Social Gamification in Enterprise Crowdsourcing

Gregory Afentoulidis
Delft University of Technology
Delft, The Netherlands
gregafent@gmail.com

Zoltán Szlávik
IBM Netherlands
Amsterdam, The Netherlands
zoltan.szlavik@nl.ibm.com

Jie Yang
University of Fribourg
Fribourg, Switzerland
jie@exascale.info

Alessandro Bozzon
Delft University of Technology
Delft, The Netherlands
a.bozzon@tudelft.nl

ABSTRACT

Enterprise crowdsourcing capitalises on the availability of employees for in-house data processing. Gamification techniques can help aligning employees' motivation to the crowdsourcing endeavour. Although hitherto, research efforts were able to unravel the wide arsenal of gamification techniques to construct engagement loops, little research has shed light into the social game dynamics that those foster and how those impact crowdsourcing activities. This work reports on a study that involved 101 employees from two multinational enterprises. We adopt a user-centric approach to apply and experiment with gamification for enterprise crowdsourcing purposes. Through a qualitative study, we highlight the importance of the competitive and collaborative social dynamics within the enterprise. By engaging the employees with a mobile crowdsourcing application, we showcase the effectiveness of competitiveness towards higher levels of engagement and quality of contributions. Moreover, we underline the contradictory nature of those dynamics, which combined might lead to detrimental effects towards the engagement to crowdsourcing activities.

KEYWORDS

Enterprise Crowdsourcing; Gamification; Social Incentives

ACM Reference Format:
Gregory Afentoulidis, Zoltán Szlávik, Jie Yang, and Alessandro Bozzon. 2018. Social Gamification in Enterprise Crowdsourcing. In *WebSci '18: 10th ACM Conference on Web Science, May 27–30, 2018, Amsterdam, Netherlands.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3201064.3201094

1 INTRODUCTION

Crowdsourcing is a computational paradigm that builds upon the idea of harnessing the collective intelligence of the crowd to overcome limitations of current technologies, which unavoidably require human intervention and intellect. Enterprises have been adopting the paradigm to bolster their business needs and processes, transferring the practices of crowdsourcing from the online environment to the internal crowd of the enterprise: the *employees*.

Enterprise crowdsourcing allows the deployment of tasks of confidential nature, and it benefits from the utilisation of employees' working capacity and knowledge for quality task contributions [19]. However, it also suffers from traditional challenges of participation, retention to the crowdsourcing endeavour, and quality of the produced work [35]. Gamification[1] is often seen as a suitable tool for engagement and retention purposes. However, it is widely accepted that introduction of gamification involves several non-trivial steps that require strong consideration and scrutiny in order to achieve its goals [18, 24, 25]. In this respect, crowdsourcing is no exception.

Problem Statement. Previous work shows that gamification can incentivise the crowd and drive its behavioural outcome towards augmented and prolonging participation and task contribution, as well as quality output [9, 33, 34]. However, it is still unclear which game mechanics are more suitable for enabling crowdsourcing within an enterprise. This is mainly because gamification techniques are not always necessarily tied to the motivations of the employees. More importantly, limited research has been focusing on evaluating the interplay between *game elements* and *social incentives*, especially in an enterprise context in which synergy and competition are concepts that play an important role.

In this work, we aim at achieving a better understanding of the motives of employees behind participation in a gamified enterprise crowdsourcing application, so as to clarify the main requirements for selection of suitable tasks within an enterprise, and to adequately inform gamification design. We seek to answer the following research question:

RQ: How can gamification techniques enhance reliability and foster engagement in enterprise crowdsourcing?

Original Contribution. We instrumented a study that involved 101 employees from two large multinational enterprises, for an observational interval that lasted two months. First, we performed a qualitative exploratory analysis of the dominant player types existent within the targeted enterprises, and highlight the importance of the social characteristics of the workforce that inform the design of enterprise gamification. We then set-up a quantitative study of gamified enterprise crowdsourcing by extending a mobile enterprise crowdsourcing application (`ECrowd` [30]) with pluggable

[1] The "process of enhancing a service with affordances" (e.g. game mechanics) "for gameful experiences to support user' value creation" (e.g. engagement) [24].

WebSci '18, May 27–30, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5563-6/18/05…$15.00
https://doi.org/10.1145/3201064.3201094

gamification elements. We implement *competitive* and *collaborative* game mechanics by designing a scoring function based on the number and quality of contributions, and task sharing capabilities within the enterprise to foster community collaboration.

We apply those two aspects of social gamification on top of traditionally employed game mechanics, to study the effects of synergistic and competitive dynamics in engagement and data quality in enterprise crowdsourcing. Finally, we compare the results obtained in the two enterprises, to gain a better understanding of the contextual effects that might exist between the relationship of gamification and crowdsourcing and how they mediate it.

Results suggest a preference of competitive game mechanics over the collaborative ones, and show the detrimental effects that their combination might yield to users engagement. As far as quality is concerned, the experiments showed that depending on the task type, we can expect higher quality contributions when competitive and collaborative game mechanics are used. Despite variations in the perception of gamification were noticed between employees of the two companies, we were not able to report significant differences.

Paper Organisation. The remainder of the paper is organised as follows. Section 2 presents related work. Section 3 describes the applied research methodology. Section 4 introduces the `ECrowd` enterprise crowdsourcing platform, and its extensions. Section 5 presents and discusses the experimental results. Section 6 concludes.

2 BACKGROUND AND RELATED WORK

Incentives in Enterprise Crowdsourcing. Enterprise crowdsourcing differs from traditional online crowdsourcing in terms of both the crowd it involves (i.e., employees) and the problem it targets (i.e., business problems) [35]. These two differentiating characteristics bring potential benefits for enterprises along with big challenges: it provides an effective way to exploit the internal knowledge profiles of employees and to leverage on their non-utilised working capacity to solve business critical and confidential tasks [15, 19], but it also faces challenges such as adherence to intellectual property legislation for the re-purposed work of employees, and minimising the risk of information leakage related to the business problems under consideration [35]. More importantly, as the main focus of the employees in a corporate environment is on accomplishing their daily duties and tasks, a strong consideration of the motivation of the crowd and fine engineered incentive mechanisms are required.

As opposed to money-based rewards in public crowdsourcing, in enterprise crowdsourcing, intrinsic motivations are mainly exploited. This is due to the conflict between money-based rewards and the already established compensation arrangements with the employees [35]. It is suggested that identifying what is the main interest of the employees in terms of personal values, causes and actions is pivotal for successful enterprise crowdsourcing [31]. Employees tend to be motivated by learning something new, improving the output of the company, contributing to their work community, improving appraisal for their work, but also by having fun [3]. Incentive mechanisms that account for these motivations are critical to worker engagement [21], which significantly affects the quantity and the quality of the performed tasks [16, 22].

