeister et al., 2009) found that material compensation was a less powerful driver for participation in idea competitions than intrinsic motivation. Similarly, again, a study (Wolf et al., 2011) undertaken in the context of so-called games with a purpose (GWAP) – a sub-category of crowdsourcing (Thaler et al., 2012) – corroborated the greater role played by non-monetary motives for participation, with enjoyment and sociability having acted as key drivers for participating in semantic quiz games. According to these results, extrinsic motivation, then, only affects efforts related to result quantity to some extent, while the importance of intrinsic motivation in the context of crowdsourcing has been highlighted by several studies. These results do not seem to be surprising, however, since jobs on crowdsourcing platforms such

as Amazon Mechanical Turk or Crowdfunder² tend to offer low wages (Ipeirotis, 2010). Our study, therefore, focuses on the role played by intrinsic motivation and, more specifically, task meaning, to find out whether workers’ creativity on crowdsourcing platforms can be stimulated by higher task meaning.

Task Meaningfulness in Crowdsourcing

One important motivational factor for workers’ performance that has received relatively little attention so far is the meaningfulness of a task as perceived by the person performing it (Kosfeld et al., 2014). Literature on the impact of meaning on behavior (Table 1) is relatively sparse, especially in the context of economics and the influence of meaning on workers’ performance (Ariely et al., 2008). ‘Meaning’ in this context is defined as the extent to which a job is “(a) recognized and/or (b) has some point or purpose” (Ariely et al., 2008). Previous literature on the influence of task meaning on worker performance in both offline and online contexts underlines the importance of this motivational factor. In the offline context, for example, a field experiment with students working in a research center in China revealed that workers apply greater effort when task meaning is high (Kosfeld et al., 2014). Two groups of students were randomly assigned to each group and asked to file survey data into an electronic database while the instructions for the task were varied for both groups to create low and high meaning conditions. The effect of meaning also had a greater statistical effect than monetary incentives (Kosfeld et al., 2014). Similar results were found in laboratory experiments (Ariely et al., 2008). When varying the meaningfulness of two different tasks - (1) assembling Legos, and (2) finding ten consecutive letters from a sheet of random letters - results show that the performance was significantly greater in the more meaningful condition than in the ignored condition (Ariely et al., 2008). Further, subjects exhibited the lowest average reservation wage in the meaningfulness condition. Field experiments on task meaning in different contexts revealed similar results, since greater task meaning increased the performance of fundraisers and the dedication and helping behavior of lifeguards (Grant, 2008).

Table 1. Studies on Influence of Meaning on Task Performance

Article	Aim of the study	Subjects	Task	Result
Ariely et al. (2008)	To investigate how perceived meaning influences labor supply	Students in a laboratory setting	Assembling Bionicle Lego models	Meaningful task increases productivity, while subjects in less meaningful conditions exhibit much higher reservation wages
Kosfeld et al. (2014)	To investigate the effect of meaning on workers’ performance and interaction with monetary rewards and worker recognition	Students in China	Data entry job	Meaning increases workers’ performance and outperforms monetary rewards, while meaning and recognition have largely similar effects but interact negatively.
Chandler and Kapelner (2013)	To explore the relationship between the “meaningfulness” of a task and worker effort.	Crowd worker on Amazon MTurk	Labeling medical images	Subjects in the meaningful framed task significantly increased the quantity of the output with an insignificant change in output quality
Moussawi and Koufaris (2013)	To investigate crowdsourcees’ perceived task characteristics and their association with both perceived performance and perceived effort	Students on an own, simulated crowdsourcing platform	Microtasks focusing on classification and ownership identification	Positive association between perceived task meaningfulness and perceived performance and effort ³
Grant (2008)	To investigate whether intrinsic motivation moderates the association between prosocial motivation and performance.	Firefighters and fundraising callers	Overtime hour of firefighters and fund raising	Firefighters who reported high levels of intrinsic motivation averaged more overtime hours while fundraiser with high intrinsic motivation raised more money

² The platform has been renamed to FigureEight

³ These results represent only a part of the overall results of this study.

The literature exploring the influence of task meaningfulness on performance in crowdsourcing is nascent but growing. One study that developed and tested a theoretical model to investigate the influence of motivational factors (Moussawi and Koufaris, 2013) showed a positive association between higher task meaning and workers’ involvement as well as between perceived meaningfulness of the task on perceived effort and their performance (Moussawi and Koufaris, 2013). Another study used a field experiment involving 2,500 workers on a crowdsourcing platform asked to label medical images, to investigate task meaning on the quantity and accuracy of workers’ performance. The purpose of the tasks allocated to the three groups varied, with the instructions for the subjects in the meaningful treatment being told they were labeling tumor cells to assist medical research, while subjects in the control group (“zero-context”) were not told the purpose of the task and, in stark contrast, subjects in the third group (“shredded treatment”) were told that their work would be disregarded (Chandler and Kapelner, 2013). Results show that in the more meaningful task framing, workers were more likely to participate, and that the meaningful treatment increased the quantity (number of images labeled) as well as the quality (accuracy of image labelling) of the output (Chandler and Kapelner, 2013). In contrast, the shredded treatment decreased the quality of output with no change in quantity (Chandler and Kapelner, 2013).

While previous literature investigated the influence of task meaning on task performance, both in the offline and in the online context, the main focus has always been output quantity and accuracy, whereas to the best of our knowledge no previous study has so far investigated the influence of task meaning on output creativity in crowdsourcing. While the number of companies using crowdsourcing for idea generation to gain competitive advantage is growing (Kosonen et al., 2014), little is known about how to positively influence the creativity of workers on crowdsourcing markets. Therefore, our aim is to shed light on this question by investigating the influence of task meaning on creativity in crowdsourcing markets.

