Gamification in Enterprise Crowdsourcing

Master's Thesis

Grigorios Afentoulidis

Gamification in Enterprise Crowdsourcing

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE
TRACK DATA SCIENCE AND TECHNOLOGY

by

Grigorios Afentoulidis
born in Athens, Greece

Web Information Systems
Department of Software Technology
Faculty EEMCS,
Delft University of Technology
Delft, the Netherlands
http://wis.ewi.tudelft.nl

IBM
Johan Huizingalaan 765
1066 VH Amsterdam, The Netherlands
www.ibm.com

Gamification in Enterprise Crowdsourcing

Author: Grigorios Afentoulidis
Student id: 4521862
Email: gregafent@gmail.com

Abstract

In an era where innovative machine learning and artificial intelligence applications are gaining popularity, enterprises steer their interest to enterprise crowdsourcing, to capitalize on their available human resources to achieve inclusion of in-house human generated data. In this setting, gamification techniques are appealing in order to align employees' motivation to the crowdsourcing endeavor. Although hitherto, research efforts were able to unravel the wide arsenal of gamification techniques to construct engagement loops, empirical studies have been limited to the experimentation of only a few. More importantly little research has shed light into the social game dynamics that those foster and how those impact crowdsourcing activities. In the current study 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 75 and 26 employees in a mobile crowdsourcing application across two large multinational enterprises, 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.

Thesis Committee:

Chair:	Prof. Dr. Ir. Geert-Jan Houben, Faculty EEMCS, TUDelft
University supervisor:	Dr. Alessandro Bozzon, Faculty EEMCS, TUDelft
Company supervisor:	Dr. Zoltán Szlávik, Center for Advanced Studies, IBM Netherlands
Committee Member:	Dr. Cynthia Liem, Faculty EEMCS, TUDelft

If you deconstruct Greece, you will in the end see an olive tree, a grapevine, and a boat remain. That is, with as much, you reconstruct her.

— Odysseas Elytis

Preface

What a journey! In all honesty if I were to verbosely express my gratitude for all people involved in this study and their valuable contributions, I might have needed to cover the length of this study once more.

Throughout my life as a student, I was truly lucky to be surrounded by great and inspirational professors and teachers, who shaped my way of thinking and molded my character. My supervisor in this work, Alessandro Bozzon, is undoubtetly in this list of people. I owe him a great deal of gratitude for his constant support and guidance and most importantly for teaching me how to think and be rigorous in my work.

I am also really thankful to my company supervisor, Zoltan Szlavik, for providing me with all the necessary resources to succeed and for teaching me by example how to operate in a business environment. I am very much obliged to Muriel Serrurier Schepper, her enthusiasm, constant motivation and invaluable help were crucial factors for the success of this project. Special thanks go to Manfred Overmeen, for his support in the technical part of this work and his gregarious and hearty attitude which instilled positiveness in me and helped me overcome many obstacles. I would also like to thank Robert-Jan Sips and Pierre De Wit who gave me the opportunity to do this project in two large companies.

Last but not least, I owe a lot to my parents and four sisters. Their support and love not only accompanied me through my studies as a master's student in TU Delft, but also throughout my whole academic life. They have put their trust in me every step of the way while teaching me the importance of working hard, ethically and honestly.

Grigorios Afentoulidis
Delft, the Netherlands
August 8, 2017

Contents

List of Figures

Chapter 1

Introduction

In 2005 the term crowdsourcing was coined by Jeff Howe to describe "the idea of outsourcing a task that is traditionally performed by an employee to a large group of people in a form of open call" [33]. This new paradigm which builds upon the idea of harnessing the collective intelligence of the crowd, has gathered since then, a great amount of interest by academia [53]. There are three main reasons which can be listed to explain this phenomenon [79] and also advertise its usefulness. First crowdsourcing provides an ideal candidate for tasks that are not hitherto completely solvable by machines (e.g. opinion, geometric reasoning tasks) but still trivial for a human to perform. Second it is a highly cost effective way for data procurement (e.g.labels, creative solutions) and third it can virtually take advantage of a huge candidate amount of people willing to contribute, thus severely expediting the process. In a more laconic way crowdsourcing is cheap, fast, scalable and also an ideal way to underpin shortcomings in current technological advances which unavoidably require human intervention and intellect.

Those benefits are not unnoticed by enterprises [75] who are eager to shift from the traditional outsourcing paradigm, which bolsters their business processes and needs, to crowdsourcing [33]. This gives rise to enterprise crowdsourcing which transfers the practices of crowdsourcing from the online environment to the internal crowd of the enterprise: the employees. The attractiveness of crowdsourcing within enterprises is justified on the premises of its inherent features [30]. Those are the ability to use tasks that are confidential and thus not suitable outside the enterpise boundaries and the benefit of utilizing employees' working capacity to the maximum, while also relying on pre-established internal knowledge about employees' expertise for quality task contributions. This increased attention has been translated to a variety of applications that have been deployed within enterprise environments with noticeable success [36] [66] [40].

Even though employees form a crowd with unique characteristics as compared to external crowds, the traditional challenges of crowdsourcing regarding participation and retention to the crowdsourcing endeavor and also quality output remain [75]. Those challenges are even more intensified by the fact that employees are engaged with their

everyday work, which results in limited free time and specific motivations not always aligned with the crowdsourcing effort. So their participation in the enterprise crowdsourcing endeavor could be severely hindered. To this end the understanding of which are the required incentives and motivations to achieve the desiderata is an active research area [76] with lots of interest.

Gamification on the other hand can be described as : "A process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation" [34]. The affordances or game mechanics are usually drawn from traditional games and the value creation is generally reported as an effect of increased engagement of the user towards the service or the system [51] [29]. It is however widely accepted that introduction of gamification involves several non trivial steps that require strong consideration and scrutiny in order to achieve its goals [51] [73] [54] [39]. In that sense adhering to a good gamification design provides the ideal interface for enterprise crowdsourcing to draw upon the benefits that the first ballyhoos. It is still however unclear, from the research point of view, what is the interaction of different game mechanics and game dynamics to crowdsourcing activities and especially in the context of an enterprise.

1.1 Problem statement

Viewing gamification as a process of constructing user engagement loops on existing services, its combination with enteprise crowdsourcing seems ideal in order to remedy the challenge of crafting incentive mechanisms. It has been reported by previous studies that gamification can indeed intrinsically incentivize the crowd and drive its behavioral outcome towards augmented and prolonging participation and task contribution as well as quality output [29] [17] [72] for their work.

However it is still unclear which game mechanics are more suitable in enabling crowdsourcing within an enterprise. This is mainly because gamification techniques are not always necessarily tied to the motivations of the crowd. Thus the full potential of gamification is stymied by a limited selection of gamification affordances in studies, which are generally viewed as one-size-fits-all, sidestepping more often than not the underlying incentives of the crowd for participation and underestimating the dynamics and their impact in human motivation. More importantly limited research has been focusing into evaluating the interplay between game elements and social incentives, especially in an enterprise context in which synergy and competition are concepts that play important role.

So there exists the need to decouple game mechanics and the dynamics that those spark, in order to examine their effect on crucial parameters of crowdsourcing activities such as engagement and data quality. In most cases the application of gamification is studied as a consolidation of different game mechanics thus obscuring how those affect the desired outcome in isolation [51].

1.2 Research motivation and objectives

Driven by our research motivation, in this work we focus on studying gamification in enterprise crowdsourcing. We deploy a crowdsourcing mobile application as an experimental instrument in two large multinational corporate environments; IBM and Rabobank. We incorporate gamification techniques to study engagement and data quality.

Our focus in this work is to address the following main research question:

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

In order to answer this research question we identify the following objectives for our work that need to be fulfilled:

1. **Objective 1: A better understanding of the motives of employees of an enterprise behind participation in a gamified enterprise crowdsourcing application**
   This objective clarifies the main requirements for selection of suitable tasks within an enterprise and also how we can inform adequately our gamification design, based on the incentives of the employees. First we need to control for confounding variables such as task characteristics and task granularity that affect motivation [51]. Second we should analyze the incentives of the internal crowd which will inform us of the gamification techniques that are suitable. This is because different persons represent different players in a gamified context with also various needs [18]. The fulfillment of this objective will be based in our literature survey on enterprise crowdsourcing and crowd incentives and also by qualitative analysis of interviews with the employees that will help us justify our gamification techniques' decisions based on the crowd's characteristics.

2. **Objective 2: Operationalization of the concepts of reliability and engagement**
   We should choose the appropriate metrics for reliability in crowdsourcing as quality assurance and also for engagement in crowdsourcing contributions. Quality assurance is dependent on the task types and their content and will be informed by Objective 1 along with state of the art metrics from our literature study. Engagement metrics will be based on relevant studies in the field of crowdsourcing.

3. **Objective 3: Development of an experimental instrument and experimental design**
   Objective 3 is dependent on the outputs of Objective 2 and 1 above. Our experimental instrument is a gamified mobile crowdsourcing application which incorporates task types on which data quality can be measured and also incorporates

tasks for support of a specific business need. Based on a devised experimental design the instrument will include different gamification techniques which will be disseminated to different treatment groups to test their effects on the measured variables.

4. **Objective 4: Analysis and understanding of which gamification techniques affect user engagement and quality assurance in an enterprise**
   This objective relates to answering our main research question. We need to be able to measure the variables of interest in our instrument and then choose suitable data analysis to gain an understanding on how different game mechanics affect engagement and data quality in enterprise crowdsourcing.

5. **Objective 5: Analysis and understanding of the effect of contextual factors in gamified enterprise crowdsourcing**
   With this objective we are aiming in addressing the potential generalization aspect of our work. We repeat our experiment in two different corporate environments and study the potential effects of contextual factors in the adoption and effectiveness of a gamified crowdsourcing system.

A schematic roadmap of our research on which the objectives that were previously defined are mapped, is provided in Figure 1.1. Our literature review study, satisfies Objectives 1 and 2 by providing us research background on the application of gamification in enterprise contexts in terms of tasks, gamification theory and also metrics of engagement and quality of task contributions based on previous studies on the field. We leverage on our literature review findings to fulfill the requirements of Objective 1, tailored to our specific use case. Namely, we deploy qualitative research methods by conducting employee interviews to recognize the predominant enterprise player types necessary for the selection of our experiment's independent variables. We also use expert interviews to identify a suitable enteprise crowdsourcing application and inform the design of our tasks. We meet Objective 3 by designing our experimental methodology, developing our experimental tool and executing our experimental protocol. We finally address Objective 5 with quantitative analysis of the experiment's results and report our findings in order to answer our study's main research question.

1.3 Contribution

In the current work, we study the application of gamification techniques in enteprise crowdsourcing contexts to gain an understanding of their impact to crucial parameters of crowdsourcing activities such as employee engagement and quality of task contributions.

The first contribution, is a qualitative exploratory analysis of the dominant player types existent within the enteprise. Based on Bartle's theory of player types, and by combining results found in previous studies on gamification in the enteprise, we highlight the importance of the social characteristics of the workforce that inform the design of gamification in an enteprise context.

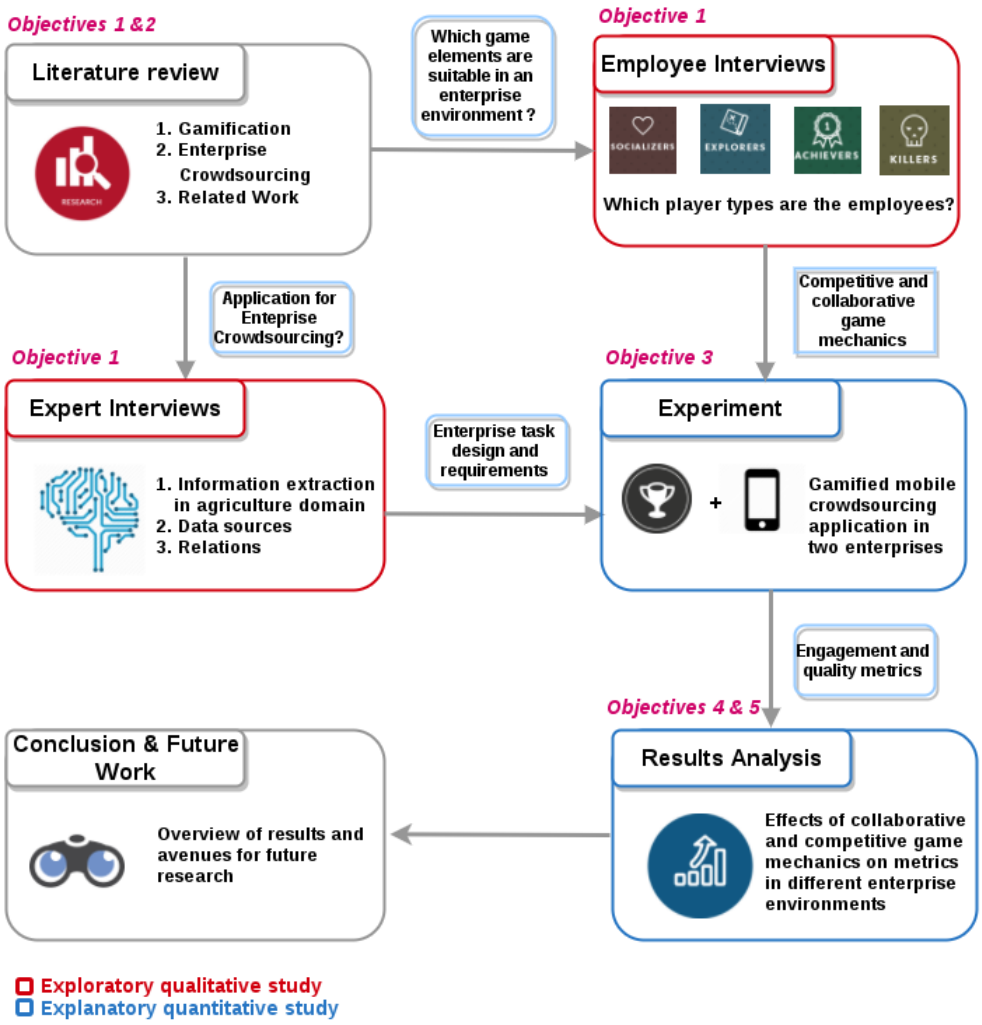

Figure 1.1: A roadmap of the research

The second contribution, is the extension of a mobile enterprise application with pluggable gamification elements. We implement competitive game mechanics by designing a scoring function based on the number and quality of contributions as well as a collaborative gamification mechanism which supports task sharing capabilities within the enterprise to foster community collaboration.

The third contribution, is the analysis of competitive and collaborative social game dynamics though experimentation with a mobile enterprise crowdsourcing application. We apply those two aspects of social gamification on top of traditionally employed game mechanics found in the literature, to study the effects of synergistic and competitive dynamics in engagement and data quality in enterprise crowdsourcing.

The fourth contribution is a novel comparative analysis of a gamified crowdsourcing application in two large enterprises. This part of the work aspires in gaining a better

understanding of the contextual effects that might exist between the relationship of gamification and crowdsourcing and how they mediate it.

1.4 Thesis outline

In the remainder of the document we first provide our literature study in the domain of enterprise crowdsourcing and gamification in Chapter 2 .

In Chapter 3 we first report our findings on the qualitative research part of our study which will be used in our discussion of our adopted research methodology and experimental design which are included in the same Chapter.

In Chapter 4 we present the implementation details and design choices for the experimental instrument which was developed as part of this study along with the tasks used for our experimentation.

In Chapter 5 we present our statistical analysis on the results of our experiments in both IBM and Rabobank and report our findings in order to answer our research question.

Finally, in Chapter 6 we reflect on the results of our study and discuss possible future directions of the current work.

Chapter 2

Related Work

In this section we provide our results elicited from our literature study in the fields of crowdsourcing, enterprise crowdsourcing and gamification . Our study in crowdsourcing and enterprise crowdsourcing will provide us with the necessary background to identify tasks usually used, incentives of the crowd and also state of the art measurements for engagement and quality assurance. Finally our study in gamification, will give us necessary information on how gamification and game mechanics in specific are used in the literature, especially in a crowdsourcing context.

2.1 Enterprise crowdsourcing

Enterprise crowdsourcing emerges from the application of crowdsourcing practices within corporate environments. A unanimously articulated definition that would stem from the specific characteristics drawn from this pairing, is yet to be expected. Instead studies that fall in this domain are mainly focusing on two aspects in their attempt to define it [30]. The first is related to the crowd on which enteprise crowdsourcing is directed, and that is the employees of an enterprise. This type of crowd differentiates it from traditional crowdsourcing, which refers to an open call to a non predefined crowd usually found via the Internet. The second refers to the problems that enterprise crowdsourcing is usually focusing in addressing. Usually these are business problems that are first translated to tasks and then solutions of these tasks can solve bigger organizational problems. Based on those two key differentiating characteristics, opportunities and challenges are emerging.

The main opportunities of enterprised crowdsourcing can be summarized as follows[30] :

1. **Use of business critical and confidential tasks**
   Since in many business problems confidential information and data are naturally existent that should only be disseminated in the internal corprorate environment, enterprise crowdsourcing has the advantage that it is inherently aligned with the confidentiality policies of the enteprise [31] [66]. This also allows for a

wider spectrum of problems in which enterprise crowdsourcing can effectively be used.

2. **Leverage on non utilized working capacity**
   Enterprise crowdsourcing can profit on non utilized working capacity by offering an alternative for employees to provide work for the goals of the enterprise.

3. **Benefit from internal business networks**
   Already established business relationships and communications between organizational units of the enterprise can prove beneficial for enterprise crowdsourcing for expert crowd selection or task recommendation purposes.

On the other hand, main challenges that enterprise crowdsourcing has to face are [76] [75]:

1. **Intellectual property rights**
   The contribution of the employees to the crowdsourcing endeavor presupposes work which adheres to intellectual properties legislation. The rights of this work should be regulated by the enterprise with suitable contract terms. As a result, crowd selection within the enterprise is conditioned to the existance of such terms.

2. **Incentives**
   In a corporate environment the main focus of the employees in on accomplishing their daily duties and tasks. In this context participation and commitment in enterprise crowdsourcing can be severely hindered [26]. To this end a strong consideration of the motivation of the crowd and fine engineered incentive mechanisms should be used to overcome this obstacle.

3. **Information security**
   Even though the usage of confidential information is a virtue of enterprise crowdsourcing, as explained before, it is also a challenge. Usually confidentiality is preserved not only with regards to the enterprise boundaries but also within internal boundaries. In that sense a careful usage of this information in enterprise crowdsourcing has to be devised to respect the information sharing policies.

Looking at the challenges above, it is clear that from a design and implementation point of view the most crucial challenge is that of understanding the incentives of the employees in order for enterprise crowdsourcing to maximize its potential.

The main dimension across which incentives are categorized in crowdsourcing is that of intrinsic and extrinsic motivations [32]. Intrinsic motivation is related to the characteristics and nature of the work which inherently drives motivation to complete it rather than expecting something in return. On the other hand extrinsic motivation is initiated by the anticipation of a form of reward which is unrelated to the nature and type of the work; this can usually be in the form of monetary rewards, prizes, recognition etc.