Gamification in Crowdsourcing. Gamification has been widely recognised as an effective way to increase motivation towards better user engagement and participation [7, 24, 27], also in crowdsourcing. Early works mainly study Games With A Purpose (GWAP) which denote the notion of disseminating tasks in a game that incites enjoyment. Notable examples include the ESP game [33], and Peekaboom [34]. While GWAP s start by defining games and afterward introducing crowdsourcing tasks, gamified crowdsourcing processes deal with existing tasks with gamification working as an added engagement layer. Tasks falling into this type include data collection [5, 23], entity and relation extraction [8], and relevance assessment [9]. The potential of gamification in enterprises has been increasingly noticed [18, 25], as it is flexible in addressing a variety of business processes and needs in an efficient manner [25]. To the best of our knowledge, little work has studied gamification in the context of enterprise crowdsourcing, which comes in contrast to the significant need for careful treatment of worker incentives required by the distinct characteristics of enterprise crowdsourcing.

Social & Contextual Factors in Gamification. Next to game mechanics, the success of gamification also depends on a set of contextual factors such as application type, task type, and user type [24]. At the application-level, Hamari [13] noted a discrepancy in the behavioural outcomes in traditional games and in a utilitarian service when badges are applied as game mechanics. Similar results were found in a gamified citizen science application [5]. At the task-level, Geiger and Schader [11] stress the need for different game mechanics used for tasks of different types. In processing and rating applications simpler game mechanics are preferred such as points, badges, and leaderboards. In tasks that require more creativity (e.g., content creation), more involved game mechanics are recommended to promote collaborative game dynamics and social influence, such as rewards, progress, social status, curiosity, and altruism. The user type relates to how the end users perceive gamification and how they react in its presence.

Bartle [2] identifies four main reasons why players typically enjoy a game, namely achievements within the game context, exploration of the game, socialising with others and imposition upon others. The categorization leads to the classical four player types: achievers, explorers, socialisers, and killers. Players might incorporate characteristics by all four types depending on the current state, but Bartle also suggests that a predominant preference to one of those four is existent in every player. The distinct incentives for different player types call for a better understanding of the effects of different game mechanics in gamified enterprise crowdsourcing. Using Bartle's taxonomy, game mechanics and dynamics for each player types have been suggested [26]. Such a necessity is confirmed in the context of online education [17] and task execution [14].

These works further suggest the importance of the social influence [14] in gamification, which motivates individuals to act in accordance with the social norms of the group. Similarly, Shi et al. [28] mention that social relatedness in a gamified context can be achieved, among others, through tagging, rating and commenting, which can be understood as social feedback. These works link to Thiebes [32], where a separate cluster called social influences was created in the taxonomy of game mechanics and dynamics, and two manifestations of social gamification were further identified, namely competition and collaboration.

Despite the literature, it remains an open question how social gamification implemented through different game mechanics, together with different contextual factors, i.e., user, task, and application types, affect worker engagement in enterprise crowdsourcing.

3 RESEARCH METHODOLOGY

Our research methodology starts with interviews with employees and experts to respectively understand employee player types and relevant tasks for enterprise crowdsourcing. Informed by their results, we design experimental conditions where specific gamification mechanics are applied, so as to understand the effects of different social gamification elements in enterprise crowdsourcing. To quantitatively analyze effects, we introduce the metrics for employee engagement and the reliability of their contribution.

To account for potential effects of enterprise environments, the research has been conducted within the Dutch headquarters of two multinational companies, which are referred to as **ET1** (a major Dutch bank) and **ET2** (one of the largest technology companies in the world). Their names are omitted due to legal requirements.

3.1 Interviews with Employees and Experts

Employee Player Types. To understand the player types of employees, we conducted 7 semi-structured interviews [10], a commonly used class of interviews to collect subjective opinions of people about their personal characteristics and also those of their peers. Interviews were structured into 5 parts: the first 4 parts contain questions related to discovering characteristics of the employees that pertain to one of the four player types suggested by Bartle's theory; the final part is designed to discover the predominant player type of the employee. Employees were also asked to provide their opinions on the player type of the general employee population that best suits the company. The employees were selected with a prior determination of the sample structure [10] based on gender, department, role and field of expertise, and considering availability.

In the following, we first describe qualitative findings obtained from the interviews, then we present the distribution of employee player types within the companies under consideration.

Socialiser Type. Respondents unanimously expressed that the ability to draw inspiration from coworkers and to develop a network within the company is of paramount importance for their ability to perform their work duties. Almost all employees expressed their personal desire to work in an environment where their feeling of relatedness is satisfied and opportunities are provided to create social connections. We were also interested to find out under which prism is social interaction expressed and preferred. All answers focused on collaborative characteristics rather than competitive, signifying that the former is much more valued in a working environment.

Achiever Type. In order to check how many of the characteristics of the achiever player type are incorporated to the employees, we set to find out how much reward oriented they are in their work. We were also interested to find out the types of rewards that are usually expected and how those are tied to their intrinsic or extrinsic motivation for doing their work. The employees' main preference resides on rewards that adhere to their intrinsic motivations. In addition, the majority of the employees recognised rewards as a main motivator for their work and also something that should be tied to their performance.

	Socialiser	***Achiever***	***Killer***	***Explorer***	***Employee***
S1	7-8	6-7	6	8	Socialiser
S2	-	-	-	-	-
S3	8	6	7-8	6	Killer
S4	6	8	9	7-8	Achiever
S5	8	7	7-8	6	Socialiser
S6	6	9	6	8	-
S7	8	7	9	8	Socialiser

Table 1: Employees' ratings on the level to which they match their personal characteristics to the 4 player types (from 1 to 10); and selection for the general population of the company.

Killer Type. To unravel characteristics of the employees that might be related to the killer player category, we asked them to comment to what extent they are finding themselves challenging their standard way of working. Most employees suggested that following a standardised way of working is in general preferred. However, they also suggested that there is a compromise between blindly accepting a specific way of operating and also being aware of opportunities where they can intervene and break the conventional order. It is also interesting to note that most employees who expressed willingness to deviate from a standard way of working would only opt for this solution when they can critically assess that this is for the benefit of their work's end result, rather than an innate personal characteristic that incentivises them to act in this specific way.

Explorer Type. To determine how much of an explorer player type are the employees, we focused on gaining an understanding on whether they like to work independently and have their own path within their working environment. The responses were balanced between employees who prefer to work in an isolated fashion and are often given the opportunity to work on new things not closely related to their main work, and those who are more focused on it.

Most of the employees in our analysis responded positively in questions that were probing whether characteristics from the 4 types can be found in them. Such a result matches the hypothesis of Bartle's theory, that the player categories are not mutually exclusive and that multiple characteristics of them can be found in a person. In the concluding part of our discussions with the employees, we asked participants to select which of player types best suits their personality. We allowed them to rate in a scale from 1 to 10 how congruent they find themselves with this type. The responses are listed in Table 1 (subject 2 did not provide rates).