Hypotheses

Prior research on workers’ motivation and performance suggests that intrinsic motivation plays an important role. One job characteristic which influences the interconnection between meaning and motivation is the experienced meaningfulness of a task to develop internal work motivation (Hackman and Oldham, 1976). Meaning itself is defined as “the output of having made sense of something, or what it signifies; as in an individual interpreting what her work means, or the role her work plays” (Pratt and Ashforth, 2003). Previous studies in the offline (Ariely et al., 2008; Grant, 2008) and the online context (Chandler and Kapelner, 2013), for example in crowdsourcing markets, suggest that people increase the quantity or their performance output when they perceive higher task meaning. Therefore, our first hypothesis reads as follows:

Hypothesis 1: A higher task meaning increases the performance of workers in crowdsourcing markets in terms of higher output quantity.

Despite focusing primarily on output quantity, prior studies mainly investigated the influence of task meaning on workers’ output quality in terms of accuracy. This metric makes sense in the context of routine tasks in crowdsourcing, such as image labeling where the number of correctly labeled images can be determined. By contrast, in the context of idea generation, it makes little sense to measure idea quality in terms of accuracy since there are no “true values” or “correct answers” to compare the results with. Therefore, in this context, workers’ output quality can only be measured by the quality of the idea generated, that is, in terms of how creative and relevant the generated idea is to solve the given problem.

Generally, perceived meaningfulness represents one of the core characteristics of a task, and is a necessary psychological component of intrinsic work motivation (Hackman and Oldham, 1980). First results indicate that worker’s higher perceived meaning of a task increases the intrinsically motivated effort (Kosfeld et al. 2014). Moreover, previous literature suggests that intrinsic motivation is conducive to creativity (Amabile, 1996; Hennessey and Amabile, 2010). According to the intrinsic motivation theory of creativity (e.g. Amabile, 1996), higher intrinsic motivation enables people to search for new and better ways of doing things, which is likely to result in high levels of creativity (Amabile, 1996). For example, Shin and Zhou (2003) found that intrinsic motivation of Korean employees working in the high-tech sector was

significantly and positively correlated with creativity (Shin and Zhou, 2003). Another study analyzed survey data from 165 employees and their supervisors from a research and development department of a large U.S. organization. Results of the supervisor-rated creativity revealed that employees’ intrinsic motivation was one fundamental antecedent for creativity (Dewett, 2007). As stated above, individuals are expected to be at their most creative when they experience a high level of intrinsic motivation for the given tasks (Amabile et al., 1990). When perceiving high task meaning, we assume intrinsic motivation to be high, since people understand what purpose their effort serves and what role their contribution plays. Thus, we formulate our second hypothesis:

Hypothesis 2: A higher task meaning increases the quality of workers’ output in crowdsourcing markets in terms of generating more creative ideas.

Summarizing, we expect to find a moderating effect of higher task meaning on the performance of a crowd in the context of idea generation in terms of both quantity and creativity of output (see Figure 1).

Experimental Design

Idea Generation

Consistent with previous studies (e.g., Chandler and Kapelner, 2013), we conduct a field experiment on an online-crowdsourcing platform to test our hypotheses. In order to avoid experimenter demand effects we designed a task that appears natural on a crowdsourcing platform (Zizzo, 2010). The different types of creativity tasks used in previous studies can be grouped into three broad domains: artistic creativity (e.g. line drawings, paintings and other forms of artistic media), problem solving (e.g. desert island survival problem) and verbal creativity (e.g. stories or word compounds) (Amabile, 1996).

We decided to design a verbal creativity task which asks the crowd on a crowdworking-platform to generate ideas for names for a specific app. This task is creative, because (1) it has little structure (there is no single correct solution), (2) it involves divergent thinking in the sense that a number of different ideas for app names need to be created, which later in the task are shortlisted by convergent thinking, (3) it involves cyclic processing by a single person (if a given set of generated ideas does not contain a satisfactory solution, the person needs to engage in another cycle of divergent and convergent thinking), and (4) it involves information combination and reorganization (John and Kundisch, 2015). Creative tasks significantly differ from routine tasks which, in contrast, only involve convergent thinking, i.e. where there is only one correct answer/solution that can only be reached by one route and with clearly available inputs (John and Kundisch, 2015).

The aim of the app is to inform people who suffer from lactose intolerance about new lactose-free products available in grocery stores, with information about product ingredients and pictures. Lactose intolerance affects about 15% of the population⁴, or about ten million people in the German-speaking countries (Germany, Switzerland and Austria) in which the contributors on the crowdworking platform were based.

⁴ <https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072452/>

Although given an identical task, we manipulated the framing of the task to generate two different conditions (Figure 2): a higher task meaning was induced by i) describing the superordinate purpose of the task in more detail, and ii) indicating the potential social impact of the developed solution. In this context, purpose means that contributors understand how their work might be linked to an objective (Ariely et al., 2008), while social impact is defined as a significant improvement to people’s well-being (Dietz, 1987). In both experimental conditions the general features of the app were described equally. However, subjects in the *low meaning* condition were told that the app developer already had a favorite name for the app but that the crowd should search for additional app names. Hence, we decreased the recognition of the work by pointing out that their work might actually be futile, since the app developer already had a favorite name. In contrast, subjects in the *high meaning* condition were told that their help is needed to find an app name and that their input would make life easier for lactose-intolerant people. Thus, we increased the social impact of the task by highlighting the superordinate purpose. All other aspects of the task were designed equally for both conditions. To make sure that subjects would read the instructions properly and not just skip them, we implemented a countdown of 30 seconds so that subjects were only able to go on to the main screen of the task after the countdown. Subjects were informed about the countdown but the time was not displayed. After 30 seconds, a button to proceed to the task after reading the instructions appeared. Subjects in both conditions could enter as many ideas for app names as they wanted, but had to submit at least one. Thus, each contributor could freely determine the number of ideas she wanted to submit and, as no additional monetary bonus was offered for each idea provided, there was no incentive to submit numerous ideas.