The same categorization of incentives is recognized in enteprise crowdsourcing as in public crowdsourcing with the main distinction being, that extrinsic motivations are usually non monetary based since this might conflict already established compensation arrangements with the employees [75]. Also usually, contradictory to the predominant money-based rewards in public crowdsourcing [19], intrinsic motivations are leveraged for successful application of enterprise crowdsourcing. In [65] they claim that identifying what is the main interest of the employees in terms of personal values, causes and actions is pivotal. By surveying employees of a company they found that the main motivations towards contributing work were to have a positive impact for the world or community, to gain recognition or rewards from the employers, to be able to learn and to be able to participate in an interesting and groundbreaking project. In a similar survey in [6] they found that the main motives of employees were to learn something new, improve the output of the company, to contribute to their work community, improve appraisal for their work and lastly to have fun. In addition to those incentives it is also crucial for the enterprise crowdsourcing endeavor to be able to deal effectively with the onboarding of the employees into the crowdsourcing initiative by offering low entry barriers and also clearly defined tasks [75].

Instigated by the employees' incentives as described before, applications of enterprise crowdsourcing are mainly falling into two categories [30]. The first category contains applications of innovation competitions and software testing and development. As an example, in [66] enterprise crowdsourcing has been used to collect translations of sentences in various languages (e.g. Chinese, German, Italian) to English, with the purpose of training statistical machine algorithms for the same task. In this category learning and rewarding incentives are mainly employed. In the second category applications that are part of developing a new business process or strategy or developing an already existing one are contained. Example of these, among others, are supporting document translation [65] and IT inventory management [76]. In this category incentives that are related to ameliorating the enterprise's performance and also contribution of the employees to their work community are used.

Summarizing our findings in the domain of enterprise crowdsourcing and merging employees incentives found in previous studies as well as typical enterprise crowdsourcing applications we can conclude, that in most cases intrinsic motivations are leveraged in a corporate environment. Those are mainly related to the need of employees to express altruism in their immediate environment by providing crowdsourced work. Also important is for the employees to feel that there is an alignment of the EC initiative to their own interests. The main interests is the feeling of learning something or getting a reward or appraisal for their contribution.

2.2 Engagement in crowdsourcing

User engagement is an elusive notion which can not be easily captured and expressed quantitatively with confidence, for the needs of empirical studies. In a laconic and simple definition authors in [44] define engagement as "the quality of user experience that emphasizes the positive aspects of the interaction, in particular the phenomena

associated with being captivated by an application". By decomposing the previous definition we understand that the positive aspects are application dependent. That means that conditionally to the application on which user engagement is measured, we would expect different metrics to be used.

In this work we are interested in defining metrics for user engagement in crowdsourcing so we are focusing only on metrics found in relevant literature. There are two positive aspects that are crucial for a successful crowdsourcing initiative and which also denote augmented engagement from different perspectives. That is the quantity of the tasks solved by the workers and also the quality of the output that they produce. Metrics found in the literature are trying to operationalize engagement on the basis of those two parameters.

Regarding quantity of work produced, in [35] they measure the total number of tasks submitted by each worker. A closely related metric is used in [73] where they define throughput as the average number of tasks contributed by a worker within a specific time frame (e.g.human hour,session [45]). In [20] instead, they use the average amount of labels produced per task. The difference of the latter to the previous, is that it accounts for the total contribution of the crowd for each task instead of measuring it per worker. Finally retention curves are proposed in [41] which are graphs of the total amount of games played against the number of players. Less steep lines in this graph indicate better retention of the crowd towards the system. In the crowdsourcing setting this can be easily adopted for number of tasks contributed.

Engagement which is based on quality is measured by assessing the output of the workers. In [35] they use the number of correct answers for all contributed tasks for each worker. In the same study they also propose the use of information gain which tries to capture the "valuable" contribution of the worker by combining quality of work with quantity. In [73] they use average lifetime play (ALP) which measures the average amount of work time of each worker across all participants. In [45] the dwell time is used to denote the amount of time spent for each task as an indication of contribution being a product of adequate amount of consideration from the worker. Finally in [17] the intra-annotator agreement of the labels produced is calculated as an indication of consistent quality work produced by the workers.

In this section we studied related literature which used metrics of user engagement in crowdsourcing. In general three categories of metrics for user engagement can be recognized [44] and those are self-reported engagement (i.e. questionnaires, interviews), cognitive engagement and online behaviour metrics. The latter are objective metrics of user engagement as the ones we discussed previously, and are measuring interactions of the user with the system by recording users' activities. For the purposes of our study in engagement we use metrics pertinent to those discussed in this section. The metrics used for our study are further explained in Section 3.3.3.

2.3 Quality assurance in crowdsourcing

In general, quality control in crowdsourcing is a term that covers all the available techniques with main purpose a better and less noisy result from the contributions of the crowd. In that sense it can be applied in different stages of the pipeline from the task requester to the crowdworker which contributs data with different forms [1]. Depending on the stage of this pipeline on which quality control is focusing, diverse techniques which consider different aspects are suitable for an application. For example quality control can be considered as part of the task design from the perspective of the requester. In the latter, efforts are made for the tasks to be as comprehensive as possible and easily solved, in order to minimize the possibility of extracting fallible annotations. There are also techniques that focus on workers' profiles and their reputation on which a requester can base his decision for high quality worker selection. Such solutions are generally provided by the crowdsourcing platform and either by manually assessing the quality of each worker either by the requester or by funneling tasks for which ground truth is already known as a form to empirically test the quality of the worker.

Label procurement for the purposes of training models is traditionally performed by a handful of experts in a domain [2]. Although the quality of the labels provided is preserved in this way, this task is cumbersome and usually needs a significant amount of time. Although crowdsourcing can remedy this bottleneck, it comes with a tradeoff. That is the quality of the labels produced by the crowd. Several reasons for this phenomenon exist, such as the lack of appropriate knowledge from workers, loss of interest or dedication [60], to even biased interests which can lead to adversarial behaviors and contributions [1]. This is why a vast amount of studies in crowdsourcing are devoted in adressing the problem of quality.

However to be able to assess with a post-hoc analysis the quality of the produced work by the crowd, there is the need to know the ground truth or golden standard of the tasks offered. Usually the tasks are annotated by several workers therefore a series of labels is provided for each instance. In that sense the problem is translated into inferring ground truth labels from a series of uncertain or noisy labels against which the quality of the workers is evaluated. A considerable amount of work is devoted in solving this problem which we analyze further here.

A simplistic and intuitive approach in inferring a label from a noisy set of labels is by applying the majority vote rule to establish a ground truth [60] [2] [63] [52]. This rule dictates that the final label is the one that is encountered most frequently among the answers of the crowdworkers for each given task. In case a tie is encountered, then random assignment of the correct label between the tied ones is opted. Mathematically the rule can be expressed as shown below for the binary label case that can easily be generalized to multiple possible labels:

$$y_{est_i} = \begin{cases} 1, if & (1/R)\sum_{j=1}^{R} y_i^j > 0.5 \\ 0, if & (1/R)\sum_{j=1}^{R} y_i^j < 0.5 \end{cases}$$

Where in the formula above y_{est_i} is the final estimated label, R is the number of labels available per instance and y_i^j is the label provided by the j worker for the instance i. Although it has been proven robust in numerous cases in inferring ground truth labels [2] [60] [52], majority voting fails to capture important aspects of the labeling process. More importantly it makes the strong assumption that all workers are equal in terms of quality (i.e. the probability of supplying the correct label) and their "vote" has the same weight in estimating the final label [55]. We however already stressed that in realistic crowdsourcing scenarios, a variability of skills and competences is usually encountered that majority voting fails to accommodate.

A slightly more ingenious idea attempts to cope with the equal quality assumption of the workers by introducing weights for their votes [63]. Weights can better describe the contribution of each worker for the final label estimation by promoting higher weights to those who are deemed to be more skilled and have higher quality in general and assign lower weights to those who do not match these criteria. The final label can be found again by using the threshold formula found above, by multiplying the label of each worker with his weight. For example, in this setting zero weights can be assigned to workers who are deemed as spammers (i.e. they contribute labels at random) and negative weights to workers who are purposefully providing adversarial labels, so that their labels can be switched to the correct ones. However there are two disadvantages related to this. The first is that estimation of the quality of each worker has to be tested empirically on a set of instances or tasks that a ground truth is already provided. In general this is not always existent or it is costly and tedious to provide. What is more, those qualities are calculated only once and cannot alter the weights of the votes to account for the possible learning effect of the workers. The latter is describing the case where a worker can become more and more skilled in a specified task as his experience with it is increasing. Finally, there exist methods that target to worker selection for the tasks with the expectation that majority rule will yield higher quality ground truth labels and more robust results when majority is considered among a carefully selected high quality set of individuals. For example in [70] the problem of worker selection is formulated as an integer programming problem while in [15] the problem is viewed as an optimal action selection one.

Aside from the majority voting method for inferring ground truth labels, several other techniques exist that are based on the Expectation Maximization algorithm (EM). The EM algorithm is a well known heuristic algorithm that tries to iteratively discover the joint distribution of a set of random variables when some of them are directly observable and some are latent (i.e. their values are not known but their contribution is acknowledged). For the EM algorithm to be able to be applied in the problem of label inference in crowdsourcing, a probabilistic model needs to be devised that accounts for the stochastic dependencies of the labeling parameters to the final label

provided and the unknown true label. For example in [78] the quality of each worker (i.e. the probability of producing a correct label) along with the difficulty of the task are modeled and their interaction produces the observable labels conditionally to the true ones which are latent. In [77] expertise and bias are added to the previous model to describe a more detailed dependency of the labels from the workers abilities. The final outcome of the EM algorithm after the appropriate modeling of the labeling procedure is an estimation of the true labels as well as the individual qualities of the workers. Two main problems arise in those methods however; the first is that EM does not guarantee optimal solution and thus the estimated true labels are still noisy and also a great amount of labels is needed for an acceptable solution which is not always feasible.

Another way of analyzing the quality of the labels provided by the crowd instead of inferring the ground truth, is to calculate the annotator agreement between the workers as an indication of quality of the labels provided by each worker. Inter-annotator agreement [12] is calculated for each worker as the average pairwise agreement of him and and every other worker across all workers [52], [63]. Those methods indirectly calculate the performance of a worker against the majority vote but they offer a higher granularity as compared to the methods explained previously. In [4], however the authors claim that low agreement between worker might not necessarily entail low quality rather task clarity and label ambiguity might also be the source of this phenomenon. Although their study is focusing in relation extraction from sentences the same principles apply in any interaction of a worker (interpreter) and a sign (e.g. image, video ,sentence) that requires a referent for that sign (e.g. label, annotations). To this end, three categories of metrics are proposed [3]. In the first category metrics related to the workers are defined and signify the quality of each worker based on label disagreement with the rest of the workers. In the second category task related metrics are defined which assess the difficulty of the task in terms of comprehension and in the third category metrics that measure the ambiguity of the possible labels are considered.

Summarizing the results of this section we looked on state of the art quality assurance techniques. Those are based either in determining the true labels of the instances or tasks and then assessing the quality of each worker on the inferred ground truth or in calculating the inter annotator agreement or disagreement for each worker relative to the labels provide by the rest of the crowd.

2.4 Gamification

Although gamification does not have a well established definition in academia [59], two seminal works synergistically provide the main characteristics of the field that are widely accepted. According to [14], "Gamification is the use of game design elements in non-game contexts". In this definition a clear separation between gamification, playing and games is attempted. Gamification is differentiated from playing in the sense that it does not relate to the free-form and improvisational characteristics of the latter. At the same time it is distinct from a game because it targets in producing a gameful experience borrowing game design principles, rather than becoming a fully developed

one. This allows gamification to not restrict its use cases to only entertaining contexts. The second definition coming from [34] states that, "Gamification is a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation". In this definition gamification is understood in a more abstract level by targeting to the outcomes of the gameful experience produced by it as perceived by the actions of the users. In contrast to the first definition, here the gameful experience and the augmented value creation that this entails are the only criteria that distinguish gamification, free from any specific context or game design element restrictions.

Regardless of the point of view from which gamification is described, both definitions agree upon the existence of two important aspects of gamification. The first is the gameful experience in the interaction of the user with a gamified system. This experience is encompassed in the psychological outcomes that are perceived by the user from the application of motivational affordances. Those can be increased motivation towards an action, enjoyment or a specific attitude [29]. Moreover those motivational outcomes are the main cause for specific behavioral outcomes from the users that can be summarized into increased engagement and positive patterns in service use [29], such as increased user participation [59]. The dynamic of this interaction has been proven to be strong enough to turn tasks that are inherently monotonous and repetitive, into being more enjoyable and fun to undertake [21].

The second is the application of motivational affordances in the form of game elements or game mechanics which work as the originating force for the interaction described before and are inspired from games. A taxonomy of the game mechanics is dependent on the axis chosen to implement it. In [58] a separation of the game mechanics is attempted on the basis of being in-game or in-person mechanics. The first category relates to mechanics that are directly implemented in the gamified application such as achievements, points and countdown (i.e. time pressure). The second category relates to mechanics that work jointly with the specific characteristics of the user and its motivations. Such mechanics can, for example be, envy, epic meaning (i.e. the sense of contributing for a larger cause or the greater good) and loss aversion. Main difference is that for the latter, their existence is dependent on the current disposition of the user. In [67] game mechanics are described as, "functional components of a gamified application that provide various actions, behaviors and control mechanisms to enable user interaction" and game dynamics as those that "determine the individual's reactions as a response to using the implemented mechanics". Contrary to the first taxonomy in this study dynamics or in-person mechanics are thought to be potentially sparked by the interaction with the game or in-game mechanics existent in the system and not independently. According on what behavioral outcomes those target and their meaning, they can be incorporated in one of the clusters of system design, challenges, rewards, social influences and user specifics.

Of paramount importance for the success of gamification is the consideration of the context in which it is applied [29], thus signifying that gamification goes beyond the mere utilization of game mechanics. The context can be separated on the application type and the user type. On the application level the intentions of use of the users is a

crucial factor for the effectiveness of gamification. For example in [27] a discrepancy in the behavioral outcomes is noticed in traditional games and studies in a utilitarian service when badges are applied as game mechanics. In a similar manner in [10] they stress the fact that different motivations for using a gamified citizen science application emerged than those initially expected, that focused on fun and community involvement rather than competition. The user type level relates to how the end users perceive gamification and how they react in its presence. In [34] they underline the subjective nature of gamefulness and the means to produce it for different users. Recalling the importance of gamefulness for behavioral changes from our previous discussion, the effect that this entails to the outcome of gamification becomes apparent. There are also studies that show that gamification is not necessarily perceived by the users that interact with gamified systems [82] thus invalidating their usage, while they also report that in a pilot test in Foursquare badges were preferred as a game mechanic by the users over points.

In this section we introduced gamification and discussed its connection to augmented user engagement towards gamified systems and also the concepts of game mechanics and game dynamics which are important in order to gain understanding on how it is applied. Finally we discussed how contextual factors such as the application domain and user types are critical for the employment of gamification. The findings of this section lead the literature study and the discussion that follows. We first look into relevant literature to understand how gamification is applied in crowdsourcing, then how gamification is applied in an enterprise environment and then we discuss the different gamer types and their relation to different game mechanics. Finally, in the last two sections we concentrate in more detail on some game mechanics that will be considered in the current study.

2.4.1 Gamification in crowdsourcing

The union of the research fields of gamification and crowdsourcing follows an interesting course. That is because the first studies attempted in the intersection of these two, emerged even before the terms of crowdsourcing and gamification have appeared. Those studies focused on solving traditionally difficult tasks for computers with input from humans as a side effect of a playing activity that they participated in [73]. From those studies the term Games With A Purpose (GWAP) emerged to denote the novel concept of disseminating tasks that required humans to be solved, concealed in a game that incites enjoyment. Notable examples of this paradigm were the ESP game [72] which collects labels for images from the web by creating a game where matching tags provided by two players are rewarded, TagATune [42] that works similarly with tags for tunes and also Peekaboom [74] which helps locate objects in images. Those seminal works were the forebears of gamification in crowdsourcing unraveling the highly promising results in terms of user participation and quality of output [72], but also provided some useful guidelines for the proper application of gamification. According to [73] they suggest that a successful gamified crowdsourcing endeavor has to have clear rules and winning conditions for the participants and adequate challenge to achieve its maximum potential. This is in accordance with the theory of flow [13] which indicates that a balance between the skills of an individual and the challenge of the task has to be

struck in order to achieve maximum immersion. What is more, they also provide some useful suggestions for game mechanics to be used in the context of crowdsourcing, such as time challenges, score keeping, player skill levels in the form of badges, and high score lists or leaderboards.

A fundamental difference of the studies in GWAP from the rest in gamified crowdsourcing, is that in the former their approach starts from defining the game and introducing the crowdsourcing tasks in it. Typically the problem that is to be crowdsourced is already defined and gamification works as an added layer to enhance engagement and ameliorate crowdsourcing activities. In a literature review study they identify the ten more used game mechanics among studies in the intersection of gamification and crowdsourcing as listed below [51]:

1. Points/Score
2. Leaderboards/Ranking
3. Badges/Achievements
4. Levels
5. Progress
6. Feedback
7. Rewards
8. Storytelling
9. Missions
10. Virtual Territory

In the same study, based on the crowdsourcing categories suggested in [23], they stress the differences that exist in game mechanics used. Namely in crowdprocessing and crowdrating applications where a worker processes or rates individually tasks, simpler game mechanics are used that target in competitiveness and collecting achievements. For this reason game mechanics such as points, badges and leaderboards are mostly employed. In more creative versions of crowdsourcing, where the crowd is asked to provide content, or contribute a novel solution to a problem more involved game mechanics are suitable which promote collaborative game dynamics and social influence. In this setting rewards, progress, social status, curiosity and altruism are widely selected.

We are also interested to find out from previous studies useful insights that would potential yield design principles for gamification in crowdsourcing. As far as the task types is concerned increased motivation for task contribution attributed to gamification has been witnessed for a variety of tasks. Gamified crowdsourcing applications have

been used for entity and relation extraction in medical texts to resolve semantic ambiguity [16], for labeling public locations [25], for relevance assessment [17], to collect data for environmental community activism [47] and to collect phenology data in a citizen science application [10] among others. Some studies also collected qualitative data to understand the motivations of users behind using a gamified crowdsourcing application. Namely in [10] having fun is a driving factor expressed as creativity, exploration and relaxation. Also they noticed the need for learning something new, being part of a community while playing, competing and socializing. Similarly in [16] they found that learning and competing in an enjoyable way where the most important motivations in participating.

Comparing with the incentives discussed in Section 2.1 for enterprise crowdsourcing we see that there is not a significant alteration of the incentives reported in enterprise crowdsourcing and gamified crowdsourcing with the exception that users expect an intriguing gameful experience as an addition. What is more, cheating is a consideration when applying gamification [17] [73] [72] where users will try to exploit rules of the gamified application for their own benefit in spite of reports that gamification in general attenuates such phenomena [51]. For example in [17] users that keep disagreeing to the consensus labels are not allowed to continue participating while in GWAPs, by design the game is able to not accept strange labels that match from the players that play together. Finally as mentioned in [51] task granularity and characteristics as well as perceived motivational affordances are important covariates for the motivation of the users in a gamified crowdsourcing application. That means that for gamification to be pervasive in this context tasks should be clearly structured and easy to understand while any usability issues should be resolved [10].