We can observe that there is a preference in the *Socialiser Killer* player type. Interestingly, when the employees were asked to provide their opinion on which of those categories best suits the general company population, most of the answers indicated the *Socialiser* type. It is also interesting to denote that the responses for the categories of *Socialiser* and *Achiever* had greater consensus compared to that of the *Killer* and the *Explorer.* Almost all respondents recognised qualities found in the first two categories in them while there was some dispersion in the answers we collected for the latter two. Those observations are in accordance with previous studies regarding gamification conducted at another Dutch company [30].

Gamification Elements	Contr.	Comp.	Collab.	Mixed
Score + Progress Bar	✔	✔	✔	✔
Leaderboard	–	✔	–	✔
Task Sharing	–	–	✔	✔

Table 2: The four experimental groups.

Enterprise Crowdsourcing Tasks. We conducted expert interviews to identify relevant tasks for enterprise crowdsourcing. For each company, we identified a use case in the domain of news analysis and summarisation.

Enterprise crowdsourcing would provide human-generated data used to train a machine learning model. Following the interview guide for expert interviewing [10], our interview addresses the following questions: 1) How do the experts conduct their research and produce their reports for which they need AI support? 2) Which are the data sources they use in their work? 3) What are some possible aspects of their work which could be automated by a machine learning model?

As a result of the interviews, several possible tasks were identified, out of which one was chosen as a focal point for experimentation: extraction of market information, i.e. the identification of key companies and corresponding relevant information in a specific domain. The objective of the machine learning algorithm utilising the crowdsourced data is to extract possible relations found between entities in unstructured online text data. The selection was promoted as a result of the easily accessible data sources which are mainly online articles and news as compared to the other options who involved proprietary data sources. We focused on 2 relations: identifying a CEO of a company, and extracting affiliation relation between companies (e.g. subsidiary company or acquisitions).

3.2 Experimental Conditions

Based on our interviews, we selected gamification elements which mostly adhere to the Achievers and Socialisers player types, namely: points, progression, leaderboard and community collaboration. Four experimental groups are created (Table 2):

Control, offering essential feedback gamification mechanics such as score and a progress bar. The scoring mechanism (described in Section 4) is based on the contributions and the quality of the work while the progress bar provides a visual representation of the amount of tasks completed from the total available.

Competitive, which promotes competitive dynamics by offering a leaderboard (based on the scoring mechanism), in addition to the progress bar.

Collaborative, which provides collaborative social gamification by means of two options for task submissions: 1) individually submit the task for one's own benefit (i.e. increase score and progress) or collaboratively (i.e. solve the task with a peer of choice). For the latter, the employee is able to submit a task and also assign it to another participant to annotate it (this takes place asynchronously).

Mixed, which provides all the previously mentioned game mechanics. This allows us to study the interaction effect of competitive and collaborative social gamification.

Legal and Privacy Aspects. The presence or real personal information about the employees participating in it is of paramount importance: the ability of an employee to relate an account to one of his peers strengthens feelings of relatedness, community acknowledgment, synergy and competitiveness.

However, the enterprise environment might pose stricter requirements in terms of privacy: employees' personal information is sensitive and confidential; also, the logging functionalities that are necessary to obtain usage metrics, and the storage and use of personal and application usage information might not in accordance with enterprise privacy policies. This condition occurred with our experiments, as we were limited to the usage of anonymous users participating in each experimental group. This constraint unavoidably introduces limitations to our experiments, which is evident in ***Competitive*** and ***Mixed*** groups – where we are restricted to use leaderboards with not realistic user names – and in ***Collaborative*** and ***Mixed*** groups, where the task sharing functionality has to be based, again, on the same user names.

3.3 Measuring Engagement and Reliability

We operationalise employee engagement and reliability of the outcomes in enterprise crowdsourcing to quantitative metrics collected and logged through the interaction of the employees with the mobile crowdsourcing application.

Employee Engagement Metrics. Engagement metrics are used to evaluate the level of interaction of the employees with the application. We measure: 1) *Number of task executions*, i.e. the average number of tasks contributed by an employee, normalised by the observation interval during the experiment duration. 2) *Number of sessions*, i.e. the amount of times an employee opened and interacted with the app during the observation period. A session start is determined when the application starts or resumes. 3) *Session time*, i.e. the time each employee spends in interacting with the application. This is signified by an application start or resume event in their mobile device until an application paused or closed event. We average the total session time by the number of sessions an employee has had within a normalised time span. 4) *Task dwell time*, i.e. the amount of time elapsed since an employee selects a task until she submits or selects to collaborate a task. We average the dwell time across all task executions contributed by the employee. Since collaboration is implemented as a task sharing function, which might affect metric across different experimental conditions, we denote the end of task execution at the time an employee presses the submit or collaboration button.

Work Reliability Metrics. Due to the absence of golden standard labels for the tasks used in our experiments, we rely on agreement metrics. Depending on the input requested per task category we use different quality metrics elaborated as follows. 1) *Plurality answer agreement*: for tasks with numerical input the following formula is used to calculate plurality answer agreement:

$$S_p(e) = \frac{f}{F} \tag{1}$$

where e represents an employee, F is the total amount of tasks the employee has provided annotations and f is the number of the tasks for which the employee's annotations are in accordance with those produced from the majority vote. This metric assumes majority vote as the golden standard on which the employee's annotations are directly assessed. 2) *Average worker-worker agreement*: for tasks where the employee is requested to annotate relations found in text, we use the average worker-worker agreement [1]:

$$avg\_wwa(e_i) = \frac{\sum_{i \neq j} |S_{i,j}| * wwa(e_i, e_j)}{\sum_{i \neq j} |S_{i,j}|} \quad (2)$$

where e_k denotes the employee k and $S_{i,j}$ is the set of common task annotated by both employees. $wwa(e_i, e_j)$ is the pairwise worker-worker agreement for all the tasks s annotated in common:

$$wwa(e_i, e_j) = \frac{\sum_{s \in S_{i,j}} RelationsInCommon(e_i, e_j, s)}{\sum_{s \in S_{i,j}} NumAnnotations(e_i, s)} \quad (3)$$

in which $RelationsInCommon(e_i, e_j, s)$ is the number of annotated relations that are in common between two employees in a specific task s and $NumAnnotations(e_i, s)$ is the total annotations produced by employee e_i for the same task.

4 THE ECROWD PLATFORM

The experiments have been enabled by ECrowd, an enterprise crowdsourcing platform [3]. This section discusses the design choices related to its extension with the functionalities required by the experimental setting. Figure 1 shows screenshots from the deployed application. Users can navigate the functionalities of the application via the main menu list (Figure 1a), accessible after authentication.

Task Types. The study includes three tasks, selected according to 1) their relation to the domain of enterprise crowdsourcing, 2) the incentives of the crowd, and 3) our research requirements. We introduced variability across those dimensions to isolate as much as possible the effect of gamification in worker engagement.