Access to the idea generation task was limited to contributors from German speaking countries, i.e. Germany, Switzerland and Austria, since the target market of the app are primarily German speaking countries. Instructions were therefore presented in German. When subjects on the platform had decided to take on the task, they were randomly assigned to either the low or the high meaning condition. Assignment to both conditions was not possible because one subject could not assign themselves multiple times to the same task. We controlled the unique worker ID of each subject provided by the crowdsourcing platform after the experiment was finished and did not find any subject who worked on the task more than once. Before the main task of generating the app name was started, we asked each subject to enter demographic information about their age and gender. Following Chandler and Kapelner (2013), we asked some control questions after participants had completed the name generation task, measured on a seven-point Likert scale from 1 = “I don’t agree at all” to 7 = “I absolutely agree”. To check whether contributors had adequately considered these questions we applied reverse scoring (e.g. ‘I enjoyed working on the task’ was changed to ‘I did not enjoy working on the task’), after which all results remained qualitatively unchanged. Our control questions asked subjects whether they had enjoyed performing the job, whether they perceived high meaning compared to other jobs available on the platform, whether they felt good performing the job and how much effort they put into solving the task. Contributors in both conditions earned a fixed wage of 0.15€. In total, 144 subjects took part in the experiment, 71 in the low meaning condition (of which 15 were

female) and 73 in the high meaning condition (of which 24 were female). The low meaning condition generated 146 ideas for app names, while 154 names were generated in the high meaning condition⁵.

Idea Evaluation

To evaluate the ideas generated for the app names, we offered an additional task on the crowdsourcing platform. Following Dahan and Hauser (2001), we designed blocks of ten app names each because it would be unreasonable to ask each contributor to evaluate all app names in a row. We randomly assigned each generated app name to one block of names. Next, we controlled whether the distribution of the number of app names generated in each experimental condition in each block was approximately equal. Regarding the instructions, we first described the purpose of the generated app as well as the features the app offers to its users. Figure 3 represents the procedure of the idea evaluation.

Previous empirical studies on creativity adopted different measures for creativity, using various scales or creativity tests (Amabile, 1996). Although different definitions and measures for creativity exist, a content analysis of 90 different articles including from business, education and psychology journals found that the combination of novelty and usefulness were the most prevalent facets (Dean et al., 2006). Hence, similarly to the existing literature (e.g., Amabile, 1996; Baer et al., 2010) we asked the crowd to consider whether each generated name was creative to the extent that it was both novel and useful. However, we did not use a holistic one scale measure since this would have several disadvantages in our context. For example, when asked to rate a generated name for the app, the name could be very novel since the rater has never heard it before but at the same time not useful since the name does not match the features and the superordinate goal of the app. Therefore, following the recommendation by Dean et al. (2006), we used two separate scales to enable the rater to evaluate the name in respect of novelty as well as usefulness. Scales are measured on a seven-point Likert scale.

Results

Idea quantity

We first report the results of the idea generation task in terms of the number of ideas generated in each experimental condition, which involved counting the number of names generated by each participant. For the purposes of correlating effort to higher task meaning, and to avoid distorting the analysis of our key variable for idea quantity, it was important that we counted all the names generated, even if the same name was suggested independently by several participants. Indeed, participants could not observe the names

⁵ These numbers also include redundant names generated by different participants.

generated by other participants, and we had to ensure that this variable accurately reflects the effort invested by each individual.

Table 2 shows the descriptive statistics for both experimental conditions including the variable of interest (Number_Names) as well as the set of control variables. Further, we controlled whether contributors in each condition knew a person who suffers from lactose intolerance – which they were asked before they started on the task. We did not find a significant difference between the conditions in this respect: the percentage was 51.4 of people in the low meaning condition and 52.1 in the high meaning condition. As we can see, the high meaning condition shows a slightly higher average number of generated ideas. All control variables, apart from the perceived meaning, show comparable values for both conditions. In order to assess whether the manipulation was successful, we checked whether subjects in the high meaning condition perceived a higher level of task meaningfulness. In contrast to previous studies (e.g. Ariely et al., 2008), we varied the instructions much more subtly. For example, we omitted a condition (“shredded condition”) in which subjects were told that their work would be completely disregarded. We decided to vary task meaning more subtly since from our point of view it would be unnatural (and non-sensible) to insert a job on a crowdsourcing platform where the crowd is told that their work is worthless anyway. The p-value of a non-parametric Wilcoxon rank-sum test (Wilcoxon, 1945; Mann and Whitney, 1947) corresponds to the allowable upper boundary of the p-value = .10, indicating that the perceived meaning of the task was statistically higher (p-value = .10) in the high meaning condition than in the low meaning condition. Hence, we conclude that the manipulation of our experiment regarding the perceived meaning was successful.