2.4.2 Enterprise gamification

The virtues of gamification are not unnoticed by organizations and enterprises which apply it to increase engagement of their workforce in their daily tasks and their interaction with information systems. One main reason for this increased interest is the gradual transformation of the age demographics of the employees in recent years where millennials tend to form the majority of employees in companies [54] [39]. This generation is nurtured and entertained with video games, so gamification is more appealing to them. Another reason is that gamification is flexible in addressing a variety of business processes and needs, in an efficient manner [54]. Although those two reasons indicate an enterprise environment is a fruitful one for gamification, the unique characteristics that this also bears call for careful consideration while applying it, to avoid nullification of its benefits. Merging best practices suggested in literature [54] [39] we list them below:

1. Clear definition of the business need and mission
2. Gaining a clear understanding of the players, their motivations and also their context
3. Facilitate onboarding for novices and let user evolve while playing

4. Selection of the appropriate game mechanics

5. Design appropriate game rule according to the selected game mechanics

Notwithstanding those guidelines, the employees' affinity to games in a corporate environment is even more crucial where the cultivated business culture and norms could bridle motivation to emerge in a gamified system [67] in the first place.

Also strong consideration has to be given in misusing gamification which could have detrimental effects in its result and also in the experience of the employees. Synthesizing the main points from literature [39], [67] we list and discuss the following challenges:

1. Legal issues
   A gamified application usually monitors the activities of the players and collects useful information for the needs of the game. Those personal data should conform with the actual data privacy policies.

2. Ethical issues
   A main critic on gamification is that it manipulates intrinsic motivation in a way that is not beneficial for the person which produces work. This is characterized by the term exploitationware [59]. To avoid ethical implications the gamified system should be adequately enjoyable for the employees.

3. Task quality
   Task quality output could be sacrificed when gamification is applied in a way that distracts from the main goal of the application.

4. Cheating the system
   Cheating should be considered to avoid exploiting the gamified application without producing valuable input.

5. Declining effects
   Declining effects refer to ephemeral interest of the employees in their interaction with a gamified system. This so called novelty effect [29] should be hindered by providing a continuous engagement loop with the right game rules and game mechanics.

2.4.3 Player types

As discussed in Section 2.4, the types of users or players for which gamification is designed heavily influence the expected outcome of gamification and the behavioral changes that this achieves. Incentivized from the design principles extracted from literature and presented in Section 2.4.2 above, in this section we look into literature to find out the distinct user types that affect how gamification is perceived by the users and how they positively react to its presence. What is more we want to know if there is

a mapping from user types to specific gamification mechanics that will inform us for game mechanics selection step.

A notable work for player types is conducted by Bartle in [5]. In this study four the main reasons mentioned which players typically enjoy in a game are, achievements within the game context, exploration of the game, socializing with others and imposition upon others. In a higher level of abstraction these factors are emerging as a combination of two dimensions of playing style. The first dimension spans across the notions of action and interaction and the second between the world and the players. The first refers to how much players are tending to be active and take actions within the game or how much they prefer interacting with it. The second dimension refers to whether the action/interaction is directed to the game world or the fellow players participating. In the four quadrants that these two dimensions form, if the first is placed vertically on the second, the four categories of players can be found with the following names and characteristics:

1. **Achievers**
   Achievers are types of players who are mostly concentrated in collecting points or getting rewarded for achievements that they accomplish during the game. Also viewing their progression is important by advancing levels.

2. **Explorers**
   Explorers tend to have an innate interest in unraveling the way the game functions and the potential hidden or in rare things that this might offer.

3. **Socializers**
   Socializers are inclined to establish interplayer relationships within the game. They seek interaction by chating, commenting or helping others.

4. **Killers**
   Killers are player types that seek dominance over other players through their actions. It is important for them to compete by hindering the progress of others.

In his study Bartle mentions that players might incorporate characteristics by all four types depending on it current state and gaming preference but he also suggests that a predominant preference to one of those four is existent in every player. Using the same player type taxonomy in [58] they suggest appropriate game mechanics and dynamics for each player type given the previously mantioned characteristics. This mapping is recreated here from this study in Table 2.1. Using the definition of game mechanics and game dynamics from 2.4 we see that in this table a mix of both is found. What is interesting, is that there is a significant overlap in many of them with the exception of the player type of socializers which is mostly interested in specific game dynamics such as community collaboration, virality and envy.

Similar efforts [46] propose six different categories of players which are related to Bartle's but provide a higher granularity that is inspired by another dimension which

	Achiever	Explorer	Socializer	Killer
Achievements	■	■		■
Points	■	■		■
Bonuses	■	■	■	■
Levels	■	■		■
Progression	■			■
Appointments	■	■	■	
Countdown	■	■		■
Leaderboard	■			■
Extinction	■	■	■	■
Community Collaboration	■	■	■	
Virality	■		■	■
Casc. Information	■	■	■	■
Envy	■	■	■	■
Loss Aversion	■	■		
Epic Meaning	■	■	■	■
Free Lunch	■	■	■	■

Table 2.1: User types and their preferred game mechanics (from [58]).

is based in Self Determination Theory [7]. The level of extrinsic and intrinsic motivation existent in each user creates another criterion from which subcategories emerge from Bartle's model. This mapping subdivision to further categories is illustrated in Table 2.2 below. For example Achievers from Bartle's model are expanded to Achievers and Player in the Hexad model. These two new categories both relate to players that are motivated by rewards and achieving accomplishments but Players focus on extrinsic rewards while Achievers focus on intrinsic rewards in this new model. Based on the Hexad model in [68] a survey instrument which was validated and checked for reliability was created that is able to identify different user types based on the model. What is more, a correlation analysis was conducted to map self-reported game element preferences to the user types found by the instrument.

Bartle's model	Hexad model
Achievers	Achieveres, Players
Socializers	Philanthropists, Socializers
Explorers	Free Spirits
Killers	Disruptors

Table 2.2: Correlation of Hexad model Bartle's model user types

2.4.4 Points and leaderboards

Points and Leaderboards are two of the most widely used game mechanics [81], which is also something that we confirmed in our literature study for gamification in crowdsourcing. From an abstract points of view, their main purpose is to provide feedback for the user activities that take place through its interaction with the gamified system. In [35] the importance of feedback mechanisms for the actions of users has been

highlighted, however the type of feedback differentiates among this dyad of game mechanics. In this section we discuss independently their functionality and also some useful points of consideration when applying them as extracted from previous studies.

The main goal of points is to provide a quantifiable feedback mechanism to the user. As stated in [48] they create "a clear connection between user effort and performance". The effectiveness of points as a game mechanic has been established empirically in several studies [48] [49]. They are implemented through a scoring mechanism that comes with a set of principles to assess user's performance based on the context and the goal of the gamified system. More specifically in crowdsourcing, scoring mechanisms are dependent on the desirable output in terms of quantity or throughput of tasks and their quality. For quantity, task contribution (i.e. the amount of tasks) is the main focus for awarding points while for quality it is mainly based on the level of agreement of a worker to the rest of the participants on the level of each task. In [35] a more involved scoring mechanism is proposed which is based in the expected information gain from each worker. In this, the probabilistic quality of each worker as a measure of quality and the level of contribution are consolidated in one measure to indicate the overall value of contribution of each worker.

Leaderboards work as a showcase of the progress of the user in a gamified system. They are inspired from multiplayer games [24] where scores of each player are entered in a leaderboard and the relative performance of a user to other participants is illustrated. As such, they work as powerful social motivators since the users receive feedback for their progress compared to others [47] and they foster competition [67] among them. Although several studies mention the motivational benefits that leaderboards yield [47] [51] they also mention the performance discrepancy that those might unravel and the demotivating results that this phenomenon can entail. Namely users that are at the top of the leaderboard or close to it are highly motivated to retain their lead or challenge themselves to reach it, while for users situated in the middle or lower ranks of the leaderboard the perceived formidable effort that is required to advance to the top, has a demotivating effect. What is more, the context in which competition is promoted might not be always suitable. For example in [67] they stress the fact that competition based gamification with leaderboards might not be suitable in a working environment. To this end several suggestions are found in the literature to diminish this effect. In [35] all time leaderboards are discouraged by the authors who instead suggest that weekly leaderboards should be adopted which keep track of the progress of the users within a weekly time span. In that way the high performance differences can be attenuated and the motivation to achieve one of the top places in the leaderboard can be reignited every week. For similar reasons in [39] they propose cross-situational leaderboards that allow performance comparisons only among players who have the same skill level. Finally in [67] they propose the shadowing effect which creates leaderboards where users compete with their own records.

2.5 Social gamification

Social gamification is a term for which a standard definition that would clarify which aspects of gamification it encompasses is not yet provided. However scattered indications of it, with similar names emerge in the literature that will help us collate its main characteristics.

According to Self Determination theory [57] intrinsic motivation of an individual is achieved through the feelings of autonomy, competence and relatedness. The main reasons for adopting social gamification are to lessen feelings of isolation and the lack of interactivity [38]. Therefore it includes those game mechanics and dynamics that foster the feeling of relatedness. The sense of relatedness can be created when an individual becomes part of a group or community, which usually operates under certain social norms, and where she is exposed to a certain social influence [28]. This social influence might further motivate her to act in accordance to the social norms of the group.

In [67], in their proposed taxonomy of game mechanics and dynamics they create a separate cluster called social influences, which includes those that encourage altruism, competition, gain of status and user high scores. Based on that, they identify two manifestations of social gamification. One that is based on competition and one that is based on collaboration. Additionally in [61] they mention that relatedness in a gamified context can be achieved ,among others, through tagging, rating and commenting which can be understood as social feedback which is part of the social influence and comes as a result of the compliance of the individual with the social norms of the community [28]. Another manifestation of social gamification is social proof [27], which is achieved by allowing individuals to showcase their own achievements or progress or empowering them to inspect those of others. Unifying from literature the instances in which social gamification is encountered we can identify and discuss three main manifestations of it:

1. Social actions
   They refer to actions that a person can take in a gamified system to establish relationships and support the motivational driver of connection [39]. As described before those actions can be of the nature of commenting, discussing, rating or showcasing the achievements . Those actions are not necessarily tight with the gamified logic and rules of the system but can work as strong motivators. Although, to our current knowledge, this aspect of social gamification is underresearched in crowdsourcing, the are several indications of its effectiveness and high usefulness to gamification. In [10], from interviews they found out that a strong motivation to participate in a gamified application for citizen science is the ability to socialize, get in touch with friends and be part of the community. Same findings were reported in the context of an enterprise where games were created by the employees for the employees to share knowledge and for leisure [26]. Also in [20] they used Activities Widget in a gamified crowdsourcing context which allowed workers to see the progress and contributions of others and

proved its effective as a furtherance incentive mechanism for task contribution.

2. Competition
   Competition is an aspect of social gamification that promotes actions which faciliatte comparison of individual's achievements and progress with that of their peer. In that way it fulfills the urge for social status, reputation and fame [67]. A widely used game mechanic that falls within the competitive nature of social gamification is the leaderboard. Leaderboards encourage users to take specific actions that will allow them to advance in the current ranking [43]. The sense of competitiveness and the motivation that this cultivates have also been reported in studies where leaderboards were applied [47]. Another instance of competitive social gamification was showed in studies in the context of e-learning environments [38]. In this study competition was implemented by matching up students to compete in knowledge quizzes, that yielded points, for lectures of the course and also proved the significant effects of it in student retention.

3. Collaboration
   The collaborative nature of social gamification is applied through mechanics that advance cooperative actions between peers in order to achieve a collective goal. To this end it stimulates synergy effects and increased motivation through altruism [67]. The effectiveness of collaboration has been proven in the studies of GWAPs described in Section 2.4.1 where players were paired to produce common tags and annotations for tasks. Furthermore in [69] team-based annotations for linguistic resources were found to increase the user activity significantly.

2.6 Chapter conclusion

In this chapter we focused on studying relevant literature in the intersection of gamification and crowdsourcing within the enterprise and also studied metrics that are related to engagement in crowdsourcing and also quality. Our intention was to first understand how crowdsourcing is applied in an enterprise environment and elicit main challenges and typical applications of it as well as the dominant incentives for participation. Second, we were interested in studying metrics used in the literature for engagement and quality in crowdsourcing that will help us address the main research question of this study. Then we moved on to study gamification. In this part we discussed the main aspects of gamification, how it is applied in an enterprise and how it is combined with crowdsourcing. We also underlined the importance of the context in which gamification is applied and studied literature for user types and their main motivations in participating in gamification. Finally in the last two sections we first discussed two of the most commonly used game mechanics which are points and leaderboards and then defined social gamification and highlighted its main characteristics and aspects.

Chapter 3

Research methodology

In this chapter we present the design of our research methodology which has been adopted and used as a roadmap to address the motivation and the main research question of this study.

In the following we first provide an overview of our research approach and the different parts that this consists of, in order to fulfill the objectives described in Section 1.2. We continue with the design and analysis of the qualitative part of our research. We motivate and describe the selection of the experimental design which was used and which also includes the operationalization of the dependent and independent variables of the experimental part of the study. Finally we delineate the research protocol which we followed to implement the crowdsourcing experiment in the two enterprise environments.

3.1 Overview

Our research focuses on the probable causal relationship between different gamification elements and the quality and user engagement in enterprise crowdsourcing. We are also interested in how different enterprise environments mediate the effect.

We first conducted a literature review in the conjunction of enterprise crowdsourcing and gamification. This helped us identify gaps in current research regarding the experimentation of gamification and motivated us to explore the social incentives that underpin the behavioral changes that gamification fosters. Through our literature study we also became aware of best practices in applying gamification in an enterprise environment which mainly involve gaining a better understanding of the player types that exist. What is more, we understood the potential confounding variables that might exist such as crowd incentives, task types and task complexity which helped us control for them by design in order to isolate the interaction effect between gamification and user engagement and quality. The insights drawn from our acclimatization with the field of our study, actuated us to separate our research methodology into two parts: an exploratory and an explanatory.

For the exploratory part, we deploy qualitative research procedures to gain an understanding of the player types that exist in Rabobank and also discover a possible enterprise application for which crowdsourcing could be used. Along with the player types in Rabobank, we complement our findings with previous research done within IBM [64], to motivate and validate the selection of the gamification elements which will be used for our experimental study. We also use the exploratory part of our study to inform the selection of an enterprise crowdsourcing task, which adheres to common crowd incentives in this context, and also falls within typical applications used in enterprise environment as explained in Section 2.1.

Our explanatory part consists of realizing our experimental design, which is discussed in detail in Section 3.3, and deploying our experimental tool to answer the main research question posed for this study. We operationalize the concepts of engagement and worker quality with measurements that have been extracted from previous studies in the field. The analysis and discussion of our results between the selected experimental conditions concludes the explanatory part of this research.

3.2 Qualitative research design

The qualitative research part of the study focuses in the exploration of the player types within Rabobank and also the identification of an enterprise crowdsourcing application that would be used as part of our empirical study. This involved semi-structured interviews with employees of Rabobank and also with experts in the Food and Agriculture department of the company. In the following two sections we describe the design of the interview guides that were used as well as the analysis of their results.

3.2.1 Employee player types interviews

As already discussed in 2.4.3, in Bartle's player type taxonomy there are four main categories of players [5]. Those consist of the achievers, the socializers, the killers and the explorers. Based on this taxonomy, but with a further segregation of Bartle's categories, in [68] they propose a validated instrument with items to study the player types. In these interviews we follow the taxonomy of Bartle and we base the formulation of the relevant questions on the items proposed in [68] using those that where proved as significantly correlated to each player type. In order to analyze the player type of a person, information about his qualities and personality characteristics are required. In our case those information should be extracted within the context of a company and as such, the questions posed are related to the working qualities of the employees.

To this end, we use the semi-standardized interview type. According to [22] the purpose of this category of interviews is to collect subjective theories of the people about a specific topic. In our case the subjective theories reflect the opinion of the employees about their personal characteristics and also those of their peers. We also select this type of interview because it will enable us to extract information from the interviewee that is immediate and based on their knowledge and will also allow us to address theory-driven questions elicited from relevant literature.

Based on those insights we segmented our interviews into 5 core topics with the addition of a final one which is introduced in order to obtain information on the preferred time and duration that employees could participate in crowdsourcing. The first four main topics contain questions that are related to discovering characteristics of the employees that pertain to one of the four player types suggested by Bartle's study. The fifth category contains questions to discover the predominant player type of the employee and also generalize to the Rabobank population. For the interested reader the interview guide can be found in Appendix B.1.

Of utmost importance for the reliability of our research was the case sampling method that would be used. This refers to the sample of persons that would be selected to interview. Since the context in which the research takes place is a working environment with thousands of employees and many different departments an exhaustive sampling method could not be easily applied. For the purposes of our research we promoted a priori determination of the sample structure based on certain prespecified criteria [22]. Those criteria were the gender and age of the interviewee and also the business unit in which he or she belongs. To this end we opted for 7 interviews with a variety of demographic information and also job positions. Those details are listed in Table B.1. In the following we present our findings and observations from those interviews per topic discussed.

Socializer player type

In the employees' responses regarding the importance of interaction with their social environment and how much they are motivated from it, we identified two different dimensions in which is this is usually understood and also a predominant way of expressing it.

The first is the working environment itself, in which respondents unanimously expressed that the ability to draw inspiration from coworkers and also develop a network within the company is of paramount importance for their ability to perform their work duties. For example one employee noted, *"for me it is very important that I can share or that I can retrieve knowledge from my colleagues, without this I do not think I can do my job very well..."*, while in a similar manner another said, *"I could not do anything at all. Because you need people in order to get requirements and discuss the risk factors..."*. In this context, motivation in interacting with their working environment is mostly elicited from the unavoidable connection of it to their working performance. It is also an indicator, since our sample is diverse, that the employer promotes a non soloistic principle of work that primarily involves having regular iterations of their ideas and solutions with at least their immediate environment.

The second dimension is that of their personal, innate desire to be involved in a working environment that exhibits proper social characteristics which is also reflecting better their relation to the socializer player type we are interested in. Again almost all employees expressed their personal desire to be in a working environment in which their feeling of relatedness is satisfied as well as it provides the opportunity to create social connections. A particular employee said, *"I have been doing a lot of work almost*

semisoloistically and this is not good for your motivation, you need to see faces." while another also noted, *"I don't like to do it on my own, because I like to discuss with you for example, to discuss my ideas and my point of view"*. Only one employee expressed an indifferent opinion on whether his work is conducted in a social or not environment by stating, *"Honestly I would not care. Because I am really good in working In both environments."*.

We were also interested to find out from the answers we collected, under which prism is social interaction expressed and preferred and especially whether it will be a collaborative or a competitive one. All answers from the employees focused on collaborative characteristics rather than competitive, signifying that the former is much more valued in a working environment compared to the latter. One employee, when specifically asked to comment on what type of social interaction is preferred he clearly stated, *"for me personally is very much collaboration"*. Another employee also expressed the need for collaboration which springs from having proper guidance from colleagues, *"But you also need some guidance. You cannot focus for 8 hours straight, you have to have some distraction to generate some new ideas."*, while most of the employees opted for collaboration as a necessary mean to collectively generate new ideas and work towards a target : *"if we are together we can get along and discuss the problems that everybody has and we can come up with solutions."*

Achiever player 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 vast majority of the employees recognized rewards as a main motivator for their work and also something that should be tied to their performance except for one employee who underlined the importance of intrinsic motivation by saying, *"I think you have to do something from within yourself and I do need an extra salary or kudos or what their name is"*. Among those who are reward oriented, there is also a separation on the preferred way that those rewards reflect their competence. Some employees think of rewards as a way of gratification that resonates to their personal desire to obtain positive feedback for their work and to reassure to themselves a good working performance. For example an employee stated, *" I think if you work hard and you do your work very well and you get rewarded that is an acknowledgement of your good work, and then the next year you want to work even more and even better"*, while another said, *"you need short samples of feedback [meant as rewards], being able to address things that happened last week or the week before while they are still fresh and you can talk about that and the things that you really do good you make bigger."*. On the other hand there are employees who position the rewards in the social context as a requirement to exemplify hard working employees from their working milieu. In this direction one employee states that, *"we don't reward people that do exemplary good work and we do not tell people who do mediocre work to step up. So whatever you do*

it is ok and we need to change that." while another complements, *"So today it does not make any difference any more how do I perform in my job, I still get the same salary and that is strange. Because if you see the people from the other side of the table, sales people who I have to work with every day, they do get paid by results or revenue or whatever."*.