The *Information Extraction* task (Figure 2a) have been selected and designed according to the outcome of the qualitative research study (Section 3.1). Users are required to annotate the relations (i.e. being the CEO of a company, being a subsidiary of another company) and the participating entities. The task addresses incentives of the crowd related to their participation in innovative projects that also the improvement of the output of the company.

The *Moral Machine* task (Figure 2b) is a survey tasks, based on a research on the morality of future Artificial Intelligence [4].[2] The task mainly addresses learning incentives – as it helps to raise awareness about the importance of programming moral decisions in AI – and fun. It has low complexity, as it involves only the selection of one of the scenarios depicted in an image.

Finally, the *Cell Count* task (Figure 2c) involves the annotation of the number of human cells that are visible in a medical image. The input of the employees is used for the development of machine learning application in the medical domain, thus addressing incentives regarding participation in interesting and useful projects. They are also tasks of intermediate complexity, as it requires some basic knowledge in identifying cells in images.

Task Sharing. The task sharing capabilities underpin the requirements for the collaborative social gamification experimental condition.Users can choose between submitting the task individually and claiming their score (described below) when submitting it and *also* choosing a colleague in their group to share the task. If a user selects to collaborate then an action sheet slides up (Figure 1e) with all the available names of the colleagues in his group.

Upon selection of a peer from the list, the task execution is concluded as normal. After a task has been shared it is stored and forwarded to the receiver which can then choose to complete it asynchronously. To complement the social incentives of collaboration we also used a feedback mechanism, to allow the user who completes a shared task to get a brief notification of the answer of the sender, to check whether his/her annotation is the same or not with that of the sender (Figure 1f).

Scoring. Scoring is used as a feedback mechanism to inform participants about their progress while contributing tasks, and to rank them in leaderboards. The score should reflect the quality of the work; and should be fair, so to foster interest to the user. In the context of crowdsourcing, it might not be possible to reward users for their answers on the basis of a ground truth or a gold standard. We address this by means of a scoring function that rewards both the amount of contribution and the quality of their answers [24]. The quality of the contributions is measured according to the level of agreement to the annotations of other users [8, 12]: the more annotations from previous contributors are in accord with that of the user, the greater the rewarded score. The scoring function is defined as in Equation 4:

$$f(x, C) = \begin{cases} [log(x+3) * g(C)] - 50, & \text{if } C \neq 0 \text{ and } x > 0 \\ [log(x+3) * g(6)] - 50, & \text{if } C \neq 0 \text{ and } x = 0 \\ 50, & \text{if } C = 0 \end{cases} \quad (4)$$

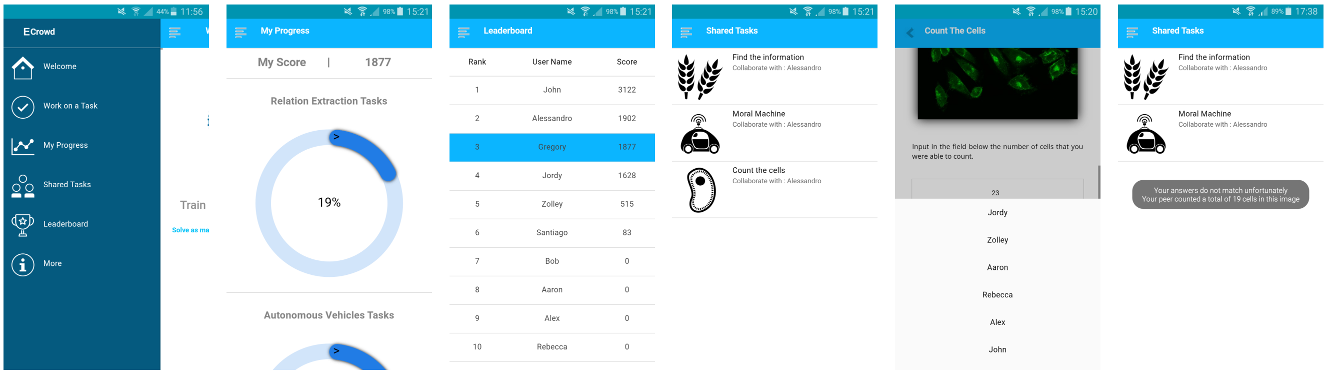
where x is the number of answers equal to the user's, C is variable that represents the level of majority (i.e. majority answers get $C = 1$, the next group $C = 2$, etc.) and the function $g(C)$ is a selection of constants, that parameterise the scoring mechanism on C with different scoring functions (i.e. $g(1) = 65$, $g(2) = 60$, $g(3) = 55$, etc.). Intuitively, each user's annotation is rewarded a higher score depending on whether it belongs to higher levels of majority and also dependent on the number of annotations which form this specific majority. A score is rewarded for each annotation regardless of whether it is in agreement with previous ones, so that we can reward continuous contributions irrespective of their quality. 50 points are rewarded when there are no previous annotations.

A bonus of 30% of what would be normally awarded is given when two users are in agreement for a specific shared task. We penalise disagreement by awarding them 0 points for a collaboration that ended up with disagreement. In this way sharing a task with a peer introduces, from a scoring point of view, a risk of either being awarded bonus points or not being rewarded any points at all. This collaborative scoring strategy is also in accordance with popular gamified crowdsourcing application used in previous studies [20].

5 EXPERIMENTAL RESULTS

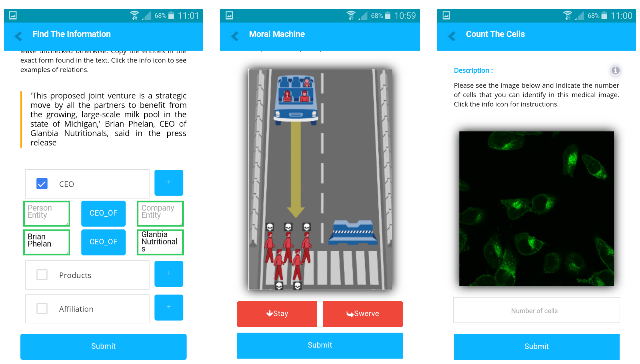
To answer the main research question, we structured the analysis into three research sub-questions: **RQ1)** What is the effect of competitive and collaborative game mechanics to employee engagement in enterprise crowdsourcing? **RQ2)** What is the effect of competitive and collaborative game mechanics to the quality of employee contributions in enterprise crowdsourcing? And, **RQ3)** What is the effect of competitive and collaborative game mechanics across different enterprise environments?

[2] Authors deploy crowdsourcing to collect opinions regarding moral decisions for autonomous vehicles.



(a) Main menu **(b) My Progress** **(c) Leaderboard** **(d) Shared tasks** **(e) Task Sharing** **(f) Task feedback**

Figure 1: Screenshots from the ECrowd application deployed in both companies.



(a) Info. extraction **(b) Moral Machine** **(c) Cell Count**

Figure 2: Implemented Crowdsourcing Tasks.