Table 2. Descriptive Statistics						
		N	Mean	Std. Dev.	Min.	Max.
Number_Names	Low_Meaning	71	2.05	2.56	1.00	20.00
	High_Meaning	73	2.10	1.79	1.00	12.00
Time_Working (in Seconds)	Low_Meaning	71	80.98	93.21	6.66	657.78
	High_Meaning	73	78.22	100.17	3.68	566.79
Feeling_Good Solving Task	Low_Meaning	71	5.85	1.33	3.00	7.00
	High_Meaning	73	6.00	1.17	2.00	7.00
Effort	Low_Meaning	71	5.57	1.70	1.00	7.00
	High_Meaning	73	5.49	1.70	1.00	7.00
Perceived_Meaning	Low_Meaning	71	5.07	1.57	1.00	7.00
	High_Meaning	73	5.50	1.35	1.00	7.00
Fun_Solving Task	Low_Meaning	71	5.92	1.15	3.00	7.00
	High_Meaning	73	6.04	1.25	2.00	7.00

Next, we checked whether the number of generated app names in the two conditions were normally distributed. The result of a Kolmogorov-Smirnov test was highly significant (Combined K-S (D) = .3113, $p < .0001$) indicating that the number of generated names significantly deviate from normality. We therefore decided to use a non-parametric test (Wilcoxon, 1945; Mann and Whitney, 1947) to compare the number of app names generated in both conditions. Results (Table 3) indicate that the number of ideas generated in both conditions does not significantly differ.

Table 3. Comparison of Idea Quantity					
	N	Rank sum	Expected	Z	P -Value
Low_Meaning	71	4957	5147.5	0.839	.401
High_Meaning	73	5483	5292.5		
Combined	144	10011	10011		
Note: *p<.10; **p<.05; ***p<.01					

Next, we analyzed the influence of different control variables on the number of names. We conducted two linear regressions, one for the high and one for the low meaning condition, using the number of names generated as the dependent variable and the set of control variables as independent variables. Following previous studies (e.g. Angrist 2001), we created dummy variables for the number of responses generated as either “equal or more than 2”, “equal or more than 4” and “equal or more than 5” (other cut offs led to similar results) and conducted a logit regression. For all independent variables we also checked for multicollinearity, without finding any. Additionally we summarized several items by conducting a principal component analysis (Dunteman, 1989) and repeated our analysis, however, results remained qualitatively unchanged.

Table 4. Analysis of the Influence of Control Variables on the Number of Generated Ideas					
		Number of names generated	Generated ≥2	Generated ≥4	Generated ≥5
High Meaning	Feeling_Good Solving Task	-.379 (.303)	-.122 (.533)	-1.39 (.886)	-1.42 (1.12)
	Effort	.347*** (.114)	.616*** (.223)	1.00** (.521)	1.23* (.715)
	Perceived Meaning	.048 (.148)	.047 (.272)	.046 (.393)	-.497 (.441)
	Fun_Solving Task	.385 (.274)	.662 (.521)	1.26 (.825)	1.32 (1.00)
	Constant	-.117 (556)	-7.15***	-7.75 (3.67)	-7.05 (4.58)
Low Meaning	Feeling_Good Solving Task	.094 (.148)	.433 (.380)	.287 (.735)	-.053 (1.37)
	Effort	.099 (.111)	.216 (.187)	-.054 (.227)	.373 (.814)
	Perceived Meaning	.104 (.121)	.167 (.211)	.343 (.367)	.365 (.697)
	Fun_Solving Task	.141 (.147)	-.094 (.397)	.434 (.778)	.241 (1.51)
	Constant	-.423 (.736)	-4.44 (1.70)	-8.19 (3.75)	-9.04 (6.70)
Controls ¹		✓	✓	✓	✓
Notes: *p<.10; **p <.05; ***p<.01; Robust Std. Err.; ¹ Gender, Age, Knowing a lactose intolerant person					

The results of the robust linear regression indicate that the reported effort has a positive significant effect on the number of ideas generated, but this relationship, however, only applies to the participants in the high meaning condition. In contrast, none of the independent variables in the low meaning condition has a significant positive influence on the number of generated names. Further, the logistic regression for participants who generated more or equal than 2, 4 or 5 names support this result, since the effort significantly affects the probability of generating more than two or even more ideas only for participants in the high meaning condition.

The results can be interpreted as follows: First, the difference between the two experimental conditions provides additional evidence for the successful manipulation in the experiment. Second, although participants in both experimental conditions report a comparable level of effort (Low: 5.57, High: 5.49), this effort only leads to more ideas generated in the high meaning condition. Hence, the reported effort for people in the low meaning condition might be inefficient somehow and mainly driven by the low meaning of the task since participants perceived that their work might be futile anyway. In contrast, the same level of reported effort for the people in the high meaning condition seems to be more efficient, leading to a larger number of ideas. We further conducted a Wilcoxon rank-sum test and compared the number of names generated by participants in both conditions who reported an effort higher or equal to six. Results reveal that the number of names in this comparison was significantly higher for participants in the high meaning condition ($p=.0955$). This result extends previous literature showing a positive association between perceived task meaningfulness and perceived effort (e.g., Moussawi and Koufaris, 2013).

In sum, the results do not support our first hypothesis, thus higher task meaning seems not to increase the output quantity for the given task. One explanation could be that our task strongly differs from tasks previously used in studies which investigated the influence of task meaning in offline or online contexts. In contrast to prior work, our task represents a creative, non-routine task including both a convergent as well as a divergent thinking phase. In contrast to routine tasks, for example identifying correct values within a given amount of data, where people can follow a considerable structure to find a single correct solution, the creative task we used in our experiment is a different type of task as it also includes a divergent thinking phase, in which participants first generate potential solutions and then shortlist these solutions by convergent thinking (Mumford et al., 1991). This process of divergent and convergent thinking involves cyclic processing if the generated ideas do not contain a satisfactory solution. However, individuals in the high meaning condition might go through this process more often, as they are more critical in the convergent phase due to the high perceived meaning of the tasks and therefore look for the best solution, from their point of view. One might argue that the time working on the task should give an indication of whether participants in the high meaning condition go through this cyclic process more frequently. However, although we controlled for the time worked on the task, this metric could not necessarily indicate the number of divergent and convergent thinking phases, since this cyclic process takes place very quickly, almost simultaneously in a person’s brain (Hommel, 2012). Therefore we cannot directly observe whether people in the high meaning condition went through this recurring process more often than people in the low meaning condition. Further, although if we were to assume that people in the high meaning condition would go through the process of divergent and convergent thinking more often than participants in the low meaning condition, this might not necessarily lead to a higher number of submitted ideas since, from the participants’ point of view, they only stop the recurring process when they have generated a potential satisfactory solution and submitted it.