Regarding the types of rewards, the employees' main preference resides on rewards that adhere to their intrinsic motivations. To this end an employee focused on the importance of being valued by your colleagues, *"For me the most important rewards are...and I like it and I feel good when they say this a good relation I think you contribute to the subject we are busy on and you are doing your job good."*. In another response an employee stressed the importance of relatedness through positive feedback and reanimation, *"...I also believe in a very positive way to motivate people, to keep their energy and keep them enthusiastic and passionate about what they are doing."*. Only one employee stated his clear preference in extrinsic rewards (e.g. monetary rewards) while another focused on the importance of balancing between types of rewards to keep the employees satisfied.

Killer player type

In order to unravel characteristics of the employees that might be related to the killer player category we asked them to comment to what extend they are finding themselves challenging their standard way of working. This is because that is one of the core qualities that characterizes players in this category, who find themselves intrigued by questioning systemic aspects of their environment. Most employees suggested that following a standardized way of working is in general preferred. For example one employee stated, *"Yes, I think it is important but in my role we need more policies and standard way of working"*, while another added *"I think it is good and practical for most people that you do need some guidelines. But I only use those guidelines if they are necessary"*. However most of the employees 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 incentivizes them to act in this specific way. For example an employee who was questioned whether he finds it personally interesting to act in a disruptive and free form way in his work responded, *"For me it would be breaking the loop. It depends on whether my analysis of the future situation and where we are heading is congruent with the direction that is being given."* while in the same direction another employee added, *"...if I don't confirm to the process and I choose another way, does it add as much value to the customer as if I were to work by the process? I try to work as much as possible within the processes but I try to look always through my lean glasses to see is it possible to skip one step or do something different so the time will be shortened or in general if I can skip a few steps."*. Only one employee stated in a more clear way that this springs as an inner motivation by saying, *" It is*

outside of my profession, but I do this where I can, I challenge the processes and I ask people what do they think and challenge them for an improvement.".

Explorer player type

To determine how much of an explorer player type are the employees, we focused in 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. There were employees who prefer to have a broader scope, *"Of course I have a path and I have to do my work, but I do not thing that I am the regular Rabo type. I like doing a little bit of this and then a little bit of that and I am all over the place."* and also those who tend to follow a specific working path, *"I would say that I am focusing on my job and maybe on the side I do some other stuff, but It is not that I am finding my own way, I just do my job"*. Finally there were employees who prefer to have a balance between those two. For example one employee stated, *"I think a combination. Many times I like doing things on my own but most times I like doing them as part of a team. Because I think if you do things as a member of a team you can grow."*.

Summary of observations

We should proceed with caution while interpreting our results from the interviews on player types within Rabobank. This is because, it is not necessary that responses on questions regarding working behavior can safely give strong evidence in which player category an employee is. However from our qualitative results discussed earlier, there is an indication, which 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. Indeed 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.

To this end in the concluding topic of our discussions with the employees we focused in reviewing the different categories and asking them to select which of them best suits their personality. To make this easier we allowed them to rate in a scale from 1 to 10 how congruent they find themselves with this category. The responses are listed in Table 3.1. Looking at the category with the highest score for all employees we see that there is a preference in the Socializer player type and also the Killer type. Interestingly enough when the employees were asked to provide their opinion on which of those categories best suits the general Rabobank population most of the answers indicated the Socializer type. It is also interesting to denote that the responses for the categories of Socializer and Achiever had greater consensus compared to that of the Killer and the Explorer. Almost all respondents recognized qualities found in the first two categories in them while there was some dispersion in the answers we collected for the later two.

Those observations provide credibility to the hypothesis that the dominant player types in Rabobank are the Socializers and Achievers. This is also in accordance with previous studies regarding gamification conducted at IBM [64]. In this study leaderboards and badges, which are the main preferences for Achievers, were used to study engagement and social interaction (i.e. inviting others to the game and sharing news) in an enterprise quiz game. When combined it was proven that they are significantly more engaging and also lead to higher social interaction. We capitalize on our findings here to select the proper gamification elements for the experimental part of the study as discussed in Section 3.3.

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

Table 3.1: Employees' ratings on the level to which they match their personal characteristics to that of the 4 player types (on a scale from 1 to 10) and also their selection for the general population of Rabobank.

3.2.2 Expert interviews

A main output of our literature review on enterprise crowdsourcing are the dimensions that differentiate it from online crowdsourcing. Among others, as stated in Section 2.1, those are the specific incentives of the enterprise crowd and also the types of applications that this is used for. Also, more often than not, those two are intermingled, meaning that the proper selection of tasks for enterprise crowdsourcing also affects the incentives of the crowd regarding participation and also engagement. In order to tie our research to previous studies in the field it is important that we identify a relevant task for enterprise crowdsourcing.

To this end, we identified a use case in the Food and Agriculture Research (FAR) department of Rabobank, to train a machine learning model reinforced with human generated data provided from enterprise crowdsourcing. In order to gain an understanding on how a machine learning approach could help experts working in the FAR department in their everyday work, and also reformulate it as a computer science problem, we deployed qualitative research methods. Namely based on the taxonomy that is proposed in [8], the interviews fall into the systematizing expert interview category. Based on the guidelines provided in [22] regarding expert interviewing we used the interview guide, which is found in AppendixB.2. To attain our goal, we focused in addressing the following questions through the expert interviews:

Q1: How do the experts conduct their research and produce Food and Agriculture Research reports?

Q2: Which are the data sources they use in their work.

Q3: What are some possible aspects of their work which could be automated by a machine learning model?

In the following we focus on the part of our analysis which is related with question Q3 and how this analysis informed the requirements for the enterprise task design.

Enterprise task requirements

A main discussion topic in the experts' interview was to identify work related, repetitive and time consuming tasks that can be automated and trained with crowdsourcing. From our analysis we identified three potential tasks which we separated across three different dimensions in order to describe them. That is their core function, the data sources that those use and also the ideal output that is expected from the experts. Those information are gathered in Table 3.2 below.

Problem Description	Function	Data sources	Output
1. Relevance assessment of news sources	Categorization of relevant and non relevant sources given a specific information need	Factiva, newsletters	A ranked output of documents based on the information need
2. Extraction of market information	Extraction of key companies operating in a specific domain and relevant company specific information	Company websites, annual reports, conference websites	Company names within a market and associated information such as no. of employees, locations, CEO name, volume etc.
3. Data validation	Validation of data in relational databases in order to exclude erroneous data entries	Department's internal database	A validated view of the database

Table 3.2: Description of candidate tasks extracted from expert interview analysis

From the tasks listed above, we opted for the one related to extracting market information from data sources. 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. From a computer science point of view we formulated the problem, as an information extraction one in the agriculture domain. The objective of the machine learning algorithm is to extract possible relations found between entities in unstructured online text data.

To narrow the scope of the problem we focused on 3 possible relations on which the algorithm should be trained based on the requirements posed by the experts for this problem. The first relation is that of identifying a CEO of a company in the text, the second is related to extracting the products produced by a company and the third one is about affiliation relation between companies (e.g. subsidiary company or acquisitions).

In similar studies where crowdsourcing has been used for information extraction [37] tasks have been used to attenuate noisy training examples that were generated from distant supervision methods [50]. In this way the crowdsourcing task was transformed in a validation task of knowledge triples in the form (e1,r,e2) where e1 and e2 are entities automatically extracted and r is the hypothetical relation between them for a given sentence. To the best of our knowledge, no existing knowledge base contains entities such as companies, products and CEOs specifically for the agriculture domain, and thus our task has to incorporate two functions. The one is entity extraction and the second is relation extraction. To this end, the main requirement for our task is to extract knowledge triples where all involved parts of the triple are requested to be annotated from the crowd. The data sources used for the generation of the tasks, which is further discussed in 4.7, are in accordance with the data sources used by the experts and are revolving around online news sources in the agriculture domain.

3.3 Experimental design

Our experimental design was motivated from gaps in the experimentation of gamification, identified in our literature study and also the results of our exploratory qualitative research on player types inside the enterprise. The latter validated the strong influence of social incentives inside the enterprise (i.e. Socializers and Achievers). Those, however, can be expressed in a collaborative but also in a competitive form and there is need to understand how gamification elements that foster them, affect crowdsourcing activities. This inquiry is further inspired from our literature study findings in which we recognized that previous studies were focusing mainly on the experimentation of limited gamification elements (e.g. points,leaderboards,badges), while there is not still a clear understanding of the how social characteristics of gamification relate to positive patterns of service use. Especially collaborative techniques have been noted as something that calls for further research [51]. Finally it is also still unclear how contextual factors such as the corporate environment influences the effects of gamification in enterprise crowdsourcing.

To this end, the experimental design of our study consists of two experiments which are designed as true experimental between subject post-test only four treatment comparison. The four treatment conditions are formed based on different gamification elements which are incorporated as seen in Table 3.3 and further elaborated in Section 3.3.1. We deploy our experimental tool as a mobile crowdsourcing application with the appropriate gamification elements per experimental condition and measure the engagement and quality based on application usage information that are logged for each participant. We replicate our experiment using the same conditions into two enterprise environments, that of IBM and Rabobank, to facilitate the comparative study of the effects of gamification in different corporate contexts.

3.3.1 Experimental conditions

To form our different experimental conditions we selected gamification elements which mostly adhere to the Achievers and Socializers player types as listed in Table 2.1 while they are also relevant to each other so that meaningful gamification groups are created.

Enterprise Context	IBM/Rabobank			
Gamification Elements	Control Group	Group 1	Group 2	Group 3
Score + Progress Bar	X	X	X	X
Leaderboard	-	X	-	X
Social Gamification (Collaborative)	-	-	X	X

Table 3.3: The experimental groups used in the experiments, based on the gamification elements deployed

Namely, we have chosen points, progression, leaderboard and community collaboration. We elaborate further in each experimental group below:

1. **Control Group**
   The control group contains basic feedback gamification mechanics such as score and a progress bar which are deemed essential as we have seen in our literature study. The scoring mechanism is based on the contributions and the quality of the worker while the progress bar provides a visual representation of the amount of tasks completed from the total available.

2. **Group 1**
   Group 1 will, additionally to the control group, contain leaderboards. As discussed in our literature study, leaderboards express the competitive nature of social gamification, so in this group competitive dynamics are promoted.

3. **Group 2**
   In Group 2 a collaborative social gamification treatment is applied. In order to cultivate synergy effects between the participants within the application we draw inspiration from the design principles of GWAPs. To this end, in this group the submission of tasks has two options. The one is to individually submit the task for one's own benefit (i.e. increase score and progress) and a second option is incentivizing the user to solve the task collaboratively with a peer of his selection. In that way the user will be able to submit a task and also assign it to another participant to annotate it (this takes place asynchronously). In this way collaboration dynamics between the users are fostered as an aspect of social gamification and we are studying potential diffusion of crowdsourcing activities via community nudging.

4. **Group 3**
   Group 3 incorporates all the previously mentioned game mechanics in one group.With the use of this group the interaction effect of competitive and collaborative social gamification is studied.

3.3.2 Legal and Privacy Aspects

An important consideration while designing our experimental methodology, was the compliance to privacy policies existent in the enterprise environment. In our case,

since our experiments took place in two different companies, he had to adhere to the strictest subset of privacy rules among the two companies in order to have identical experiments. In this Section we describe how this affected our experimental methodology and introduced limitations to our work.

Since we are focusing in the study of collaborative and competitive game mechanics in our work, which are mainly based on social dynamics, the presence or real personal information about the employees participating in it, is of paramount importance. This is because feelings of relatedness, community acknowledgment, synergy and competitiveness are intuitively strengthened by the ability of an employee to relate an account to one of his peers while using our experimental tool.

However, due to the fact that employees' personal information are sensitive and confidential and also because of the logging functionalities that are necessary to be present in our experimental tool to obtain usage metrics, the storage and use of personal and application usage information is not is accordance with enterprise privacy policies. To this end, our experimental methodology is limited to the usage of anonymous users participating in each experimental group as described in the previous Section.

This design choice unavoidably introduces limitations to our work since underlying factors that spark the game dynamics which are the focus of this study are omitted. This is evident in Group 1 and Group 3 where we are restricted to use leaderboards with not realistic user names and in Group 2 and Group 3 where the task sharing functionality has to be based again on the same user names. The repercussions to the implementation of the experimental tool due to the legal and privacy issues are discussed in more depth in Section 4.3.

3.3.3 Operationalization of engagement and worker quality

Based on our main research question, we operationalize reliability and engagement in enterprise crowdsourcing to quantitative metrics previously found in related studies, as discussed in Sections 2.2 and 2.3. Those metrics are collected and logged through the interaction of the users with the mobile crowdsourcing application. An overview of the logging information is presented in Table 3.4 below.

Engagement metrics

Engagement metrics are used to evaluate the level of interaction of the users with the application across the different experimental conditions. We used the variables logged from Table 3.4 to define the following metrics of interest.

1. **Number of task executions**
   We measure the average number of tasks that have been contributed by a worker, normalizing for the observation interval during the experiment duration, which is discussed in Section 3.4. We use this metric to compare the throughput that can be achieved for the chosen experimental conditions.

Variables logged	Description
Sign up time	A timestamp of the first time a user logs in to the application
App start time	A timestamp of when the user starts the application (i.e. app was previously closed)
App resume time	A timestamp of when the user resumes the application (i.e. app was previously paused)
App close time	A timestamp when the user closes the application (i.e. app removed from the task manager or user signs out)
App pause time	A timestamp when the user pauses the application (i.e. application is put in the background)
No. of task executions	The number of submitted tasks from a user
No. of shared tasks	The number of tasks that a user shared with his peers (i.e. for Groups 2 and 3)
No. of response tasks	The number of shared tasks a user responded to (i.e. for Groups 2 and 3)
No. of sessions	Number of times a user started or resumed the application
Session time	Time elapsed between a start/resume event and a close/pause event
Task dwell time	Time elapsed during task execution

Table 3.4: Logging information collected with the experimental tool

2. **Number of sessions**
   The number of session refer to the amount of times a user opened and interacted with the app during the observation period. A session start is determined when one of the application start time and application resume events is stored.

3. **Session time**
   We calculate the time spend interacting with the application for each user. This is signified from 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 a user has had within a normalized time span. This metric reflects the level of engagement of an individual user and an experimental group by the time spent on average using the application.

4. **Task dwell time**
   Task dwell time reflects the amount of time elapsed since a user selects a task until the user submits or selects to collaborate a task. We average the dwell time across all task executions contributed by the user. Higher average dwell time signifies higher engagement during task execution. Since collaboration is implemented by selecting a user to share the task during task execution (implementation is discussed in Section 4.6.4),which might affect metric across different experimental conditions, we take care to denote the end of task execution at the time a user presses the submit or collaboration button.

Worker quality metrics

To measure worker quality and also due to the absence of golden standard labels for the tasks used in our experiments, we rely in agreement metrics. Depending on the input requested per task category we use different quality metrics. We elaborate below on the quality metrics used, while details on the task categories used are provided in Section 4.7.

- **Plurality answer agreement**
  We measure worker quality with plurality answer agreement, similar to [71], for tasks with numerical input. The following formula is used to calculate it:

$$S_p(w) = \frac{f}{F} \tag{3.1}$$

  where w represents a worker, F is the total amount of tasks the worker has provided annotations and f is the number of the tasks for which worker's annotations are in accordance with those produced from majority vote. This metric is strict as it assumes majority vote as the golden standard on which the worker's annotations are directly assessed.

- **Average worker-worker agreement**
  For task in which the user is requested to annotate relations found in text between predefined relation categories we use the average worker-worker agreement which is proposed in similar study [3] for relation extraction tasks. We use the following formula to calculate it:

$$avg\_wwa(w_i) = \frac{\sum_{i \neq j} \left|S_{i,j}\right| * wwa(w_i, w_j)}{\sum_{i \neq j} \left|S_{i,j}\right|} \tag{3.2}$$

  where w_k denotes the worker k and $S_{i,j}$ is the set of common task annotated by both workers. Also $wwa(w_i, w_j$ is the pairwise worker-worker agreement for all the tasks s which annotated in common, and it is given by the formula below:

$$wwa(w_i, w_j) = \frac{\sum_{s \in S_{i,j}} RelationsInCommon(w_i, w_j, s)}{\sum_{s \in S_{i,j}} NumAnnotations(w_i, s)} \tag{3.3}$$

  in which $RelationsInCommon(w_i, w_j, s)$ is the number of annotated relations that are in common between two workers in a specific task s and $NumAnnotations(w_i, s)$ is the total annotations produced by worker w_i for the same task

3.4 Experimental Protocol

In this section we describe the procedures which were followed to execute our experimental design. These involve the pilot phase that has been provisioned to test the experimental instrument, the methods for recruitment of participants and also the selected observational period in which we gathered data for our analysis.

3.4.1 Pilot

To properly test our experimental instrument we scheduled for 2 pilot phases, one within IBM and one within Rabobank. Our main objectives were to collect feedback on the functionality of the application and the user experience but also to cold start gamification elements such as leaderboards and scores. The latter was deemed important so that users in those experimental groups are not demotivated by the lack of participants once they enroll in the experiment. Namely we invited 14 employees from IBM and also 10 from Rabobank during a period of 10 days. After this period we fixed bugs that have been identified and also resolved some important design issues.

3.4.2 Observation Interval

We deployed our experimental tool in the two enterprises and allowed for an experimentation period of 2 months in total. Since participants were able to join the experiment in any time within this 2 month time frame we normalize our observation interval to 1 month maximum for our analysis. This is signified by the information we log when the user logs in to the application for the first time. Namely our experiment in IBM lasted from the 17th of May until 17th of July while the experiment in Rabobank was held between 15th of May and 15th of July.

3.4.3 Recruitment Procedures

For recruitment purposes, we took care to follow the same procedures inside the two enterprises, as means to communicate our experiment. We disseminated a few hundred of flyers within the two companies and also placed posters in key locations. We have also given a few presentations on enterprise crowdsourcing and our research in particular. Finally we used corporate mailing lists to send e-mails in groups of employees inviting them to participate and also advertised our project in corporate blogs and news sites.

3.5 Chapter Conclusion

In this chapter we elaborated on our experimental methodology and how we designed our research into an exploratory and an explanatory part. The former is used to gain an understanding of the different player types existent in the enterprise and also inform us of a suitable enterprise application which could be used for crowdsourcing.

In addition with our literature study this allowed us to carefully select the independent variables of this study which focus on the study of collaborative and competitive gamification elements for crowdsourcing. Based on this we delineated our experimental design and presented the different experimental which are used in our experiments. We further operationalized engagement and worker quality which are the two main concepts of interest for our study.

Finally we discussed the experimental protocol which we followed to execute our experiments inside the two companies and we addressed possible threats in validity and

the actions we have taken to address them. In the following chapter we delve into the implementation details of our experimental instrument.

Chapter 4

RaboCrowd/IBMCrowd: The Experimental Instrument

For the aspirations of the current research, significant work has been put in the development of the experimental tool that would be used to enable enterprise crowdsourcing in combination with gamification. In order to achieve this, we developed and used the generic ECrowd platform to instantiate two mobile enterprise crowdsourcing applications, called RaboCrowd and IBMCrowd, with the intention to be used in two different enterprise environments.