5.1 Recruitment and Participation

The experiment consisted of 2 phases, and was performed in parallel in Company 1 (**ET1**) and Company 2 (**ET2**). The first phase, lasting 10 days, involved a dozen of selected employees, and helped bootstrap gamification elements, so as to prevent later participants being demotivated by the lack of previous activities. Recruitment for the second phase has been performed on a voluntary basis, through advertisement (flyers, posters in key locations, corporate mailing lists, corporate blogs). The second phase lasted from mid-May until mid-July, 2017. Participants could join the experiment at any time. To account for the variations of participation duration across different employees, we normalise our observation interval to a maximum of 1 month.

Table 3 summarised the tasks executed and employees demographics in the two companies. Despite the adoption of similar advertisement procedures, participation and attrition levels are notably different. 84 employees from **ET1** volunteered, and 75 logged in to the application.in **ET2** had an higher attrition rate in which 26 employees were active from 34 in total. We account those differences to two factors: 1) the number of employees (**ET1**) has more); and 2) the popularity of the companies' app stores, which in **ET2** was lower. The distribution of employees across different experimental conditions is acceptable.

	Task Type				Employees Demographic				
	Info.	Moral	Cell	Tot.	F	M	Man.	N/Man.	Tot.
ET1	343	601	329	1237	27%	73%	28%	72%	75
ET2	88	313	101	502	19%	81%	4%	96%	26

Table 3: Executed tasks and worker demographics.

5.2 RQ1: Impacts of Social Gamification on Employee Engagement

The analysis of employee engagement includes number of task executions, session time, number of sessions, and task dwell time.

Number of task executions. Table 4 summarises the descriptive statistics for the number of tasks executions. We omit the contributions of one employee in the control group, who contributed 68.5% of the total task, and therefore regarded as an outlier.

Previous research [29] states that both competitive and collaborative game mechanics in isolation can have positive effects on engagement, while their combination can have detrimental effects.

		ET1			ET2		
	Type	μ	σ	**Tot.**	μ	σ	**Tot.**
Contr.	*Info.*	3.2	2	51 (32%)	2	2.2	14 (15%)
	Moral	3.7	3.4	70 (43%)	10.28	12.5	72 (75%)
	Cell	2.4	2.2	41 (25%)	1.42	1.3	10 (10%)
	All	**3.1**	**3.6**	**162**	**4.8**	**8.1**	**96**
Comp.	*Info.*	5.6	7.4	90 (21%)	4.3	4.6	26 (17%)
	Moral	13.2	12.2	224 (52%)	11.8	11.8	83 (55%)
	Cell	7.1	8.5	120 (27%)	6.1	4.7	43 (28%)
	All	**8.7**	**10**	**434**	**7.6**	**8.22**	**152**
Collab.	*Info.*	2.1	1.9	25 (14%)	6	9.6	30 (30%)
	Moral	8.1	10.5	121 (69%)	7.6	12.1	38 (39%)
	Cell	2	2.3	30 (17%)	6.2	9.0	31 (31%)
	All	**4.2**	**7**	**176**	**6.6**	**9.5**	**99**
Mixed	*Info.*	1.7	1.6	29 (19%)	2.8	1.9	14 (18%)
	Moral	4.9	8.4	74 (50%)	9.5	17.6	57 (75%)
	Cell	2.9	5.1	46 (31%)	1.7	1.5	5 (7%)
	All	**3.1**	**5.7**	**149**	**5.42**	**11.56**	**76**

Table 4: Descriptive statistics of number of task executions.

We therefore test the alternative hypothesis that, compared to the control group, *Competitive* and *Collaborative* mechanics increase the amount of tasks contributed by the users, while their combination will lessen the effect. We fit a Negative Binomial regression model.[3] The coefficients of this model and their significance are summarised in Table 6, omitting the intercept. There is a significant increase in task executions when competitive gamification elements were used in isolation. In **ET2**, leaderboards have a significant positive effect ($p < 0.05$), resulting in an increase of 282% ($e^{1.03}$) compared to the control group. The *Collaborative* mechanic also proved beneficial, with a slight positive effect on tasks sharing. There is also an indication that their combination (*Mixed* might be detrimental for crowdsourcing activities. We cannot however confirm the alternative hypothesis for these gamification elements since their effects were not found to be significant. Those results were consistent for both experiments.

Table 5 shows the number of tasks shared in the *Collaborative* and *Mixed* groups. 5.7% of the total tasks contributed by the *Collaborative* group in **ET1** were due to task sharing, while for *Mixed* group the percentage climbs to 35.6%. For **ET1** those percentages are respectively 13.13% and 13.15%.

	RT1		**RT2**	
	Collab.	**Mixed**	**Collab.**	**Mixed**
Shared Tasks	10	49	13	10
Responses	0	4	0	0

Table 5: Number of shared tasks and responses.

The results, in terms of number of shared tasks, are promising. But the collaboration effect was severely hindered by the very low response rate, which was essential in completing the engagement loop of this mechanic. We believe that this was mainly due to 1) the anonymity constraints imposed by the two companies; and 2) by the absence of a notification mechanism that could inform employees about tasks shared with them. Instead, we relied on the curiosity of employees, to navigate in the application and check for shared tasks. We deliberately omitted notifications to avoid bias against the control and *Competitive* groups.

Session time. The diminished effect of combining collaboration and competitiveness is also visible in the session time, especially with **ET1**. A Kruskal-Wallis test showed that differences between the experimental groups were statistically significant ($p = .001$) in **ET1**, but not in **ET2** ($p = .951$). This is a first indication that gamification was perceived differently between the two enterprises. We perform a post-hoc analysis only for the experiment in **ET1**. A pairwise Mann-Whitney U tests with Holm-Bonferroni correction found a significant difference between the *Control* and the *Collaborative* groups ($p = .005$), and between the *Competitive* and the *Collaborative* groups ($p = .006$). This signifies that employees with both a leaderboard and task sharing spent significantly less time per session in the application compared to only having basic gamification or only the leaderboards. The difference between *Competitive* and *Collaborative* groups coincides with the results in the previous section. No significant difference was observed between the *Mixed* and any of the other groups. To further prove this assumption we looked into the net time spent by the employees interacting with the gamification elements; we found that employees in the *Mixed* group did not show higher levels of interaction.

Number of sessions. Table 8 compares the number of times employees opened the application across the different treatment groups. We observe results similar to the analysis of session time. We therefore test the hypothesis that when leaderboards or task sharing are present, employees would be motived to open the application more times, while when combined this might result in fewer times using the application. We use a Kruskal-Wallis test, and found insufficient evidence to reject the null hypothesis (**ET1**: $p = .926$; **ET2** $p = .101$). Another way of analyzing engagement of the employees is by counting the time interval between contiguous sessions, as defined in [22]. The results of the empirical cumulative distributions of inter-session times are depicted in Figure 3.