Idea creativity

To investigate whether task meaning affects the creativity of the names generated in each condition, we needed to evaluate the generated names, for which we designed another task on the online crowdsourcing platform. Once all the ideas for app names had been evaluated, we compared the average ratings for app names from each experimental condition. Another approach that could be even more relevant to comparing idea quality is to compare the very best ideas only. Not only would it be important to know which experimental condition comes up with the best ideas, but companies value more a few outstanding ideas identified through idea generation in order to develop them further (Poetz and Schreier, 2012). In line with previous literature (Zhao et al., 2016), we investigate the number of good ideas as defined by a rating equal to or above five on a seven-point scale as a threshold, rather than just the average quality of ideas (Jung et al., 2010; Poetz and Schreier, 2012). In addition, we also identify the overall best idea.

In a first step, we identified redundant names which were generated multiple times by different individuals. Since it would be pointless to ask the crowd to evaluate the same name more than once, we only assigned one version of redundant names to our evaluation list. Finally, we obtained 125 different names for our evaluation task. Since it would be unreasonable to ask each member of the crowd to evaluate all 125 names in sequence, we designed twelve blocks of ten ideas each and one additional block for the remaining five names. Ideas were randomly assigned to one block. Afterwards, we checked whether approximately the same number of names from each experimental condition were assigned to each block and could not find any systematic bias. As described above, each participant was randomly assigned to one block of names. Participants evaluated whether each name was creative in the sense that it is novel and useful, with each criterion measured separately. In sum, 174 individuals evaluated the generated names, while each block of names (following also every single name) was evaluated by at least ten different individuals. To avoid potential learning effects, individuals could only evaluate one block of ideas and could not assign themselves multiple times to the task. Each individual received a fixed fee of 0.20€ for evaluating one block of names.

After all names had been evaluated, for each idea we first calculated their average ratings on the two scales. Next, we calculated the combined average for both scales (Cronbach’s alpha = .90) as a measurement for creativity (Amabile, 1996) to compare all results with respect to the source of the name, i.e. whether the name was generated in the low or high meaning condition. For all three items the descriptive statistics of ratings (Table 5) show slightly higher averages for names generated in the high meaning condition.

Table 5. Descriptive Statistics of Ratings					
	Meaning	Mean	Std. Dev	Min.	Max.
Novelty	Low	4.09	.568	2.87	5.60
	High	4.11	.586	2.35	5.45
Usefulness	Low	4.14	.628	2.54	5.54
	High	4.17	.708	2.21	5.54
Creativity*	Low	4.11	.514	3.03	5.40
	High	4.14	.532	2.32	5.13
Note: *Consisting of the average of both items novelty and usefulness					

We use a parametric⁶ two-sided t-test to compare ratings for names generated in both experimental conditions. Results (Table 6) show no significant differences between any of the items measured. These results remain consistent if we assign the evaluation results of a redundant name multiple times to the experimental condition the name was generated in. For example, if the name “lactofree” was generated five times in the low meaning condition and its creativity was rated 4.83, we would assign this evaluation result five times to the list of evaluation results from the low meaning condition. However, even when applying this procedure, the results (-0.192 ; $p = .419$ for novelty, $.033$; $p = .397$ for usefulness and $.007$; $p = .468$ for creativity) remain qualitatively consistent.

Hence we can conclude that higher task meaning does not, on average, positively affect novelty, usefulness or creativity of ideas generated in the high meaning condition. Following previous literature (Zhao et al., 2016), we additionally investigate which experimental condition came up with the very best ideas as well as the number of good ideas generated in each condition.

⁶ Kolmogorov-Smirnov tests and Levens’s tests indicate that parametric test can be used.

Table 6. Comparison of Idea Ratings					
		Diff.	P-Value	[95% Conf. Inter.]	
Novelty	Low Meaning	-.025	.403	4.00	4.20
	High Meaning				
Usefulness	Low Meaning	-.032	.392	4.04	4.28
	High Meaning				
Creativity	Low Meaning	-.028	.378	4.04	4.22
	High Meaning				
Note: *:p<.10;**p<.05;***p<.01					

Consistent with the literature, we define a rating equal or above five on a seven-point scale as a threshold for good ideas (Jung et al., 2010; Poetz and Schreier, 2012). In addition, we also identify the overall best idea. Apart from the dimension ‘usefulness’, the number of ideas rated equal to or above five (= good ideas) only slightly differ between the experimental conditions for each item measured. Further, the overall best rated ideas as judged against the criteria novelty and creativity, respectively, were generated in the low meaning condition, while the most useful idea was generated in both conditions.

Table 7. Number of Good Ideas and Source of Best Idea			
		Number ⁷ of ideas rated ≥5	Best idea
Novelty	Low Meaning	5	✓
	High Meaning	6	-
Usefulness	Low Meaning	6	✓
	High Meaning	10	✓
Creativity	Low Meaning	2	✓
	High Meaning	5	-

Further, we conducted two separate linear regressions for each experimental condition with the creativity of ideas generated as the dependent and the control variables as independent variables. Results (Table 7) indicate that the control variables had no significant bearing on idea creativity in the low meaning condition. In contrast, “feeling good solving the task” (Column 1) had a significant, positive influence on the creativity of ideas generated in the high meaning condition. We interpret this result in two ways: first, it confirms that the treatment had an effect since this influence only holds for participants in the high meaning condition. Second, this result is in line with previous literature (e.g., Schwarz, 2000) showing that good mood positively influences individuals’ creativity.