IBMCrowd and RaboCrowd are two identical mobile applications with only small necessary differences to discern them, since they are deployed in two different companies. Those differences are only in the name of the applications and some legal and contact details that are included. From the research point of view, they are the same experimental tool, so in the discussion to follow when we refer to either IBMCrowd or RaboCrowd we are referring to the same mobile application and thus we may use those names interchangeably.

In the rest of this chapter we first provide an overview of ECrowd and its implementation details. We move on to explain the main requirements that were drawn to achieve the current research goals and explain the architecture and data model of the application. We finally discuss how we implemented and incorporated gamification in the mobile applications and also illustrate the tasks that we generated and included in our crowdsourcing experiment.

4.1 Overview

ECrowd is a platform in which mobile enterprise crowdsourcing applications can be instantiated. Its back end is built with Spring Framework to support a website front end for administrative purposes. Through its web interface administrative actions such as the generation of a new application, creation of crowdsourcing tasks and their templates and resources, definition of the look and feel of an application, addition of new users and also control of the annotations generated from various applications and users.

It is designed to support the creation of custom mobile crowdsourcing applications by allowing an administrator to have control over most of the aspects that compose a mobile application.

The modular capabilities of the ECrowd platform are provided to the front end through a RESTful api, to which requests can be made to load all the necessary modules for the mobile application. The mobile front end is using the Cordova framework [1] which allows for the usage of standard web technologies (i.e. HTML,CSS,JavaScript) and works as a container that interfaces between them and the native device capabilities. The mobile application also makes use of AngularJS [2] framework that allows for the creation of domain specific language for annotations and also the Ionic framework [3] to support the native look and feel of mobile applications which are developed in Cordova.

From the user point of view the application can be used after authenticating (Figure 4.1) and can be navigated via a main menu list (Figure 4.3). Depending on the experimental condition in which the user is added, the welcome screen (Figure 4.2) provides him basic information about the application and also its functionalities. From the main menu the user can navigate to a view where he can select one of the task categories available and work on a specific task. Gamification elements are added to separate menu items available. The users can access their progress menu to check their current score and also their progress in circular progress bars. They can also navigate to the leaderboard to see their ranking among their peers and also check from the shared tasks menu the tasks that have been shared by their colleagues with them. Finally the application includes a menu item that provides more detailed information about the application, legal notes such as a privacy statement and also a feedback and usability test survey in case the users want to provide feedback for the application.

4.2 Requirements

To successfully study the potential effects of gamification in employee engagement in enterprise crowdsourcing, it is highly important that the effects of extraneous factors are diminished. This means that the application should provide a seamless user experience, that allows for easy task execution and an interface that incorporates gamification in a way that is apparent and understandable but not obstructive to the main goal which is task execution. More importantly the deployment of the application in an enterprise environment imposes some extra requirements that are related to the specific needs of the crowd. Those are, the limited amount of time of the employees to interact with the application and an easy onboarding process. Finally it should be designed to operate as an experimental tool to fulfill the requirements of our experimental design. Based on these remarks, we identified the following main requirements for the mobile application:

[1] https://cordova.apache.org/
[2] https://angularjs.org/
[3] https://ionicframework.com/

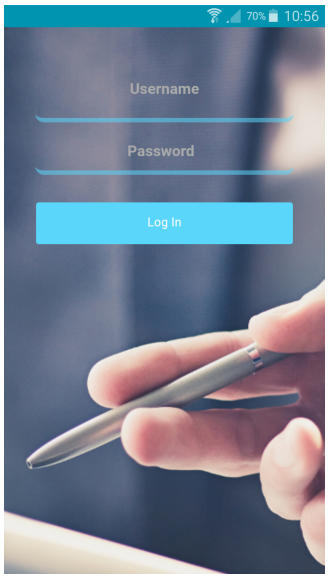

Figure 4.1: IBMCrowd login view

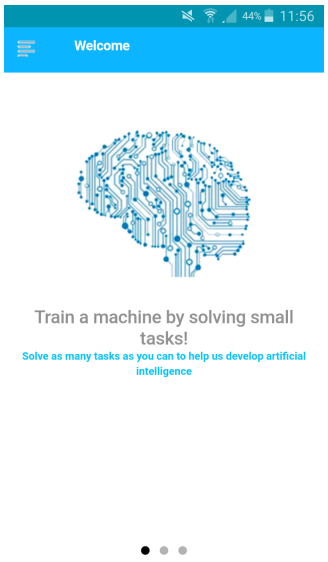

Figure 4.2: IBMCrowd welcome view

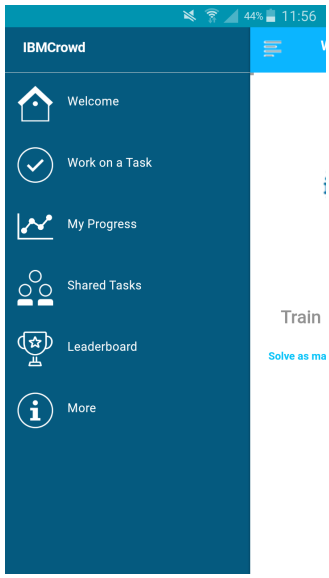

Figure 4.3: IBMCrowd main menu

- It should be able to fairly distribute users across different experimental conditions
- It should be able to load different gamification elements for the various experimental conditions
- It should include tasks that adhere to the crowd's incentives in enterprise crowdsourcing and are easy to understand and execute
- Its design should strike a balance between an enterprise application and a gameful experience
- It should be able to allow employees to share tasks to each other, earn points for contributing tasks and also review their progress by checking their overall contributions as well as their current score
- It should be able to log application usage information to support the measurements of our dependent variables
- It should be responsive and easily used

Most of these requirements were covered from carefully setting up an initial plan for the development of the application and adhering to it, to minimize the chance of irregular behaviors and bad design choices. However the final experimental tool that was deployed, has incorporated the feedback and suggestions from the two pilot phases.

4.3 Legal and privacy compliance

Before moving on to explain in detail the implementation of the experimental tool it is important to mention how the application was built in order to be compliant to legal and privacy regulations within IBM and Rabobank.

Since the experimental tool is designed to collect data related to the activity of the users when using the application, it was of paramount importance to comply with the data privacy regulations of IBM and Rabobank. This means that the application was restricted to not have access to any kind of personal information or logging information that could trace back to the actual employee using the application. Although we implemented a LinkedIn login functionality for the app, in order to have a more personalized user experience and also facilitate the process of signing up we had to change it and allow users to login in with predefined credentials randomly generated by us and provided upon request. We also included a legal note in the form of a privacy statement in the application which for the interested reader, it can be found in Appendix A.

The repercussions of the legal issues to our research can be summarized into two points. First the user sign up process was impeded since it could not be integrated in the application. This choice of signing up was necessary, to prevent users from selecting a user name that could be their true name or anything closely related to it. However this provided us with an opportunity to have full control of the assignment of the users to the experimental conditions, as well as the option to collect demographic and job related information easily. The second consideration is that since our experimental conditions are based on social incentives, the inability of a user to relate an application user account to a colleague might have effects on the results of our research, which is something we also stressed in Section 3.3.2.

4.4 Architecture

The architectural style used by the application is that of the Model, View, Controller (MVC) both in the back end as well as the front end. The user makes api calls, to the controllers which are responsible to access the data layer through the model and inform the view component which in turn responds to the user with the corresponding view of the application in json format as illustrated schematically in Figure 4.4. In that way the user can access the appropriate menu items, per application and experimental group, which contain the gamification elements and also fetch tasks for execution.

The view component is using a template engine based on Apache Freemarker [4] to compile the templates of the views and the tasks with the appropriate resources. Each view and task visible in the application, is composed from a template that defines the layout and the logic and specific resources which are available to the template. In this way we were able to construct multiple tasks and views of the same category with different resources, such as images, text etc.

[4] `http://freemarker.org/`

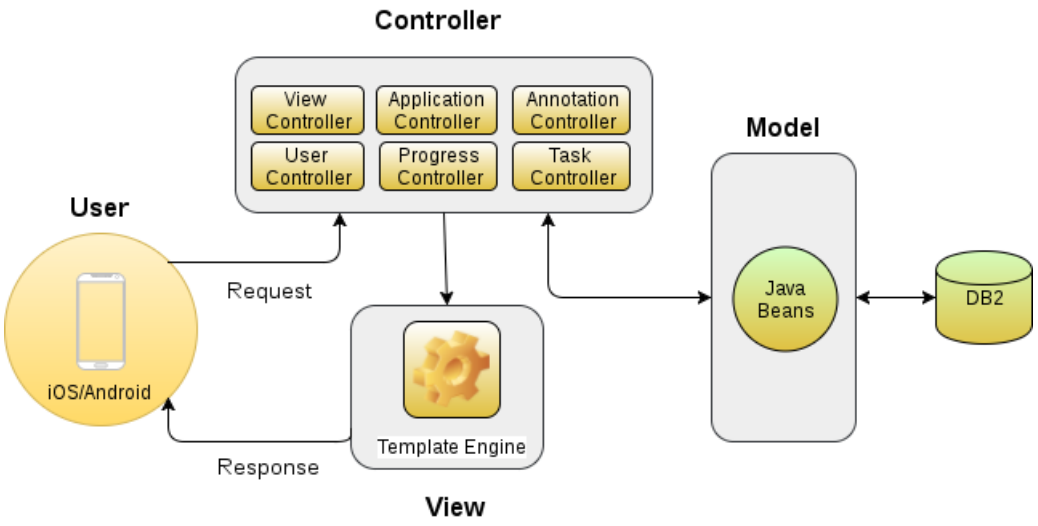

Figure 4.4: Application Architecture

In the front end AngularJS allows for the adoption of the same architectural pattern (i.e. MVC) with the added flexibility of defining an annotation domain specific language that is used to collect information from the user activities while using the application. This DSL is extending the expressiveness of standard HTML to collect annotations for various tasks, make API calls to the back end to load the correct resources dynamically, compile and handle the views and tasks requested from the back end and also submit annotations.

We proceed by discussing the main functionalities that each back end controller provides to the application.

Application Controller. Application controller is responsible for providing information regarding the application name and also the corresponding CSS styles of the application that define how it is viewed in the mobile device. Each application is packaged with the application's unique identifier and the user can refer to the application controller with this identifier to collect those information dynamically.

User Controller. User controller is responsible for all user related functions regarding the application. Those involve signing up to the application and authenticating a user to access the application with the appropriate credentials provided by us. Upon authentication the user is provided with his user id and an experimental group number from the system. All subsequent requests are conditional to those two information, to allow for different experimental conditions. What is more,, it is also responsible for storing the first time the user was signed up to the application. The latter is important in order to allow us to calculate the proper time frame in which we will analyze our results for the users. Finally it provides information about the list of colleagues that are in the same experimental condition with each user, which is needed to support the sharing task capabilities of the application as well as the leaderboard ranking.

View Controller. The view controller is one of the two controllers that is responsible for any dynamically generated content that is available in the mobile application. The main job of the view controller is to fetch upon request from the user, the correct menu items for the application. Upon authenticating and with any subsequent launch of the application thereafter the user's first request is to the view controller to fetch the corresponding menu items conditionally to his experimental condition. Since the menu items are the main navigation point that also provide the gamification elements for the application, in this way we were able to create different experimental conditions and support the needs of the research.

Task Controller. The task controller in a similar way to the view controller is able to fetch the correct task to the user. It communicates with another important module of the application which is called task distributor. The latter is concerned with fetching the correct template and resources either for a specific task or for a task type. If the second option applies then the task distributor follows one of the available strategies in order to select the proper task (e.g. random selection). We describe the task selection strategy that we adopted in Section 4.7.

Annotation controller. The annotation controller is facilitating the collection of any kind of annotations a user creates while using the application. In the broader sense of annotation this does not only include the contributions to the tasks by the user by also the logging information that are made while using the application. To this end annotation controller is primarily concerned with gathering the answers of the users for the tasks and also attaching to them some additional data such as the date and time of creation and also the time passed for the task execution as well as the score which requests from the scoring module. Additionally it collects annotations that are related to the date and time the user opened and resumed the application and also a session time annotation when he/she closes the application. Every annotation and also the information that it incorporates is differentiated by its annotation type. For our research purposes we created and used the following annotation types:

- COMPLETED : Used for annotations for tasks completed by the users
- GETTABLE : Used to denote a sharing task from one user to another
- SHARERESPONSE : A response annotation to a shared task by a user
- APPLICATIONSTART : An annotation when the application has been opened by a user
- APPLICATIONRESUME : An annotation when the application has been resumed (i.e. is active in user's screen)
- SESSIONTIME : Used to denote the amount of time between a start or resume event and a close or paused event from the application.

Progress controller. The progress controller provides the information regarding the progress of the user while using the application. It is mainly used by the leaderboard view of the application and also the individual progress view. It fetches the overall score for the user and its peers for the leaderboard and also the individual score and the percentage of completed tasks from the total available. The scores are calculated per annotation based on the scoring function, which will be discussed later, and compiled to form the overall score by the progress controller.

4.5 Data model

The database schema that was used for the purposes of the application is shown in Figure 4.5. The main entities in it are Application, Group, User, Annotation, Task, ApplicationMenu and Template. The Application entity has the fields to store the number of experimental groups and also the stylesheet data that define the look and feel of the application. Group table stores information about each experimental condition in which a user can be included. A user is defined by his credentials and its first time logging in the system.

In the annotation table the most important fields are that of the content_string that includes in json format the necessary information depending on the type of the annotation as described earlier. It also saves the server time and the device time when the annotation was created and also the score for each contributed annotation and a code which is an auxiliary data field to support the scoring function. Also the Task and ApplicationMenu tables hold information about the name of the task or the menu item and their resources as json values in the metajson field. Both of them are associated with a template which is defined in the corresponding table in the schema and includes the html content along with some metadata in json format specific for each template in the metajson field. Finally the hidden task table is an auxiliary table that stores information about the completed tasks for each user and also the forward audience table stores information about a task that was shared from one user to anothers to support the collaboration between users.

4.6 Implementation of gamification

One of the main objectives of this study is to study gamification in crowdsourcing. Informed by our experimental design we implemented different game elements that were added to the mobile application for different experimental conditions. Those gamification elements are available from the main menu that navigates between different views of the application. Namely we created a view where a user can have access to his progress by getting information about his current score and the amount of tasks that he/she has contributed at any point in time, to support the control group condition. We also implemented a leaderboard that summarizes the scores of the users in the same experimental condition and ranks them for the competitive gamification group, and finally we created a view where the tasks that have been shared to a user from his colleagues are available in order to support the social gamification experimental condition.

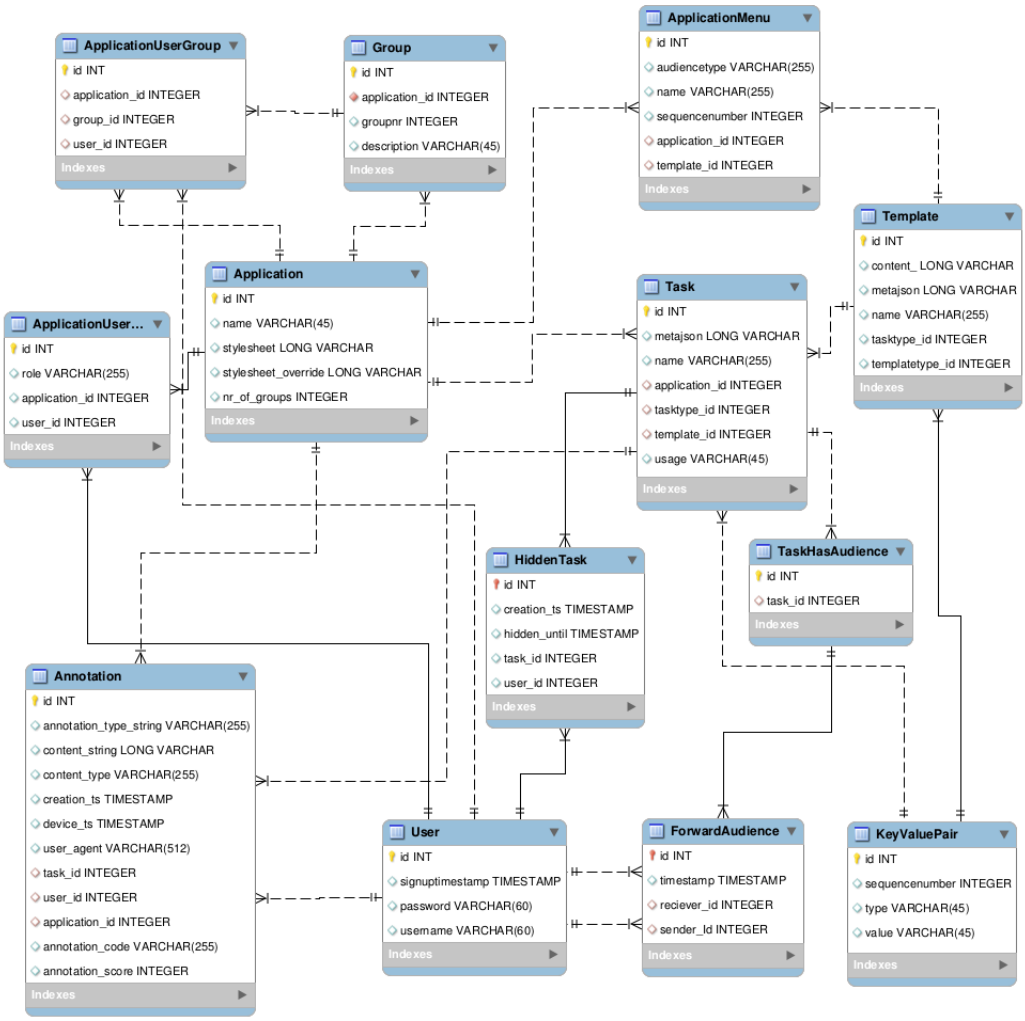

Figure 4.5: IBMCrowd/RaboCrowd Data model

In Figures 4.6,4.7 and 4.8 the three gamification views are depicted. In the following sections we provide more details on the implementation of gamification for each experimental group.

4.6.1 Annotation scores

Scoring is an important aspect of gamification and in particular for our study since it is involved in all conditions of our experiment. It is used as a feedback mechanism to inform him about his progress while contributing tasks and is visible in the application through the My Progress menu item for all experimental conditions. Additionally for those that contain leaderboards it is also ranked in the leaderboard view of the application. It is necessary that the score reflects the quality of the crowdworker and also since we are using social incentives via competition and collaboration it should be fair in order to foster interest to the user. An extra challenge is introduced when we think of scoring in the context of crowdsourcing since we are not able to reward users

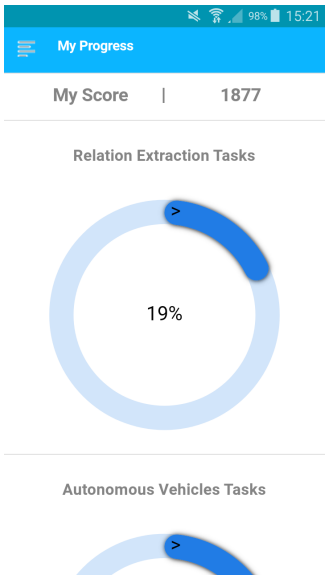

Figure 4.6: My Progress view

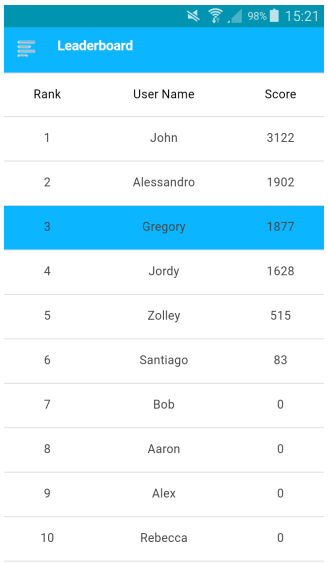

Figure 4.7: Leaderboard view

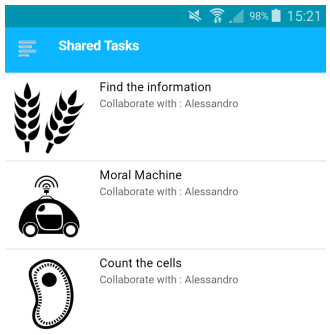

Figure 4.8: The list of shared tasks

for their answers on the basis of a ground truth or a golden standard.