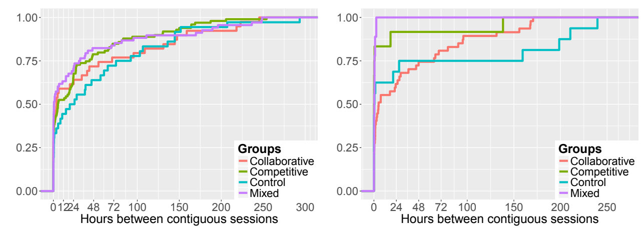


Figure 3: Empirical Cumulative Distribution of inter-session times (hours) for the application across the experimental groups and the two companies (ET1 left, ET2 right)

We notice higher probability of employees re-engaging with the application within 1 or 2 days for the conditions in which leaderboards were present. Also, small differences exist between the participants with task sharing functionality and the *Control* group. This is also an indication that employees who had leaderboards were more inclined to revisit the application.

Task dwell time. Dwell time is defined as the net time spent in task execution. The three task categories feature different levels of complexity, which influence the time spent executing them. We expect the *Information Extraction* tasks to require more time to be completed. compared to the others. Tasks also adhere to different incentives of the employees.

The descriptive statistics are illustrated in Table 9. In **ET1**, employees in the *Competitive* group spent less time executing simpler tasks; and employees in *Collaborative* and *Mixed* groups spent less time on average than the one in the *Control* group. With the *Information Extraction* tasks the differences are less profound, a result that we believe is due to the direct relevance of the task to the company's goal. In **ET2**, there is a general fluctuation of the observed values depending on the task type and experimental group.

We test the statistical significance of the observed differences. Being the dependent variable continuous, and given the positive skewness of our samples, we perform our analysis using a generalized gamma linear model.[4] First, we test the hypothesis that participants in the *Competitive* group would spend less time while

[3] The Negative Binomial regression model has been preferred to a Poisson model due to over-dispersion [6] – $p < 0.001$ – of the data

[4] Fitness to gamma distribution has been verified with a Kolmogorov-Smirnov test.

	ET1		ET2	
#Execs	Coeff.	Sig.	Coeff.	Sig.
Comp.	1.04	**0.013** *	.46	0.46
Collab.	0.25	0.56	.77	0.23
Mixed	-0.18	0.65	-0.08	0.90

Table 6: Negative Binomial regression models describing the effect of game mechanics to # task executions (*: .05 significance).

	ET1			ET2		
	μ	σ	m	μ	σ	m
Contr.	178	197	115	217	291	129
Comp.	180	194	119	226	288	96
Collab.	128	161	73	192	193	133
Mixed	109	202	42	205	286	100

Table 7: Descriptive statistics for session time (in seconds, rounded). μ**: mean;** σ**: standard deviation;** m**: median.**

	ET1				ET2			
	μ	σ	m	**Tot.**	μ	σ	m	**Tot.**
Contr.	4.25	3.04	4.5	51	4.4	3.2	3	22
Comp.	7.43	9.81	4	119	2.85	2.11	2	20
Collab.	4.5	3.42	3	54	16.66	17.78	9	50
Mixed	5.27	6.05	2	95	3.5	3	2	14

Table 8: Descriptive statistics for number of sessions. μ**: mean;** σ**: standard deviation;** m**: median; Tot.: total.**

		ET1				ET2			
	Type	μ	σ	m	**Tot.**	μ	σ	m	**Tot.**
Contr.	*Info.*	85	80	53	4320	101	81	100	1414
	Moral	38	36	24	2665	19	22	11	1340
	Cell	31	29	19	1255	36	20	37	359
	All	**51**	**58**	**32**	**8240**	**32**	**46**	**15**	**3113**
Comp.	*Info.*	101	77	81	9117	55	42	47	1427
	Moral	27	70	12	5987	20	17	13	1630
	Cell	20	20	14	2432	17	12	12	736
	All	**40**	**70**	**16**	**17535**	**25**	**26**	**14**	**3792**
Collab.	*Info.*	93	70	77	2329	99	55	78	2982
	Moral	20	20	13	2438	21	11	17	801
	Cell	32	22	26	958	27	21	19	823
	All	**32**	**41**	**16**	**5726**	**46**	**48**	**29**	**4607**
Mixed	*Info.*	90	62	79	2607	184	124	131	2580
	Moral	35	37	20	2561	28	38	14	1582
	Cell	24	21	15	1119	35	17	27	177
	All	**42**	**46**	**22**	**6287**	**57**	**86**	**19**	**4340**

Table 9: Descriptive statistics for task dwell time (in seconds, rounded). μ**: mean;** σ**: std. deviation;** m**: median; Tot.: total.**

executing tasks compared to the *Control* group – mainly focusing on gathering points and improving their position in the leaderboard faster. The hypothesis can be accepted for the *Cell count* task, where we observe a statistically significant decreasing effect for the task execution time, when only leaderboards are present, in both experiments (**ET1**: Coeff = −.74, Sig = .004; **ET2**: Coeff = −.41, Sig = .013). A similar negative effect is observable in **ET2** with the *Information Extraction* task (Coeff = −.61, Sig = .009). No significant effect could be observed for other configurations.

In **ET1**, the moral decision tasks shows a significant negative effect when task sharing functionality was present (Coeff = −.636, Sig = .029), while in **ET2** the effect is positive but significant only in the *Mixed* group (Coeff = .399, Sig = .034). In a similar way, the *Information Extraction* task features a significant positive correlation for *Mixed* in **ET2** (Coeff = .601, Sig = .023). In **ET1** the effect is positive but mild and not statistically significant.

We believe that those results are only partially explained by the use of game mechanics, as confounding factors such as employees incentives for specific tasks are also playing an important role. When such incentives were loosened, as for example for the moral decision tasks and the *Cell count*, then the role of gamification is more evident.

5.3 RQ2: Impacts of Social Gamification on Work Quality

We focus on the *Cell count* and *Information Extraction*, having more objective outcomes than the moral machine. Due to the lack of a golden standard, the quality of the contributions is calculated based on agreement metrics. To improve the robustness of agreement calculation, we also incorporate the labels obtained for the tasks from the participants in the pilot phase of both experiments in the two companies. In this way we were able to have more labels per task unit and stronger majorities which in turn leads to more robust results. Figure 4 depicts the distribution of agreement scores that we obtained for two tasks across the two experiments. Recall that different agreement metrics are used for the different tasks (Section 3.3). Results of significance tests using the Kruskal- Wallis non-parametric test indicate that there is no statistically significant evidence for a difference between the distribution of quality scores across the experimental conditions, for both task types. The following observations are therefore of interest in the context of the experiment, but not conclusive.