⁷ If the same name was generated in both experimental conditions, we counted the name twice, i.e. once for each experimental condition

Table 8. Influence of Control Variables on Creativity of Ideas		
	Creativity of Ideas	
	<i>Low Meaning</i>	<i>High Meaning</i>
Feeling Good Solving Task	.160 (.408)	.491* (.303)
Effort	.369 (408)	.0008 (.724)
Perceived Meaning	-.082 (.151)	-.290* (.167)
Fun Solving Task	-.111 (.357)	-.223 (.207)
Constant	1.97 (1.36)	4.05** (.012)
Controls	✓	✓
Notes: *p<.10; **p <.05; ***p<.01; Clustered Std. Err. by participant; ¹ Dependent variable measured by the average of novelty and usefulness; ² Gender, Age, Knowing a lactose intolerant person		

Surprisingly, the results also indicate a negative effect of perceived meaning on idea creativity for participants in the high meaning group. To further analyze this influence, we graphically investigated the relationship between the perceived meaning and the creativity of ideas generated for participants in this group (Figure 4). Figure 4 indicates that, compared to the low meaning condition, the majority of participants in the high meaning condition indicated values at the upper limit of the scale and only very few reported a task meaning below the value of 5. This significant negative relation seems to be mainly driven by the small number of participants in the high meaning condition who reported a relatively low value for perceived meaning and generated relatively uncreative ideas. To check this influence, we excluded the only participant in the high meaning condition who reported a perceived meaning of 3 and repeated the regression analysis. Results indicate a non-significant relation between perceived meaning and creativity, indicating that this influence was mainly driven by one extreme outlier.

Summarizing, we find no support for the hypothesis that high task meaning positively affects the output creativity of a crowd generating ideas on a crowdsourcing platform, overall. However, this result should be

considered in a differentiated manner. Although high task meaning does not systematically increase idea creativity across all participants, our results indicate a positive influence of a good mood e.g. when participants feel good while solving the creative task. Interestingly, however, this influence only occurs for participants in the high meaning condition. Hence, we extend the existing literature (e.g. Schwarz, 2000) on the positive influence of good mood on creativity, since our results indicate that high meaning might enhance participants’ good/positive mood, in turn resulting in higher idea creativity (e.g. Schwarz, 2000).

Conclusion

While the number of companies using crowdsourcing for idea generation to gain a competitive advantage is growing (Kosonen et al., 2014), little is known about how to positively influence the creativity of workers’ outputs in crowdsourcing markets. Previous literature on influence factors suggests that task meaning might play an important role in influencing creativity. Aiming to shed light on the question of whether output quantity and/or output creativity of workers solving a creative task in crowdsourcing markets could be influenced by task meaning, we conducted an experiment in which crowdworkers were asked to generate names for a specified app, and varied the task meaning for two different experimental conditions.

In this study, we investigate whether higher task meaning in crowdsourcing could positively affect both the quantity of idea generation as well as the creativity of the output. Our results indicate that higher task meaning neither positively affects output quantity of idea generation nor the creativity of generated ideas. However, a more detailed consideration of the perception of the task in each experimental condition reveals that, although participants in both experimental conditions report comparable values for perceived effort, this effort positively influences the number of ideas generated by subjects in the high meaning condition. Further, we found a positive effect of good mood e.g. feeling good when solving the task on idea creativity, but only for the subjects in the high meaning condition. Therefore, we conclude that high meaning might indirectly foster creativity as higher task meaning enhances participants’ good mood, resulting in higher creativity of the generated ideas (e.g., Schwarz, 2000).

Our work contributes to the literature in several ways. First, we contribute to the literature investigating the influence of task meaning on workers’ output in the specific context of crowdsourcing markets. These markets differ from classical labor markets in that crowd workers act as anonymous agents, shifting from one job to another within minutes without fixed ties to companies. Second, to the best of our knowledge, we are the first to investigate the effect of task meaningfulness on performance in the context of crowdsourcing, involving creative tasks. Creative tasks (e.g. idea generation) differ from routine tasks (e.g. image labeling), since they include both a divergent and a convergent thinking phase in contrast to routine tasks which only consist of a convergent phase.

In addition to contributing to the literature, our research also carries practical implications since it offers helpful insights for commissioners of crowdsourcing platforms, and platform designers on how to motivate an anonymous crowd on an online-platform to generate creative ideas. In contrast to previous literature on task meaning in crowdsourcing (e.g., Chandler and Kapelner, 2013) which mainly induced very high task meaning in the context of medicine, where participants could potentially contribute to saving lives by labelling images of tumor cells, we designed a more subtle and less specific cover story since it would be unrealistic to assume that many companies using crowdsourcing for idea generation would be likely to offer tasks as meaningful as labeling tumor cells.

While our study provides important insights relevant for research and practice, we acknowledge certain limitations that ought to be considered. First, our creative task of generating ideas for app names represents a specific version of a creative task, namely a verbally creative task. However, since we conducted our field experiment on a microwork crowdsourcing-platform, this type of task appears natural for this type of online workplace. Nevertheless, we suggest carrying out additional studies with other types of creative tasks to investigate the influence of task meaning in this context. Second, in contrast to other studies, we manipulated task meaning in our experiments in a very subtle way. We decided not to include a “zero-meaning” condition where participants know a priori that their work is entirely worthless because their results will be discarded. On the one hand, by only subtly manipulating task meaning, the treatment effect might be less pronounced compared to a more extreme or pronounced manipulation. However, we did not want to force any artificial treatment effect because, from our point of view, it would appear unnatural on a

crowdworking-platform to design tasks which the crowd knows from the outset to be futile. In addition, one might criticize that we did not design a baseline condition with a neutral task description. However, implementing a baseline condition would not gain relevant additional insights since our results indicate no significant differences regarding our variables of interest for relatively large differences in task meaning. Hence, it would be unreasonable to expect an increasing or decreasing effect of high or low task meaning compared to a baseline condition.