Looking in related work for scoring mechanisms for crowdsourcing we realize that the two main criteria that form a scoring function are the amount of tasks that a user contributes and also the quality of the annotations [51]. In order to measure the quality of the contributions most studies base their assessment on the level of agreement to the annotations of other users [26] [25] [16]. Namely the more annotations from previous contributors are in accord with that of the user the greater the score he/she will be rewarded.

Based on these observations we decided to develop a scoring mechanism that fulfills the following criteria and its formula is shown at 4.1 below :

- The score should be an increasing function dependent on the amount of users that are in agreement with an annotation at the time that it is submitted. It should also be steeply increased for weak majorities of annotations and then gently increased as the number of annotations are vastly in agreement. This behavior reflects our need to support weak majorities that might evolve to ground truths and also provide an almost steady reward for annotations that do not provide significant additional evidence for ground truth.
- More points should be awarded by the scoring function dependent on the difficulty of the task. We define difficulty as the level of disagreement that has developed for a task and is parameterized by the number of different types of annotations that have been provided so far.
- When a variety of different annotations is previously available, the user should

be rewarded an increasing number of points depending on the order that those are sorted based on the number of annotations. This calls for a variety of homogeneous scoring functions that will consist the scoring mechanism to express the different categories of annotations. Each of these functions except for the last is upper bounded by at least one more.

$$f(x,C) = \begin{cases} [log(x+3) * g(C)] - 50, & if \quad C \neq 0 \quad and \quad x > 0 \\ [log(x+3) * g(6)] - 50, & if \quad C \neq 0 \quad and \quad x = 0 \\ 50, & if \quad C = 0 \end{cases} \tag{4.1}$$

In the scoring funciton above, x is the number of annotations which are the same as the current, C is an indicator variable that signifies in what order is the group of annotations based on their number (i.e. a group of similar annotations which are the majority get C=1) and the function $g(C)$ is a selection of constants, defined in 4.2 below, that parameterize the scoring mechanism with different scoring functions.

$$g(C) = \begin{cases} 65, & if \quad C = 1 \\ 60, & if \quad C = 2 \\ 55, & if \quad C = 3 \\ 50, & if \quad C = 4 \\ 45, & if \quad C = 5 \\ 40, & if \quad C = 6 \end{cases} \tag{4.2}$$

In 4.9 the different homogeneous scoring functions are depicted. For each value of C, which represents the level of majority (i.e. first,second etc.), a different scoring function is selected. 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. The selection of six scoring functions is based on the variety of different categories of annotations that we expected to obtain as answers. The tasks we used had a limited amount of possible categories so a total number of six scoring bands was deemed preferable and also for the reason that we wanted to keep the different scoring values produced by each function close enough to promote competition.

It is however apparent that 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. Also in 4.1 there is a default score of 50 points rewarded when there are no previous annotations which is calculated as the mean of the lowest scoring function for 20 annotations. Since this scoring value is rewarded when a user is the first to complete a task we think it represents a reasonable value, compared to others, in order to strike a balance between the unlucky event of being the first user to annotate a task and our lack of knowledge on the quality of the annotation.

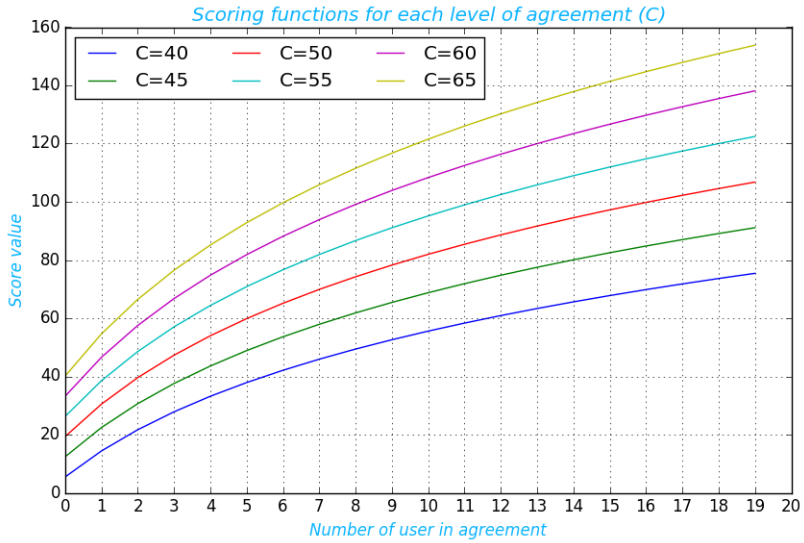

Figure 4.9: Scoring functions

To complement the collaboration incentives of the sharing task capabilities of the application and also to integrate it with the scoring mechanism, we opted to reward bonus points when two users are in agreement for a specific shared task. So for each task that is shared by a user to another, when the second completes it asynchronously, a bonus of 30% of what would be normally awarded is given to both users. To avoid users from sharing tasks and only expecting an additive value, thus making this action always favorable to just submitting a task individually, we penalize disagreement between the users by awarding them 0 points for a collaboration that ended up with disagreement in their annotations. In this way sharing a task with your 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 [41].

4.6.2 Progress bars

Progress bars are used to inform about the progress of the user regarding the amount of tasks that he/she has already contributed. In this way they incentivize users by offering them a clear goal to attain and also continuous feedback on how much is left to achieve it. In our case we implemented circular progress bars, as seen in Figure 4.6 for each task category that is available along with the score of the user in the My Progress menu item of the mobile application. Each circular bar is highlighted according to the percentage of tasks that were already completed by the user for each task category as part of the total available. It is important to notice that since we are studying engagement, tasks for each category are always available to the user even if he/she manages to annotate each of the total available at least once. This is achieved, by offering him/her the chance to work on tasks that has been submitted in the past once again. In those cases, however, the progress bars have been designed to have a 100% value to reflect that the user has already completed all available tasks.

4.6.3 Leaderboards

We have also developed leaderboards for the mobile application. As we discussed in Section 2.4.4 leaderboards can spark competitive social incentives to the users, but can also have a detrimental effect for new users onboarding to the application which are trying to compete with more experienced users who rank in the top of the leaderboard. To cope with this phenomenon we first adopted the proposal of related studies, which suggest having weekly leaderboards that summarize the scores of the users within a time span of past 7 days [35]. However during our pilot study in both IBM and Rabobank it was apparent that this could not be a probable problem since we did not notice extreme differences to the scores of the users mainly because of the scoring mechanism which awards points to the users regardless of the quality of the annotations. Another reason is that since users were able to login anytime during the timespan of the application they might have had to face a leaderboard of very low scores if our initial users were not that active during the previous week, and that would be in consequence demotivating for new users. To this end we implemented all-time leaderboards that summarize the scores obtained for the users for each annotation from the time they logged in as seen in Figure 4.7.

4.6.4 Task sharing

The task sharing capabilities of the mobile application underpin the requirements for our collaborative social gamification experimental condition as described in Section 3.3.1. Users in this condition are provided with two options when they are submitting a task. They can choose between submitting the task individually and claiming their score as described in Section 4.6.1 and the option of submitting it and also choosing a colleague in their group to share the task, with the scoring repercussions described in the same Section. If a user selects to collaborate, by pressing the corresponding button, then an actionsheet slides up in his/her screen as seen in 4.10, 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. The lifecycle of a shared task in the system is shown in Figure 4.12 along with the annotations made by the users, as described in Section 4.4. The recipient of a shared task can find what was shared with him by other users under the shared tasks menu item (Figure 4.3) of the application. In this view the user is presented with a list of tasks which are also accompanied with the name of the sender and also the task category in which they belong. This list of tasks is refreshed every time the application is started or resumed by the user.

An asynchronous execution of tasks, instead of synchronous, has been adopted in order to avoid the case in which a user who wants to collaborate cannot find another user logged in to the application at the time. This is a scenario that we would expect to encounter more frequently in an enterprise environment. 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. When

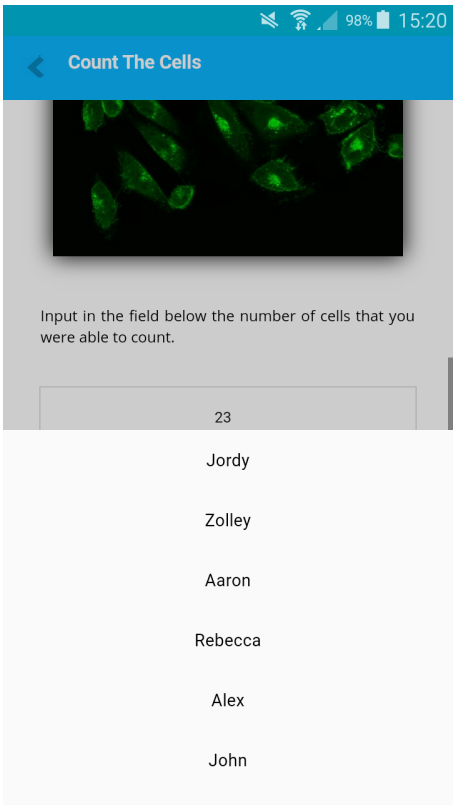

Figure 4.10: The action sheet to select users to share task with

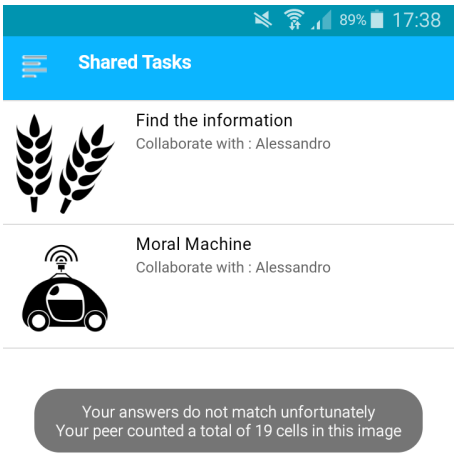

Figure 4.11: The list of shared tasks and also a toast that provides feedback on the answers of the sender after completing a shared task.

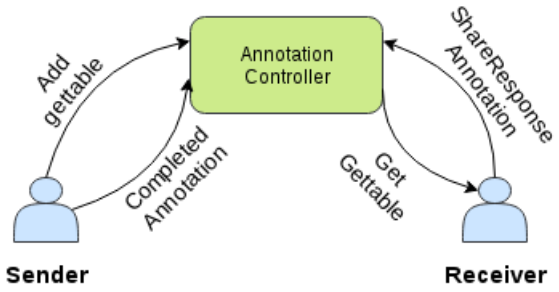

Figure 4.12: Lifecycle of a shared task

a shared task is submitted by a user a small message will pop-up to inform him/her on whether his/her annotation is the same or not with that of the sender. Additionally when the annotations of the two users differ, the pop-up message will also include the answer of the sender, as presented in 4.11.

4.7 Tasks

The selection of tasks that were included in our study are based on their relation to the domain of enterprise crowdsourcing, the incentives of the crowd and also the requirements of the current research. Since crowd incentives, task type, task complexity and design are dimensions of crowdsourcing that can play important role in the participation and engagement of the crowd [80] it is important to introduce variability across those dimensions to isolate as much as possible the effect of gamification in worker engagement. At the same time sufficient number of tasks should be generated to avoid hindering engagement due to boredom effects. To this end we selected 3 different task types that are aiming on different crowd incentives and introduce different complexity. We first describe the three task types below and how a user can select to work on each of them through the mobile application. In the following sections we discuss the process we followed to generate them and also their design.

Figure 4.13: Information extraction task

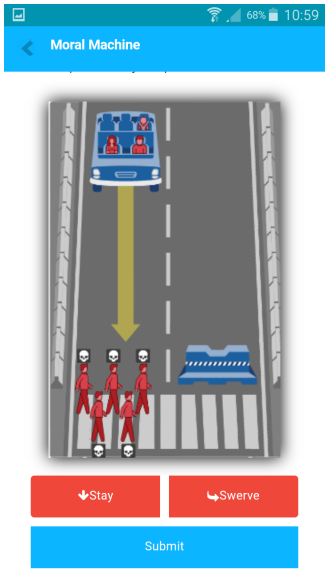

Figure 4.14: Moral Machine Task

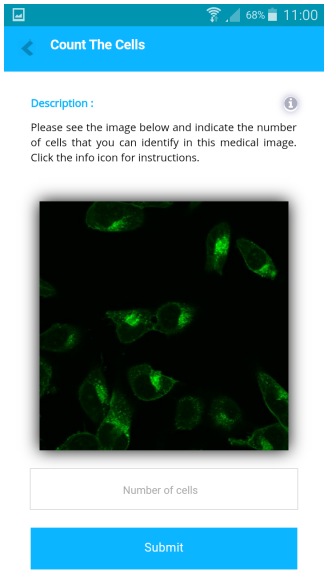

Figure 4.15: Cell Count Task

Information extraction tasks

This type of task is based on our qualitative research study that was presented in Section 3.2.2. Its main purpose is to support, with human generated data, the training phase of an automated information extraction tool which extracts relations from news sources in the agriculture domain. Since this tool is intended to be used to assist the work of Rabobank experts in the agriculture domain it addresses incentives of the crowd related to their participation in innovative projects that also the improvement of the output of the company. Moreover due to the nature of the task and also the ambiguity of language this task type is of increased complexity as compared to the rest. In those tasks the user is presented with a small paragraph which was extracted from a web news article in the agriculture domain and he is requested to annotate the relations and the entities that participate in those relations. The relations that the user

is requested to find are predetermined. The first is related to a person being the CEO of a company, the second is relation between a company and the products that this produces the last one is whether two companies are affiliated (e.g. subsidiaries). An example of this task type is illustrated in Figure 4.13.

Moral machine tasks

The moral machine tasks are survey tasks, based on a research on the morality of future Artificial Intelligence [9]. In this study the authors deploy crowdsourcing to collect opinions regarding moral decisions for autonomous vehicles. The user is presented with a scenario which involves an autonomous vehicle with a braking failure that must select between two different options that might involve the deadly (or not) injury of different characters (either as passengers or pedestrians) under different circumstances. Since the purpose of this task is to primarily raise awareness about the importance of programming moral decisions in AI it mainly addresses incentives of the crowd regarding learning, and also fun. It is also a task type with significantly less complexity than the rest, as it involves only the selection of one of the scenarios depicted in an image, as seen in Figure 4.14.

Cell count tasks

This task type (Figure 4.15) involves annotation of the number of human cells that are visible in a medical image. The input of the crowd for these tasks could potentially be used for the development of machine learning application in the medical domain. As such, they address incentives of the crowd regarding participation in interesting projects and also to work that is beneficial for the greater good. They are also tasks of intermediate complexity compared to the other two, which additionally require some basic knowledge in identifying cells in images.

The aforementioned task types are presented in a view to the user by selecting the corresponding menu item (see Figure 4.3). In this view the user can press to select between the three different task types to work on. When pressing a category a task is presented to the user based on a task selection strategy followed by the task distributor, as discussed in Section 4.4. In our case the we opted for a random assignment of tasks from those that have not been already completed by the user for a specified strategy. Every time a user submits a task it is excluded as a candidate for a subsequent execution until all tasks from a category have been completed. In the latter case all tasks are automatically becoming again candidates for selection by the task distributor.

4.7.1 Task generation

For each of the task types described in the previous section we followed a different approach to generate them. However, our primary focus, since we are studying engagement, was for the tasks to be of sufficient amount and also to introduce variability in their difficulty to keep the crowdworkers interested. To this end we generated in total 384 tasks, of which 195 tasks were for the information extraction type, 100 for the moral machine and 84 for the cell count. For the latter we used a dataset of medi-

cal images that have been already used with the CrowdTruth [5] crowdsourcing platform within IBM. For this reason in the following we focus our discussion in the information extraction and moral machine tasks.

Information extraction tasks

From the qualitative research conducted to identify an enterprise application suitable for crowdsourcing we knew that our tasks should allow users to provide training data for an information extraction algorithm operating on online news sources. As with previous studies in crowdsourcing in information extraction [37], we first opted for an automated way of extracting potential relations and their corresponding entities from news articles with the help of the AlchemyLanguage API [6]. AlchemyLanguage is a NLP api, part of IBM's Watson cognitive services, that is capable of parsing text from online sources and also extract possible entities as well as the verb words that might express a relation between them. It yields, however, highly noisy results since it is not trained specifically for the agriculture domain. Our goal was to use AlchemyLanguage to identify potential relationships and the pieces of text where those were located in order to create our tasks. To reduce the amount of noise which would be caused by choosing random news articles, we chose instead to use the searching capabilities of Meltwater [7] to target web documents specifically in the agriculture domain. To this end, we were first provided with a list of approximately 40 companies that operate in the dairy industry from the Food and Agriculture experts. With the names of these companies and keywords for the dairy domain we constructed a boolean query and used the Meltwater's search tool with which we extracted around 3000 online news articles for the dairy industry that have been written in the past year. The corpus of documents collected with this method was parsed using the AlchemyLanguage api which was able to extract candidate triplets in the form <e1,v,e2>, where e1 and e2 represent entities and r is the verb word between them. This process is schematically represented in the first 4 step in Figure 4.16.

To our current knowledge there exists no knowledge base with company entities or the relations we were interested for the agriculture domain, so the final filtering step from the candidate tasks to the final tasks was done manually. We first focused on candidate triplets that had at least one entity being in the list of the dairy companies and then examined the verbal expression found along with the second entity. Since we were interested in 3 types of relations we tried to have equal amount of tasks for each category and also included a fourth one where none of the three types of relation exist. That is, we created approximately 50 relations for each of the CEO,products and affiliation relations types and also 50 for the none relation type. The text of the news article that was used to create each information extraction task was extracted and provided by the AlchemyLanguage api and was used without any alteration.

[5] http://game.crowdtruth.org/

[6] https://www.ibm.com/watson/developercloud/alchemy-language.html

[7] https://www.meltwater.com/nl/

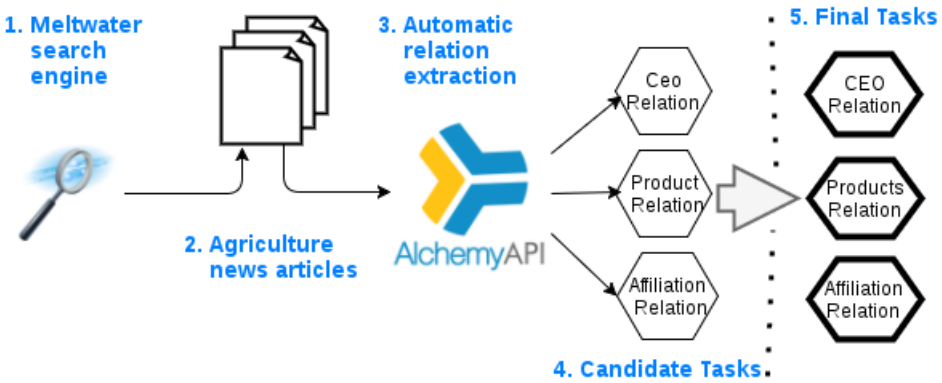

Figure 4.16: Schematic representation of the process used to generate information extraction tasks

Moral machine tasks

Although this study is not focusing on the psychological research interest that emerges from crowdsourcing opinions on the morals of Artificial Intelligence, it was however important to introduce different scenarios to keep it intriguing for the users. On top of that we wanted to have scenarios that would potentially have such value to allow for an option for a future research.