For the *Cell count* task, *Competitive* and *Collaborative* groups provide contributions of higher quality than the control group, indicating that social gamification can contribute to work quality. *Mixed* group in **ET1** yielded better results on average than in **ET2**, a result that we explain in terms of the difference in the total shared tasks. Revisiting Table 5 (number of shared tasks for each group that had collaborative game mechanics) the significant difference between **ET1** and **ET2**'s *Mixed* groups in terms of the total shared tasks might explain the difference in the quality of the contributions we observe between the two companies. Interestingly, quicker task execution times observed for *Competitive* and *Collaborative* groups, as we have seen in our results for task dwell time, comes without sacrifice of work quality.

Results for the *Information Extraction* task vary, as shown Figure 4. In both **ET1** and **ET2**, we observe an average lower agreement for the *Competitive* and *Mixed* groups, and a similar level of agreement for the *Collaborative*, compared to the *Control* one. The different results obtained for the different tasks suggest the potential benefits of such gamification elements are dependent on specific task types.

5.4 RQ3: Gamification and Enterprise Environments

In this section, the focus is on the analysis of the effect of gamification mechanics for different enterprise contexts, so to gain an understanding of how those might affect crowdsourcing activities.

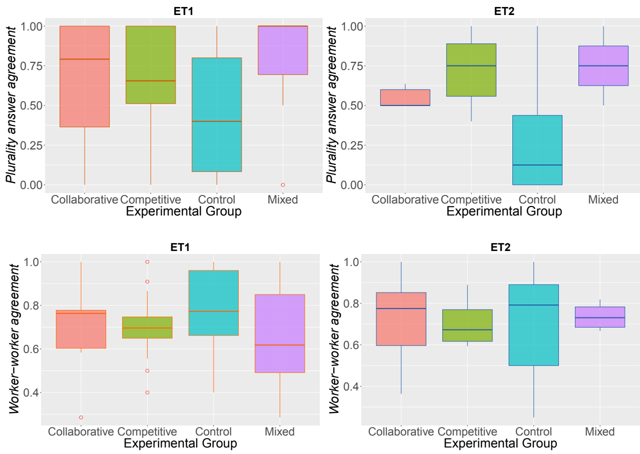


Figure 4: (Upper figures) Distribution of plurality answer agreement scores for the *Cell count* tasks. (Lower figures) Distribution of average worker-worker agreement scores for the *Information Extraction* tasks.

We hypothesise that enterprise environment plays a role in how gamification is perceived, which results in different patterns of crowdsourcing activities.

By juxtaposing the results found in the previous sections we were able to identify some similarities and some differences between **ET1** and **ET2**. Similarities are in terms of the number of task executions and the session time for the different experimental groups, where we noticed higher preference of the employees to the *Competitive* mechanics compared to the *Control*, a small increase when *Competitive* incentives were used, and a diminished effect when those were combined. On the other hand, we observe a slight increase in the number of executions for **ET1** compared to **ET2** when *Competitive* mechanics were used, while the opposite was observed when *Collaborative* mechanics where introduced and when they were combined. We were also able to notice differences in the session times calculated, where in **ET2** we had higher session times on average for all treatments used in our experiments. Furthermore, for the times that the application was opened by the employees, we noticed higher when leaderboards (*Competitive*) were introduced in **ET1** compared to **ET2**; while when only task sharing was used, **ET2**'s employees were more eager to open the application. Regarding the quality of contributions for the *Cell count* task type and the *Information Extraction* task type, by revisiting our results in Section 5.3, we see slight differences between the agreement scores calculated for the same experimental conditions across the different companies. The most profound one is noticed for the *Control* group in for the annotations collected for the *Cell count* tasks.

5.5 Discussion

Experimental results suggest a preference of *Competitive* game mechanics over the *Collaborative* ones. As far as quality is concerned, our experiments showed that depending on the task type, we can expect higher quality contributions when *Competitive* and *Collaborative* game mechanics are used. We attribute the result to the contradicting nature of combining these game dynamics which does not provide a clear goal to the employees while undertaking tasks, from a gamification point of view. Finally, although differences in the perception of gamification were noticed by comparing our two experiments, in a more in-depth analysis we were not able to suggest significant differences between the two companies.

Post-experiment interviews. The use of gamification for enterprise crowdsourcing was viewed positively by the employees who engaged with the application. Informal interviews performed at the end of the experiments revealed that the experimental tool was intuitive and easy to use. One employee stated: *"The use of the application itself and what we needed to do, so fill in a couple of things or make a choice, that was definitely clear"*. Moreover, the gamification elements were perceived as motivating and retained their interest in contributing tasks. An employee revealed: *"At first I was just like, I needed to do the tasks as many times as possible and just contribute to the project. At a certain point I came across the leaderboard and as I am quite competitive that made it a game for me. I wanted to go as high as possible to the ranking"*. Another employee said that progress bars were giving him clear goals and kept him motivated by saying: *"I started with the one with the cars and I wanted to finish this to 100% and then I tried to finish the* Information Extraction *to 75%"*. Surprisingly, even gamification elements that we assumed would not incite great interest, such as the points in the control group and the social gamification group, where leaderboards were not present, proved motivating for the employees. Specifically, an employee from the control group stated: *"it kept me motivated, I tried to reach 500 points at first and then aimed for 1000 points"*.

Validity threats. We consider validity threats related to the history effect, selection and also diffusion of treatment. The history effect is addressed by starting the experimentation almost simultaneously in the two enterprises, so such effects are the same in each participant. We also opted for an observational period which does not contain major public holidays. Flexible sign up times for the participants, however, prevented us to completely control for effects that might arise. A possible history effect could have affected the results in **ET2**, where previous experiments in enterprise crowdsourcing have been conducted in the past. Although new tasks and a new application with gamification incorporated was used in our study, we recognize that the similarity to past studies might have affected the participation and engagement levels towards our experimental tool. We addressed the selection effect by assigning participants randomly to experimental conditions. Signing up for the application was permitted by requesting credentials, and employees were assigned to experimental groups in a round robin fashion. Looking back at the demographics of our experiments, we showed that this strategy yielded acceptable results considering the number of our samples. Diffusion of treatment refers to the potential threat to internal validity in which participants from different conditions communicate with each other. Although we recognize that in an enterprise environment we cannot completely control for this threat, we took care to promote the experiment in an as wide audience as possible inside the two companies, with the intention of recruiting participants from diverse departments. We also believe that the vast amount of departments existent as well as employees working in both enterprises minimize the potential effect of a diffusion of treatment significantly.

6 CONCLUSIONS

With this work, we aimed at furthering the understanding of how gamification can effectively support enterprise crowdsourcing activities, in terms of employee engagement and also the quality of their contributions. Based on Bartle's theory, the exploratory analysis has shown a non mutually exclusive player type characteristics of employees. By combining qualitative research results with those of previous studies on gamification in the enterprise, we were able to show the preference of employees in competitive and collaborative game dynamics. These results informed our explanatory research, for which we deployed a gamified mobile crowdsourcing application that combines competitive and collaborative game mechanics. We used our experimental tool into two large multinational enterprises for an observational interval that lasted two months and involved 101 employees. Results show that competitive game mechanics can better foster engagement than collaborative ones, and that their combination can have a detrimental effect.