Future work could seek to gain additional insights into other aspects of high task meaning. Since participants in the high meaning condition perceived a significantly higher task meaning, other aspects of this effect, for example higher task involvement or a positive perception of the task designer (=company), could be investigated. Further, results regarding the quantity of output indicate that high task meaning could influence the perceived effort, influencing the number of submitted ideas. Hence, future research could also investigate whether high task meaning could influence participants’ endurance during task processing. Finally, using the “think aloud method” (e.g., Jaspers et al., 2004) by asking subjects in a laboratory setting to talk through their idea generation could help to understand the cognitive processes when solving the given creative task, enhancing the understanding of how participants perceive different task meanings and how task meaning influences the recurring process of divergent and convergent thinking.

References

- Amabile, T. M. 1996. “Creativity in Context: Update to the Social Psychology of Creativity,” Boulder, CO: Westview Press.
- Amabile, T. M. and Hennessey, B. A. 2010. “Creativity,” *Annual Review of Psychology* (61), pp. 569-598.
- Amabile, T. M., Goldfarb, P. and Brackfield, S. C. 1990. “Social Influences on Creativity: Evaluation, Coaction, and Surveillance,” *Creativity Research Journal* (3), pp. 6-21
- Angrist, J. D. 2001. “Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors,” *Journal of Business & Economic Statistics* (19:1), pp. 2-28.
- 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.
- Baer, M., Leenders, R., Oldham, G. and Vadera, A. 2010. “Win or Lose the Battle for Creativity: The Power and Perils of Intergroup Competition,” *Academy of Management Journal* (53:4), pp. 827-845.
- Bayus, B. L. 2013. “Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community,” *Management Science* (59:1), pp. 226-244.
- Chandler, D. and Kapelner, A. 2013. “Breaking Monotony with Meaning: Motivation in Crowdsourcing Markets,” *Journal of Economic Behavior & Organization* (90), pp. 123-133.
- Davis, M. A. 2009. “Understanding the Relationship between Mood and Creativity: A Meta-Analysis,” *Organizational Behavior and Human Decision Processes* (108), pp. 25-38.
- Dahan, E. and Hauser, J. R. 2001. “The Virtual Customer,” *Journal of Product Innovation Management* (19:5), pp. 332-354.
- Dean, D. L., Hender, J. M., Rodgers, T. L. and Santanen, E. L. 2006. “Identifying Quality, Novel and Creative Ideas: Constructs and Scales for Idea Evaluation,” *Journal of the Association for Information Systems* (7:10), pp. 646-699.
- Deci, E. L. and Ryan, R. M. 2000. “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions,” *Contemporary Educational Psychology* (25), pp. 54-67.
- Dewett, T. 2007. “Linking Intrinsic Motivation, Risk Taking, and Employee Creativity in an R&D Environment,” *R&D Management* (37), pp. 197-208
- Dietz, T. 1987. “Theory and Method of Social Impact Assessment,” *Sociological Inquiry* (57:1), pp. 54-69.
- Dunteman, G. H. 1989. “Principal Component Analysis,” Sage Publications.
- Erat, S. and Gneezy, U. 2016. “Incentives for Creativity,” *Experimental Economics* (19), pp. 269-280.
- Füller, J., Hutter, K., Hautz, J. and Matzler, K. 2014. “User Roles and Contributions in Innovation-Contest Communities,” *Journal of Management Information Systems* (31:1), pp. 273-308.
- Frankl, V. E. 1984. “Man’s Search for Meaning” (3rd ed.). New York: Pocket Books.
- 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.