As mentioned earlier the tasks of this type are inspired from previous research in the domain ([9]), which focused in studying psychological dimensions of the problem with specific scenarios. An example of such a scenario is how utilitarian (i.e. how many lives should it sacrifice) should an autonomous vehicle be programmed to be when the user is the passenger of this vehicle. The same researchers also created an online crowdsourcing platform where internet users are free to work on moral tasks and also create scenarios of their own [8]. We used this platform to create tasks based on scenarios were both inspired by the same study and also informal interviews that we had with graduate students in psychology. We constructed tasks based on 7 different scenarios which involve different types of characters. These types are men, women, thieves, businessmen, businesswomen, doctor (men and women), homeless, girls, boys, babies, pregnants, athletes (men and women), corpulent persons (men and women), criminals, dogs or cats. The different scenarios used are listed below:

- **Scenario 1 :** In this scenario the user has to select between sacrificing a passenger in the autonomous vehicle (AV) and a varying number of pedestrians (1 to 5) of different types.

- **Scenario 2 :** The user selects between sacrificing two dyads of pedestrians where all are the same type except for one for an AV with no passengers.

[8] `http://moralmachine.mit.edu/`

- **Scenario 3 :** The user selects between sacrificing pedestrians with antithetical types (e.g. corpulent man vs man athlete) but the same gender and amount of people per group for an AV with no passengers.
- **Scenario 4 :** The user selects between sacrificing a family (i.e. man, woman and a baby) in the AV and variable type of people as pedestrians with number equal or above 3.
- **Scenario 5 :** The user selects between sacrificing two groups of pedestrians with equal amount of people and the same type but different gender for an AV with no passengers.
- **Scenario 6:** Selection between sacrificing a passenger in the AV and a group of pedestrians with the same type who cross the road legally and illegally.
- **Scenario 7:** Selection between sacrificing the driver and injuring the rest of the passengers in the AV for a family (man driver, woman and baby) and two pedestrians of the same type crossing the road.

4.7.2 Task design

Regarding the task design we followed simple design principles to accommodate the idiosyncracy of mobile crowdsourcing in combination with an enteprise crowd. Namely across all task types we strived to keep all necessary functionality to complete a task within a single screen without scrolling accounting for different mobile devices. For example for the moral tasks the two images are presented as one flipping image having each of them in one its sides and is controlled by two buttons at the bottom. For all task types we provided a small description of the task and incentivized users to press a button to pop up a view with more detailed instructions. We chose method of providing instructions, because we recognized that those would only be needed for the first few task executions, so we excluded them from the main task view to keep it simple and uncluttered from much text. At any point in time the user can press the back button and avoid working on the task by returning to the task selection view.

For the information extraction task we require the user to check one of the available checkboxes to annotate whether a relation exists or not. When a checkbox is activated a triplet becomes visible that has the empty text boxes for the entities that need to be filled by the user and a predicate in between them to denote the relation that exists between them, as seen in Figure 4.13. In this way the user can annotate triplets found in the text and he/she can also add triplets for multiple relation types as well as multiple triplets for the same relation type by pressing the add button. We also enabled copying text from the paragraph of the task and instructed the users to copy paste the entities that they find to make it simpler and also stress the fact that we want the entities in the exact form found in the text. We also employed some input validation techniques such as not allowing for the input of the cell count task to be non numerical.

For the functionality of the collaborative button, which was explained in Section 4.6.4, we added in every task a static text that explains its functionality. We decided to

add this description to avoid users in the corresponding conditions from not comprehending its functionality compared to the submit button. Finally the task execution is concluded with a thank you pop up message to the user.

4.8 Chapter conclusions

In this chapter we presented the implementation details of the instrument that was used for the experimental part of the study. We discussed the architectural and data model choices that were made to facilitate the research goals set by our experimental design. We also delineated the user experience for the different experimental conditions by describing how gamification elements were designed and incorporated in the mobile application. Finally we discussed in depth how we selected the three different types of tasks that were used in our experiments as well as the process followed to construct them and also the design choices we made. In the next chapter we move on to present the results of deploying our experimental instrument described here to study the effects of gamification in enterprise crowdsourcing.

Chapter 5

Experiment Results

The results of the experimental part of this study are included in this chapter. We first provide an overview and the demographics of our experiments in Rabobank and IBM, in Sections 5.1 and 5.2. We then move on to address the main research question of our work, as posed in 1.2, based on the selection of competitive and collaborative game mechanics which was input from our exploratory research in player types. To achieve this, we divide our main research question in three subquestions as follows:

RQ1: What is the effect of competitive and collaborative game mechanics to worker engagement in Enteprise Crowdsourcing?

RQ2: What is the effect of competitive and collaborative game mechanics to the quality of worker contributions in Enteprise Crowdsourcing?

RQ3: What is the effect of competitive and collaborative game mechanics to worker engagement and worker quality in different enterprise environments?

Based on the operationalization of the concepts of worker engagement and worker quality presented in Section 3.3.3, first is Section 5.3, we focus in analyzing engagement metrics in order to answer RQ1. Further, in Section 5.4 we steer our interest in studying worker quality for different gamification strategies to answer RQ2, while for answering RQ3, in final Section 5.5 we look upon differences regarding both engagement and quality between different enterprise environments

5.1 Overview

The experimentation period lasted two months, from 15th of May until 15th of July in Rabobank and also from 17th of May until 17th of July in IBM. We selected a 2 months interval in which employees could request credentials and log in to the application. In our analysis of the results, we normalize the observational interval to one month, from the time an employee first logged in to the application.

In this interval, for RaboCrowd application, a total of 1273 tasks were contributed by the employees, of which 343 were relation extraction tasks, 601 were moral decision tasks and 329 were cell count tasks. Employees interacted with the application for a total of 19.48 hours of which 15.1 hours (77.5%) were devoted purely in task execution. Similarly for IBMCrowd, employees contributed 502 tasks overall, with 88 of them being relation extraction, 313 moral decision tasks and 101 cell count tasks. They have spent 7.05 hours using the application of which 4.69 hours (66.5%) was task execution time.

As an overview of how task executions were performed during this observation interval with regard to the time and the day across the two companies, in Figure 5.1 we first illustrate the number of tasks executed in different times of the day (separated into hour intervals), taking into consideration only those that took place in weekdays. Similarly in Figure 5.2 we demonstrate the distribution of task executions for different days of the week for Rabobank and IBM respectively.

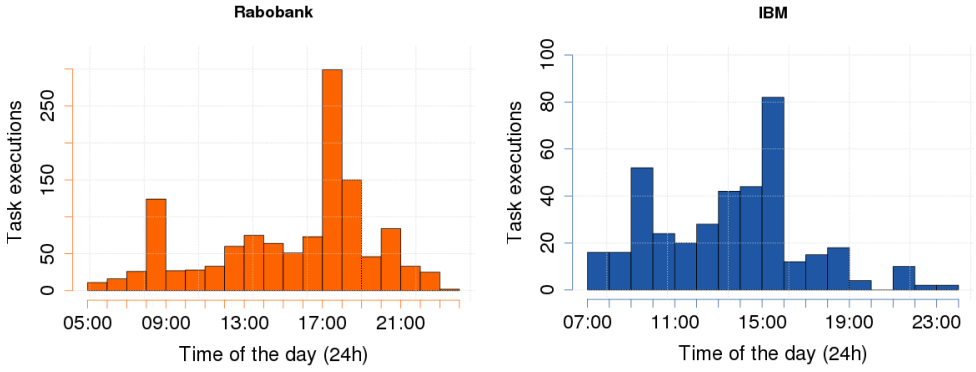

Figure 5.1: Number of task executions for different times of the day

It is interesting to notice some similarities and differences in Figure 5.1, regarding the time in which employees chose to work on tasks. One similarity is that we notice high concentration of task executions early in the morning where employees start working. This is the time slot 8:00-9:00AM for Rabobank and 9:00-10:00AM for IBM. We believe that those time slots do not coincide exactly, mainly because of the different working hour habits between the employees of the two companies. What we also notice in common, is that there is an amount of task executions outside working hours, such as late in the evening and also early in the morning, signifying that there were some participants who were willing to contribute tasks during their free time in weekdays.

An evident difference between the two is that in Rabobank high task contributions were observed close to the end of the working hours between 17:00-19:00PM, while for IBM there is a preference between 14:00-16:00PM. Those results for Rabobank are in accordance with the answers we collected from employees from the interviews for the exploratory part of our research, where the majority of them stated a preference in interacting with the application either early in the morning while arriving at work

or in the evening close to the end of their working hours. For IBM as well, those high frequency task executions time slots are in perfect alignment with responses provided by the employees in a survey for their preferred time in which they would contribute tasks, from a past research regarding enterprise crowdsourcing [6]. It is finally worth mentioning that across the two companies there is no specific preference between the observed peaks, especially for the lunch break between 12:00-13:00PM which is common for both.

Regarding the time of the day in which employees preferred to work on tasks, we notice that in both companies the vast majority of contributions was performed during Mondays. This difference is much clearer in IBM as compared to Rabobank where we notice a declining effect as the week approaches to its end. In IBM, there is almost an equal distribution of task executions across the other days of the week. This observation along with the fact that the observation interval was one month, denotes the renewed interest of the employees in crowdsourcing every week.

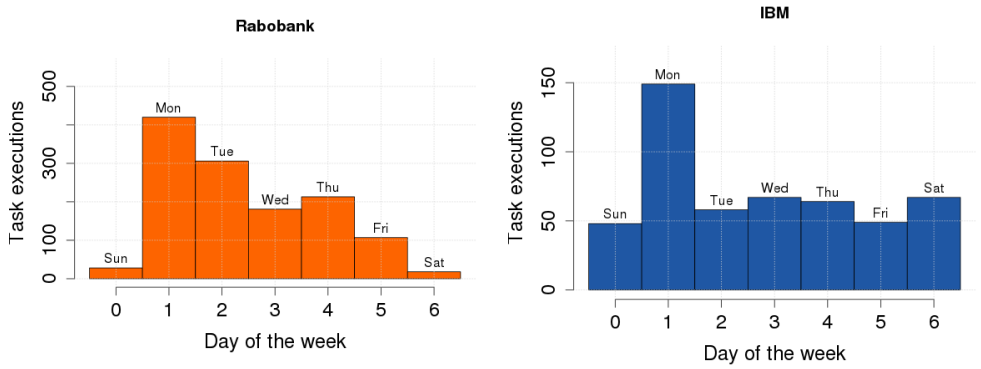

Figure 5.2: Number of task executions for different days of the week

With a few exceptions, the main observation for Rabobank is that task executions were performed mainly within working days and working hours. This is an indication that the experimental instrument was perceived as an enterprise application, for which employees were willing to devote a portion of their working time in contributing tasks. This is not the case however for IBM were we had a significant amount of task executions performed during the weekends, a phenomenon that was repeated in previous relevant studies in the company [6]. A reason behind this difference might be the different incentives between the employees of the two companies. Namely in Rabobank, task executions were more tightly connected to the development of the company (i.e. an information extraction algorithm for the Food and Agriculture department) as compared to IBM, and as such employees might have viewed it as work performed for the company which should take place during working days. We recognize however that more research is needed to understand how employees' incentives might affect the day and time of task executions.

5.2 Demographics

In Tables 5.1, 5.2 we illustrate the demographics of the employees who participated in the two experiments in Rabobank and IBM. In those tables we list only employees who logged in at least once to the application, and we separate them according to their gender, managerial position and also the department in which they are employed.

Although a total of 84 Rabobank employees volunteered to participate in the experiment only 75 of them logged in to the application, resulting in an attrition rate of 10.7%. A higher attrition rate of 23.5% was noticed in IBM in which 26 employees were active from 34 in total. Although we followed the same procedures to promote the experiment in both companies the difference in the participation levels between them is high. There are several reasons that might explain this phenomenon. One reason is the difference to the number of employees working at IBM in the Netherlands, as compared to Rabobank. Another could be due to the fact that both applications (i.e. IBMCrowd and RaboCrowd) were available in the corporate mobile stores of the companies, which for IBMers it was less often that they would have it installed in their devices compared to the Rabobank employees, thus stymying their onboarding. Finally we believe, that due to the fact that similar studies in enterprise crowdsourcing were conducted in IBM, the present research was less attractive for the employees due to the history effect, while for employees of Rabobank it was a novel and innovative study that intrigued them to participate.

Rabobank					
	Number of participants				
Gender	**Control Group**	**Group 1**	**Group 2**	**Group 3**	**Total**
male	14	13	12	16	**55**
female	6	5	4	5	**20**
Job Position Type					
manager	3	4	7	7	**21**
regular employee	17	14	9	14	**54**
unknown	0	0	0	0	**0**
Department					
SYS	2	4	4	3	**13**
Data Analytics	2	1	2	4	**9**
OPS	4	1	2	1	**8**
Human Resources	0	4	1	1	**6**
BEDR	1	1	0	2	**4**
Other	10	7	7	9	**33**
Unknown	1	0	0	1	**2**
Total	**20**	**18**	**16**	**21**	**75**

Table 5.1: Rabobank's group demographics

Specifically in Rabobank, approximately 26.6% of the employees are females and the remaining 73.4% consist of males. In general we notice a fair dissemination of users in the 4 experimental conditions both for the number of participants as well as the

genders. Slight discrepancies are mainly noticed due to the attrition effect, which was something we could not control for. Managers form 28% of the total sample and the rest 72% are employees holding non-managerial positions. We also notice that employees' departments are diverse with 44% of them being unique and the rest 56% consists of 5 different departments and some that we were not able to collect information for.

For our sample in IBM, 80.7% of the participants were males while only 19.2% were females, which is an analogy observed also in previous studies within the company [64]. We notice that due to the attrition effect there is misrepresentation of females in Group 3. There is also lack of participants holding a managerial position in our sample, with only 1 being assigned in Group 2. As far as the departments are concerned, in symphony with our sample in Rabobank, 34.6% of the departments of the employees are unique while the rest 65.3% is either from one of the four departments listed in Table 5.2 or we were unable to retrieve this information.

Overall there was a fair distribution of the employees across different experimental conditions for both experiments. Considering also that with the round robin group assignment method we opted for, we mainly controlled for equal distribution of number of employees and gender between the groups, while managerial positions and department were left random. What is more, the high dispersion of participants in different departments for the two companies minimizes the probability of the diffusion of treatments as a validity threat of our experimental design.

IBM					
	Number of participants				
Gender	**Control Group**	**Group 1**	**Group 2**	**Group 3**	**Total**
male	5	6	4	6	**21**
female	2	1	2	0	**5**
Job Position Type					
manager	0	1	0	0	**1**
regular employee	5	6	6	6	**23**
unknown	2	0	0	0	**2**
Department					
Global Business Services	2	2	2	0	**6**
Global Technology Services	2	0	1	2	**5**
IBM Cloud	0	1	0	2	**3**
IBM Executive Staff	1	0	0	0	**1**
Other	0	4	3	2	**9**
Unknown	2	0	0	0	**2**
Total	**7**	**7**	**6**	**6**	**26**

Table 5.2: IBM's group demographics

5.3 Worker Engagement

In order to address RQ1, we separate our analysis of employee engagement across four dimensions. Namely for our experimental conditions and the two enterprises, we study the number of task executions, the session time, the number of sessions and also the task dwell time.

Number of task executions

As an indication of user engagement we first look into the number of tasks contributed by the employees of the two companies during the experimentation period.

We start first by displaying in Table 5.3, the descriptive statistics for the number of tasks executions for the two experiments, separated across the three task types that we used. In those results we omit the task contributions of a user in the control experimental group, who contributed approximately 68.5% of the total task contributions of the group. The level of contributions from this specific user is regarded as an outlier in our study and as such it will be excluded from the analysis of the results that follows.

It is evident from these results that moral decision tasks were the most preferred tasks for the employees in all experimental conditions and for both experiments. We also notice higher dispersion of our samples for this task type as compared to the other two, which signifies wider level of interests. Since this task type was selected in order to adhere to fun incentives of the crowd and also introduces less task complexity, this remark highlights the importance of those two parameters in enterprise crowdsourcing with regard to task contribution. It is also interesting to notice that the average contributions of the participants for the relation extractions do not appear to have noticeable differences comparing all cases, despite the fact of different incentives between the employees of the two companies. More specifically in Rabobank we do not notice considerably higher contributions, compared to IBM, for the information extraction tasks, although from informal interviews after the end of the experiment employees stated that their main motivation was to contribute for the development of machine learning applications within the company.

We are now interested in aggregating the task contributions to analyze the effects of gamification for all task types. From previous research, which compared competitive with collaborative game mechanics in crowdsourcing [62], there is a theoretical underpinning which states that both competitive and collaborative game mechanics in isolation can have positive effects to engagement while when combined they can have detrimental effects. Due to the similarity of this study we form our hypothesis correspondingly, in order to test this theory. Namely, our alternative hypothesis is that leaderboards (i.e. competitive) and task sharing (i.e. collaborative) will increase the amount of tasks contributed by the users in those groups compared to the control group while their combination will lessen the effect. The descriptive statistics, calculated for all the tasks across our experimental conditions, shown in Table 5.3 provide an initial credibility which is in alignment with our alternative hypothesis for both experiments.

		Rabobank			IBM		
	Task type	**mean**	**sd**	**total**	**mean**	**sd**	**total**
Control Group	*Relation extraction*	3.18	2	51 (31.5%)	2	2.16	14 (14.6%)
	Moral decisions	3.68	3.44	70 (43.2%)	10.28	12.55	72 (75%)
	Cell count	2.41	2.23	41 (25.3%)	1.42	1.27	10 (10.4%)
	All tasks	**3.11**	**3.58**	**162**	**4.57**	**8.14**	**96**
Group 1	*Relation Extraction*	5.62	7.41	90 (20.8%)	4.33	4.63	26 (17.1%)
	Moral Decisions	13.17	12.16	224 (51.6%)	11.85	11.81	83 (54.6%)
	Cell Count	7.05	8.52	120 (27.6%)	6.14	4.74	43 (28.3%)
	All tasks	**8.68**	**9.99**	**434**	**7.6**	**8.22**	**152**
Group 2	*Relation Extraction*	2.08	1.92	25 (14.2%)	6	9.61	30 (30.3%)
	Moral Decisions	8.06	10.49	121 (68.8%)	7.6	12.09	38 (38.4%)
	Cell Count	2	2.33	30 (17%)	6.2	9.01	31 (31.3%)
	All tasks	**4.19**	**6.99**	**176**	**6.6**	**9.59**	**99**
Group 3	*Relation Extraction*	1.7	1.61	29 (19.4%)	2.8	1.92	14 (18.4%)
	Moral Decisions	4.93	8.4	74 (49.7%)	9.5	17.56	57 (75%)
	Cell count	2.87	5.09	46 (30.9%)	1.66	1.52	5 (6.6%)
	All tasks	**3.1**	**5.65**	**149**	**5.42**	**11.56**	**76**

Table 5.3: Descriptive statistics of number of task executions

To further assess the significance of the observed result, we first fit a Poisson regression model to our data obtained from the experiment conducted in Rabobank. However, the results obtained from our model suggest violation of the model's assumption of equal mean and variance, which holds for Poisson distributed data. We hypothetize that this might be an indication of overdispersed data. To further examine whether our hypothesis for overdispersed data holds, we use an overdispersion test statistic [11], which tests the null hypothesis that the data at hand do not suffer from overdispersion. The results of this test, leads us to reject the null hypothesis ($p < 0.001$), thus we opt to model the data with a Negative Binomial regression model. The coefficients of this model and their significance are listed in Table 5.4, omitting the intercept.