As part of future work, we plan to investigate how personalisation can strengthen the competitive as well as the collaborative incentives of the employees especially when task sharing functionality is concerned. It would also be beneficial to study more intricate schemes of gamification such as competitiveness between collaborative groups of employees for crowdsourcing campaigns in the enterprise; and which task parameters mediate the effect of gamification in enterprise crowdsourcing and whether there are some which possibly negate its merits.

ACKNOWLEDGEMENT

The authors would like to thank Robert-Jan Sips for the helpful discussions in the early stage of the project.

REFERENCES

[1] Lora Aroyo and Chris Welty. 2013. Measuring crowd truth for medical relation extraction. In *2013 AAAI Fall Symposium Series*.
[2] Richard Bartle. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research* 1, 1 (1996), 19.
[3] Sarah Bashirieh, Sepideh Mesbah, Judith Redi, Alessandro Bozzon, Zoltán Szlávik, and Robert-Jan Sips. 2017. Nudge Your Workforce: A Study on the Effectiveness of Task Notification Strategies in Enterprise Mobile Crowdsourcing. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 4–12.
[4] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. 2016. The social dilemma of autonomous vehicles. *Science* 352, 6293 (2016), 1573–1576.
[5] Anne Bowser, Derek Hansen, Yurong He, Carol Boston, Matthew Reid, Logan Gunnell, and Jennifer Preece. 2013. Using gamification to inspire new citizen science volunteers. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*. ACM, 18–25.
[6] A Colin Cameron and Pravin K Trivedi. 1990. Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics* 46, 3 (1990), 347–364.
[7] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference*. ACM, 9–15.
[8] Anca Dumitrache, Lora Aroyo, Chris Welty, Robert-Jan Sips, and Anthony Levas. 2013. Dr. Detective: combining gamication techniques and crowdsourcing to create a gold standard in medical text. In *Proceedings of the 1st International Conference on Crowdsourcing the Semantic Web-Volume 1030*. 16–31.
[9] Carsten Eickhoff, Christopher G Harris, Arjen P de Vries, and Padmini Srinivasan. 2012. Quality through flow and immersion: gamifying crowdsourced relevance assessments. In *Proceedings of the 35th international ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 871–880.
[10] Uwe Flick. 2009. *An introduction to qualitative research*. Sage.
[11] David Geiger and Martin Schader. 2014. Personalized task recommendation in crowdsourcing information systems—Current state of the art. *Decision Support Systems* 65 (2014), 3–16.
[12] Ido Guy, Anat Hashavit, and Yaniv Corem. 2015. Games for crowds: A crowdsourcing game platform for the enterprise. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1860–1871.
[13] Juho Hamari. 2013. Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications* 12, 4 (2013), 236–245.
[14] Juho Hamari and Jonna Koivisto. 2015. "Working out for likes": An empirical study on social influence in exercise gamification. *Computers in Human Behavior* 50 (2015), 333–347.
[15] Matthias Hirth, Tobias Hoßfeld, and Phuoc Tran-Gia. 2013. Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. *Mathematical and Computer Modelling* 57, 11 (2013), 2918–2932.
[16] Panagiotis G Ipeirotis and Evgeniy Gabrilovich. 2014. Quizz: targeted crowdsourcing with a billion (potential) users. In *Proceedings of the 23rd International Conference on World Wide Web*. The International World Wide Web Conference Committee, 143–154.
[17] Markus Krause, Marc Mogalle, Henning Pohl, and Joseph Jay Williams. 2015. A playful game changer: Fostering student retention in online education with social gamification. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 95–102.
[18] Janaki Kumar. 2013. Gamification at work: Designing engaging business software. In *Proceedings of the 2nd International Conference of Design, User Experience, and Usability*. Springer, 528–537.
[19] Gioacchino La Vecchia and Antonio Cisternino. 2010. Collaborative workforce, business process crowdsourcing as an alternative of BPO. In *Proceedings of the 10th International Conference on Web Engineering*. Springer, 425–430.
[20] Edith Law and Luis Von Ahn. 2009. Input-agreement: a new mechanism for collecting data using human computation games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1197–1206.
[21] Janette Lehmann, Mounia Lalmas, Elad Yom-Tov, and Georges Dupret. 2012. Models of user engagement. In *Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization*. Springer, 164–175.
[22] Andrew Mao, Ece Kamar, and Eric Horvitz. 2013. Why stop now? predicting worker engagement in online crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*.
[23] Elaine Massung, David Coyle, Kirsten F Cater, Marc Jay, and Chris Preist. 2013. Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 371–380.
[24] Benedikt Morschheuser, Juho Hamari, and Jonna Koivisto. 2016. Gamification in crowdsourcing: a review. In *Proceedings of the 49th Hawaii International Conference on System Sciences*. IEEE, 4375–4384.
[25] Marta Rauch. 2013. Best practices for using enterprise gamification to engage employees and customers. In *Proceedings of the 2013 International Conference on Human-Computer Interaction*. Springer, 276–283.
[26] Maik Schacht and Silvia Schacht. 2012. Start the game: Increasing user experience of enterprise systems following a gamification mechanism. In *Software for People*. Springer, 181–199.
[27] Katie Seaborn and Deborah I Fels. 2015. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies* 74 (2015), 14–31.
[28] Lei Shi, Alexandra I Cristea, Suncica Hadzidedic, and Naida Dervishalidovic. 2014. Contextual gamification of social interaction–towards increasing motivation in social e-learning. In *Proceedings of the 13th International Conference on Web-Based Learning*. Springer, 116–122.
[29] Kristin Siu, Alexander Zook, and Mark O Riedl. 2014. Collaboration versus competition: Design and evaluation of mechanics for games with a purpose.. In *Proceedings of the 9th International Conference on the Foundations of Digital Games*.
[30] Laurentiu Catalin Stanculescu, Alessandro Bozzon, Robert-Jan Sips, and Geert-Jan Houben. 2016. Work and play: An experiment in enterprise gamification. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 346–358.
[31] Osamuyimen Stewart, Juan M Huerta, and Melissa Sader. 2009. Designing crowdsourcing community for the enterprise. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*. ACM, 50–53.
[32] Scott Thiebes, Sebastian Lins, and Dirk Basten. 2014. Gamifying information systems-a synthesis of gamification mechanics and dynamics. In *Proceedings of the 22nd European Conference on Information Systems*.
[33] Luis Von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 319–326.
[34] Luis Von Ahn, Ruoran Liu, and Manuel Blum. 2006. Peekaboom: a game for locating objects in images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 55–64.
[35] Maja Vukovic and Claudio Bartolini. 2010. Towards a research agenda for enterprise crowdsourcing. In *Proceedings of the 3rd International Symposium On Leveraging Applications of Formal Methods, Verification and Validation*. Springer, 425–434.