- Gough, H. G. 1979. "A Creative Personality Scale for the Adjective Check List," *Journal of Personality and Social Psychology* (37), pp. 1398-1405.
- Grant, A. M. 2008. "The Significance of Task Significance: Job Performance Effects, Relational Mechanisms, and Boundary Conditions," *Journal of Applied Psychology* (93), pp. 108-124.
- Hackman, J. R., and Oldham, G. R. 1976. "Motivation through the Design of Work: Test of a Theory." *Organizational Behavior and Human Performance* (16), pp. 250-279.
- Hackman, J. R., and Oldham, G. R. 1980. "Work Redesign," Reading, MA: Addison-Wesley.
- Hennessey, B. A. and Amabile, T. M. 2010. "Creativity," *Annual Review of Psychology* (61), pp. 569-598.
- Hommel, B. 2012. "Convergent and Divergent Operations in Cognitive Search," in *Cognitive Search: Evolution, Algorithms, and the Brain*. Strüngmann Forum Reports, Vol. 9, eds P. M. Todd, T. T. Hills, and T. W. Robbins (Cambridge, MA: MIT Press), pp. 215-230.
- Howe, J. 2008. "Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business," Crown Business, New York.
- Huang, Y., Singh, P. V. and Srinivasan, K. 2014. "Crowdsourcing New Product Ideas Under Consumer Learning," *Management Science* (60:9), pp. 2138-2159.
- Ipeirotis, P. 2010. "Demographics of Mechanical Turk." New York University. Working Paper
- Jaspers, M. W. M., Steen, T., van der Bos, C. and Geenen, M. 2004. "The Think Aloud Method: a Guide to User Interface Design," *International Journal of Medical Informatics* (73), 781-795.
- Jung, J. H, Schneider, C. and Valacich, J. 2010. "Enhancing the Motivational Affordance of Information Systems: The Effects of Real-Time Performance Feedback and Goal Setting in Group Collaboration Environments," *Management Science* (56:4), pp. 724-742.
- John, T. and Kundisch, D. 2015. "Why Fit Leads to Surprise: An Extension of Cognitive Fit Theory to Creative Problems," *Proceedings of the 36th International Conference on Information Systems*, Fort Worth, USA.
- Kaufman, J. C. 2002. "Creativity and Confidence: Price of Achievement?," *American Psychologist* (57), pp. 375-376.
- Kaufman, N., Schulze, T. and Veit, D. 2011. „More than Fun and Money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk," *Proceedings of the Seventeenth Americas Conference on Information Systems*, Detroit, USA.
- Kittur, A. 2010. "Crowdsourcing, Collaboration and Creativity," *ACM Crossroads* (17:2), pp. 22-26.
- Kittur, A., Nickerson, J., Bernstein, M., Gerber, E., Shaw, A., Zimmermann, J., Lease, M. and Horton, J. 2013. "The Future of Crowd Work," *Proceedings of the 2013 Conference on Computer Supported Cooperative Work and Social Computing*, pp. 1301-1318.
- Kosfeld, M., Neckermann, S., and Yang, X. 2014. "Knowing that You Matter, Matters! The Interplay of Meaning, Monetary Incentives, and Worker Recognition," *IZA Discussion Paper*, No. 8055.
- Kosonen, M., Gan, C., Vanhala, M. and Blonqvist, K. 2014. "User Motivation and Knowledge Sharing in Idea Crowdsourcing," *International Journal of Innovation Management* (18:5), pp. 1-21.
- 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.
- Mann, H. B., and Whitney, D. R. 1947. "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *Annals of Mathematical Statistics* (18), pp. 50-60
- Mason, W. A. and Watts, D. J. 2009. "Financial Incentives and the Performance of Crowds," *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pp. 77-85.
- Martindale, C. 1989. "Personality, Situation, and Creativity," In: J. A. Glover, and Reynolds R. (Eds.), *Handbook of Creativity*, pp. 211-232.
- Meinel, M. and Voigt, K. I. 2017. "What Do We Really Know about Creativity Techniques? A Review of the Empirical Literature," In: Brem, A. Puente-Diaz, R. and Agogu, M., "The Role of Creativity in the Management of Innovation," World Scientific.
- Moussawi, S. and Koufaris, M. 2013. "The Crowd on the Assembly Line: Designing Tasks for a Better Crowdsourcing Experience." *Proceedings of the 34th International Conference on Information Systems*, Milan, Italy.
- Mumford, M., Mobley, M., Reiter - Palmon, R., Uhlman, C. and Doares, L. 1991. "Process Analytic Models of Creative Capacities," *Creativity Research Journal* (4:2), pp. 91-122.
- Mumford, M. D. and Gustafson, S. B. 2007. "Creative Thought: Cognition and Problem Solving in a Dynamic System," In: Runco, M. (Ed.), "Creativity Research Handbook, Vol. 2, Hampton, Cresskill, NJ.

- National Center for Biotechnology Information:
<https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0072452/>.
- Oldham, G. R. and Cummings, A. 1996. "Employee Creativity: Personal and Contextual Factors at Work," *Academy of Management Journal* (39:3), pp. 607-634.
- 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.
- Pratt, M. G., and Ashforth, B. E. (2003). "Fostering Meaningfulness in Working and at Work." In: K. S. Cameron, J. E. Dutton, and Quinn, R. E. (Eds.), *Positive Organizational Scholarship*, San Francisco: Berrett-Koehler Publishers, Inc.
- Prendergast, C. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* (37:1), pp. 7-63.
- Schwarz, N. 2000. "Emotion, Cognition and Decision Making." *Cognition and Emotion* 14, pp. 433-440.
- Shin, S. J. and Zhou, J. 2003. "Transformational Leadership, Conversation, and Creativity: Evidence from Korea," *Academy of Management Journal* (46:6), pp. 703-714.
- Silva, P. J. 2008. "Creativity and Intelligence Revisited: A Latent Variable Analysis of Wallach and Kogan," *Creativity Research Journal* (20:1), pp. 34-39.
- Thaler, S., Simperl, E. and Wölger, S. 2012. "An Experiment in Comparing Human-Computation Techniques," *IEEE Internet Computing* (16:5), pp. 52-58.
- Torrance, E. P. 1974. "Torrance Tests of Creative Thinking: Norms - Technical Manual," Lexington, MA: Ginn.
- Van der Lugt, R. 2002. "Brainsketching and how it differs from Brainstorming," *Creativity and Innovation Management* (11:1), pp. 43-54.
- Wilcoxon, F. 1945. "Individual Comparisons by Ranking Methods." *Biometrics* 1, pp. 80-83.
- Wolf, L., Knuth, M., Osterhoff, J. and Sack, H. 2011. "RISQ! Renowned Individuals Semantic Quiz: A Jeopardy Like Quiz Game for Ranking Facts," *Proceedings of the 7th International Conference on Semantic Systems*, pp. 71-78.
- Zhao, Z., Renard, D., Elmoukhli, M., and Balague, C. 2016. "What Affects Creative Performance in Idea Co-Generation: Competitive, Cooperative or Cooperative Climate?," *International Journal of Innovation Management* (20:4), 1640002.
- Zheng, H., Li, D. and Hou, W. 2011. "Task Design, Motivation, and Participation in Crowdsourcing Contests," *International Journal of Electronic Commerce* (15:4), pp. 57-88.
- Zizzo, D. 2010. "Experimenter Demand Effects in Economic Experiments", *Experimental Economics* 13, pp. 75-98.