	Rabobank		IBM	
#TaskExecutions	**Coefficients**	**Significance**	**Coefficients**	**Significance**
Leaderboards	1.0395	**0.0127 ***	0.4595	0.464
Task Sharing	0.2547	0.5568	0.7715	0.236
Leadrboads+Task Sharing	-0.1837	0.6522	-0.0794	0.904

Table 5.4: The coefficients and their statistical significance of a Negative Binomial regression model fit to describe the effect of different game mechanics to the number of task executions in Rabobank (***: .001 significance, **: .01 significance, *: .05 significance

From the coefficient results of the Negative Binomial regression model, we can confirm the alternative hypothesis that leaderboards have a significant positive effect ($p<0.05$) to the expected number of task executions, resulting in an increase of 282% ($e^{1.03}$) compared to the control group. Additionally, we notice that there is a slight positive effect of sharing tasks, while combined with the leaderboards there is a negative effect as compared to the control group. We cannot however confirm the alternative hypothesis for these gamification elements since their effects were not found to be significant.

Similar to our analysis for Rabobank, we opt for a Negative Binomial regression model for the data obtained from our experiment in IBM. What we see is that there is a positive effect for leaderboard and task sharing alone while combined there is a negative effect to the expected number of tasks executions. Although those results are in accordance with the ones observed for Rabobank, there is no compelling evidence that would lead us to accept our hypothesis.

We are also interested to see how many tasks where shared for the groups that had the collaborative option (i.e. Group 2 and Group 3). The number of tasks shared by the employees and also the responses to those shared tasks are listed in Table 5.5. We notice that in Rabobank 10 and 49 tasks were shared from the employees to their colleagues asking for collaboration for Group 2 and Group 3 respectively, however very few responses were noted for Group 3 while for Group 2 all shared tasks remained unanswered. In IBM only 12 tasks were shared in the collaborative group and 10 in the group that had both task sharing and leaderboards combined and in all cases no responses were collected for those tasks. In general 5.7% of the total tasks contributed for Group 2 in Rabobank were due to task sharing while for Group 4 this percentage climbs to 35.6%. For IBM those percentages are 13.13% for Group 2 and 13.15% for Group 3.

	Rabobank		IBM	
	Group 2	**Group 3**	**Group 2**	**Group 3**
Shared Tasks	10	49	13	10
Responses	0	4	0	0

Table 5.5: Number of shared tasks and responses for the two groups and companies

Those results, considering the analogy to the total executed, are promising regarding the adoption of the task sharing capabilities of the application, however the collaboration effect was severely hindered by the very low response rate to the tasks shared, which was essential in completing the engagement loop of this mechanic. We believe that this was mainly because of the absence of a notification mechanism that could inform employees when they had a list of tasks that were shared to them. Instead, we relied in their curiosity to navigate in the application and check for shared tasks when they opened it. We however deliberately avoided notifications since this could potentially bias the specific groups against the control and leaderboard ones. We would also expect higher percentages of tasks shared if the application was personalized and the employees could see to whom they shared the tasks, since this feature would be important for the social dynamics that this mechanic targets.

In [62] they stress the fact that users prefer selecting clearly either a collaborative or competitive strategy rather than having to select between them, with an inclination to competitive. This is something that we can notice both for Rabobank and also for IBM where the majority of tasks contributed by the employees in Groups 3 were not selected to be shared.

Session length

To further understand the level of engagement of the employees to the mobile crowdsourcing application, we measure the time that they have spent interacting with it. This metric aggregates the time spent executing tasks and also the time interacting with the application. The descriptive statistics across the four experimental conditions and the two companies are illustrated in Table 5.6 below.

	Rabobank			IBM		
	mean	**sd**	**median**	**mean**	**sd**	**median**
Control Group	178.18	196.61	115.39	217.53	290.63	129.56
Group 1	179.69	193.83	119.2	226.31	288.54	95.78
Group 2	127.93	161.15	73.18	191.97	193.07	132.73
Group 3	108.78	202.07	42.34	205.26	286.03	100.06

Table 5.6: Descriptive statistics for session time (secs) per group and company

Focusing on the first order statistics for session time for Rabobank, it is interesting to notice that Control Group and Group 1 have similar values. Since in our previous results we showed that Group 1 had higher mean task executions compared to the Control group, we would expect to also have higher mean session time for this group. This leads us to believe that participants in Group 1 were, on average, more swift in executing tasks during their sessions. This is an argument, on which we will return when we will analyze the task dwell time for our experimental conditions. What we also observe for this experiment is that participants in Groups 2 and 3 had lower mean session times compared to the first two groups. Regarding IBM we see that the values of the first order statistics are much more closer to each other across the different experimental conditions with small differences. This signifies that there was little effect of the different gamification techniques to the mean session time for this experiment.

We now turn our interest in discovering whether any of the observed differences found for the time spent in the application are statistically significant. Our hypothesis is that leaderboards and task sharing will have positive effects to the session length observed while when combined the effect will lessen. We look into the distribution of our dependent variable for the two experiments and for the different groups, to inform the selection of a proper statistical test. Using a gaussian kernel we estimated the distribution of our dependent variable as depicted in Figure 5.3. It is readily seen from this figure that we cannot rely on test statistics which assume normal distributions. For this reason we opt for Kruskal-Wallis test which is a rank-based nonparametric test, which tests for statistically significant differences between different distributions. A Kruskal-Wallis test conducted for our experiment in IBM did not show statistically significant differences between the experimental groups (p=.951), whereas for Rabobank we were able to show that there is a statistically significant difference (p=.001). This is a first indication that gamification was perceived differently between the two enterprises.

Based on our previous results, we perform a post-hoc analysis only for our experiment in Rabobank. We conduct pairwise Mann-Whitney U tests with Holm-Bonferroni correction to account for Type I error inflation to our results. The results of our tests are

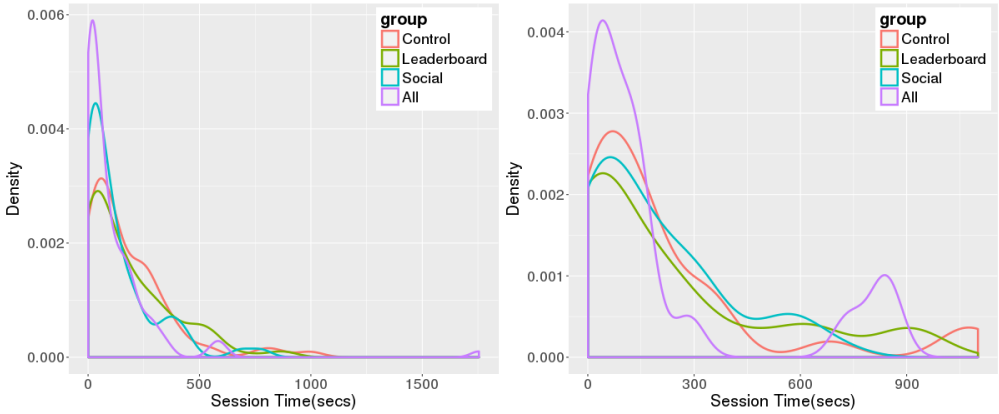

Figure 5.3: Gaussian kernel density estimation of application session time for Rabobank (left) and IBM (right)

summarized in Table 5.7. We can see that there is a significant difference observed in the comparison between the Control Group and Group 3 and also between Group 1 and Group 3. This signifies that participants in the group that incorporated both a leaderboard and the task sharing functionality spent significantly less time per session in the application compared to only having basic gamification or only the leaderboards. The difference between Group 1 and 3 coincides with our results for task contribution in the previous section and also provides evidence for our hypothesis. No significant difference were observed between the social gamification (i.e. Group 2) and any of the rest groups.

In our analysis for the task contributions, we hypothetized that the diminished interest of the users in groups where task sharing was included was due to their little interaction with the menu item under which shared tasks could be found by their peers. In order to verify this assumption we subtracted from the total session time of the participants the task execution time to find the time spent interacting with the rest features of the application and specifically the gamification elements. For the Rabobank experiment we found that on average, more time was spent to the application, excluding task execution time, for Group 1 compared to the rest of the groups where this time was almost the same. This might be an indication that users in this experimental condition have spent time checking their position in the leaderboard that might have motivated them more to continue contributing. This also signifies that participants in Group 2 were not necessarily checking the shared tasks menu item to engage with collaborative contributions and their interaction with the application was more closely related to that of the baseline group. In a similar analysis for IBM experiment we did not notice such differences between the groups. This remark once more, underlines that gamification elements introduced in our experiment were perceived differently in IBM and might have been less motivating for this population overall.

	Rabobank			
	Control Group	**Group 1**	**Group 2**	**Group 3**
Control Group	-	.853	.201	**.005**
Group 1	.853	-	.268	**.006**
Group 2	.201	.268	-	.494
Group 3	**.005**	**.006**	.494	-

Table 5.7: Mann-Whitney-Wilcoxon pairwise significance tests (with Holm-Bonferroni adjustment) for session time (in seconds) across treatment groups and the two companies

Number of sessions

In this section, we compare the number of times employees opened the application across the different treatment groups for our experiments. In Table 5.8 we summarize the descriptive statistics of this variable.

	Rabobank				IBM			
	mean	**sd**	**median**	**total**	**mean**	**sd**	**median**	**total**
Control Group	4.25	3.04	4.5	51	4.4	3.2	3	22
Group 1	7.43	9.81	4	119	2.85	2.11	2	20
Group 2	4.5	3.42	3	54	16.66	17.78	9	50
Group 3	5.27	6.05	2	95	3.5	3	2	14

Table 5.8: Descriptive statistics of the number of sessions per group and company

Specifically for the Rabobank experiment we observe that the application has been opened more times in total and on average for Group 1 compared to the Control Group. A similar amount of times is observed for Group 2 compared to the baseline while slightly higher values are obtained for Group 3. For IBM we do not observe the same effects, where for both groups containing leaderboards we have counted less times where the application was opened while for Group 2 we have higher values compared to all other groups. Studying in depth the results for this group we understood that the value for this specific group, is merely because of the increased interest of a specific user.

Similar to our analysis regarding the session time, we hypothetize that when leaderboards or task sharing are present this would motivate the users to open the application more times, while when combined this might result in less times using the application. In order to ascertain this hypothesis we first test the assumption of normally distributed data using a Shapiro-Wilk test for all experimental conditions and the variable of interest. In all our tests we were able to reject the null hypothesis of normally distributed data ($p<.05$), thus we use a Kruskal-Wallis test to test for possible statistically significant differences between our experimental conditions. The obtained results for the Rabobank experiment does not provide us with enough evidence to reject the null hypothesis ($p=.926$). Similarly for IBM we are not able to show any statistically significant differences for the number of the times the application was opened for the different experimental conditions ($p=.101$).

Another way of analyzing engagement of the employees is by counting the time in-

terval between contiguous sessions, as defined in [45]. In that work they measure the amount of disengagement as the time it takes for workers to return to the application accounting for breaks of more than 5 minutes and they analyze it in terms of a cumulative distribution expressing the probability of disengagement under a specific threshold. We use the same way to analyze way to analyze the distribution of disengagement time with the difference of not using the threshold of 5 minutes to define a contiguous session since in our case the time between sessions is usually in the order of hours, which clearly signifies two independent contiguous sessions. The results of the empirical cumulative distributions of inter-session times are depicted in Figure 5.4 for Rabobank and IBM.

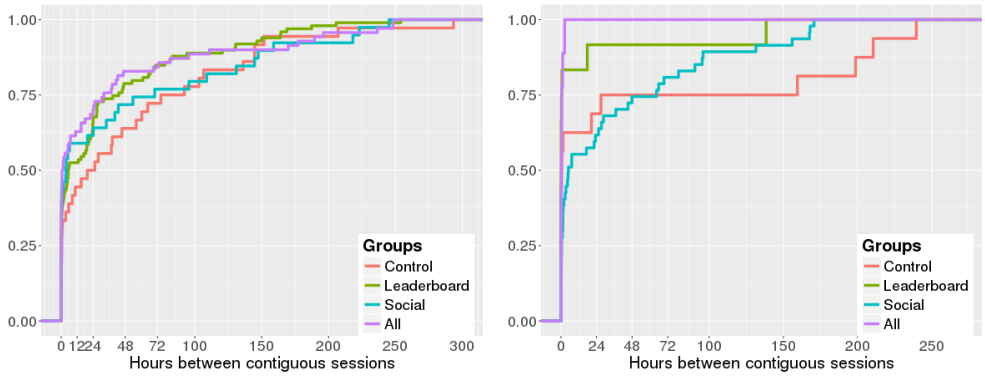

Figure 5.4: Empirical Cumulative Distribution of inter-session times (hours) for the application across the experimental groups and the two companies (Rabobank left, IBM right)

What is common in the Figure 5.4 for both experiments, is that we notice higher probability of users 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 users who had leaderboards were more inclined to revisit the application compared to the other conditions. Merging our observations from the times that the application has been opened with the current ones, we can assume that the users with the competitive and collaborative game mechanics although they had the same interest in reopening the application, as with the users who had only leaderboards, they only did it a few times and then quickly disengaged from the application.

Task dwell time

We compare task dwell time across the different treatment groups, defined as the net time spent in task execution by the employees. From our experimental design, we know that the three task categories in our experiments, introduce different levels of complexity which severely influence the time spent executing them. Namely, by design, we expect the information extraction tasks to require more time to be completed compared to the cell count tasks and the moral machine tasks to be the less complex compared to the other two. What is more, they adhere to different incentives of the employees which might as well affect the specific variable.

For the reasons mentioned, we separate our analysis per task type and experimental

condition, since an aggregation of the variable on the dimension of task type would reveal little information. The descriptive statistics of task dwell time are illustrated in Table 5.9 below, for each different task category and experimental condition for the two companies.

		Rabobank				IBM			
	Task type	**mean**	**sd**	**median**	**total**	**mean**	**sd**	**median**	**total**
Control Group	*Relation extraction*	84.69	80.12	52.78	4319.65	100.97	81.4	100.43	1413.63
	Moral decisions	38.07	36.49	24.22	2665.45	18.6	21.69	11.44	1339.57
	Cell count	30.6	29.37	19.14	1254.93	35.94	19.91	37.55	359.4
	All tasks	**50.86**	**57.59**	**32.25**	**8240.03**	**32.4**	**46.2**	**15.35**	**3112.6**
Group 1	*Relation Extraction*	101.29	76.95	81.07	9116.66	54.87	42.07	46.83	1426.64
	Moral Decisions	26.72	70.34	12.51	5986.58	19.63	16.96	13.33	1629.71
	Cell Count	20.26	19.64	14.28	2431.72	17.11	12.09	12.48	736.14
	All tasks	**40.40**	**69.65**	**16.46**	**17534.97**	**24.95**	**26.02**	**14.51**	**3792.5**
Group 2	*Relation Extraction*	93.17	70.43	77.54	2329.36	99.4	54.75	78.19	2982.28
	Moral Decisions	20.15	20.44	12.77	2438.37	21.07	11.43	16.92	800.80
	Cell Count	31.94	21.85	26.29	958.27	26.56	21.3	19.12	823.54
	All tasks	**32.53**	**40.95**	**16.26**	**5726.01**	**46.53**	**48.05**	**28.63**	**4606.63**
Group 3	*Relation Extraction*	89.91	61.98	79.15	2607.39	184.31	124.4	130.63	2580.41
	Moral Decisions	34.61	36.69	20.29	2561.18	27.74	38.19	13.79	1581.61
	Cell count	24.32	20.79	14.64	1118.87	35.45	16.62	26.79	177.28
	All tasks	**42.19**	**45.78**	**22.03**	**6287.45**	**57.09**	**86.56**	**19.47**	**4339.31**

Table 5.9: Descriptive statistics for task dwell time (in seconds) per group and company

From the first order statistics for the Rabobank experiment, we can observe that comparing the Control group with Group 1, there is on average less time spent to task execution especially for the tasks that were more easy to complete. This ascertains our assumption set when we analyzed the session time for the different groups that although between those groups a similar amount of session time was observed task executions for the leaderboard group were significantly more due to the fact that users in this group had a tendency to execute them more quickly. For the same experiment we also observe less time spent on average for Groups 2 and 3 compared to the Control. Interestingly enough for the information extraction tasks the differences are less profound between the groups compared to the other two task types. We believe that this is the case, because the specific task type is relevant to the development of an application specifically for Rabobank and as such the effects of gamification for this specific task might have been overshadowed by willingness of the employees to perform well for this task. For the experiment in IBM although we observe similar behavior between the Control group and Group 1 as with the experiment in Rabobank, there is in general fluctuation of the observed values depending on the task type and the experimental group.

Our hypothesis, concerning the task dwell time, is that participants in the leaderboard condition would spent less time while executing tasks compared to the Control group, mainly focusing in gathering points and improve their position in the leaderboard faster. We further hypothetize that users with the task sharing functionality would on average have higher task execution times because of their incentive to match their answers with that of their peers when sharing tasks in order to gain more points. Finally we also hypothetize that the combination of the aforementioned game mechanics in Group 3 would lead again in less task execution times since the competitive game dynamics would prevail.

To test our hypothesis, we are interested in modeling our data with a regression model. Due to the nature of the observer dependent variable being continuous and also as a result of an exploratory analysis which revealed a consistent positive skewness for our samples, we believe that the assumption of a gamma distribution fits well for our dependent variable. To reinforce our argument we empirically estimate the gamma distribution parameters of our data for each experimental condition and task type and perform a Kolmogorov-Smirnov test. The specific test, tests the null hypothesis which claims that our data are not from the assumed gamma distribution. With the few exceptions, which are mainly due to the relative small samples at hand, our tests showed that there is a statistically significant evidence that leads us to reject the null hypothesis. Therefore we proceed with our analysis to model the task dwell time variable using a generalized gamma linear model.The results of the coefficients and their significance after fitting the model for the two experiments are shown in Table 5.10 for Rabobank and Table 5.11 for IBM.

Looking first at the results obtained from the Rabobank experiment we can see that those are in alignment with our hypothesis for the cell count tasks, where we indeed observe a statistically significant decreasing effect for the task execution time, when leaderboards only are present. For the same task type we also see that there is a positive effect when task sharing functionality is included while when this is combined with leaderboards the effect is again decreasing. For the latter two however we were unable to prove that they were statistically significant. Similar results are observed for Group 1 and Group 3 for the moral decision tasks, although contrary to our hypothesis we observe a significant negative effect for our variable when task sharing functionality was present, which is for Group 2. Interestingly enough for the relation extraction tasks there is a positive effect for all groups compared to the baseline which negates our hypothesis for Group 1 and Group 3. We believe that those results are partially explained from the use of game mechanics and confounding factors such as employees incentives for specific tasks are also playing an important role. This is evident for example in the relation extraction tasks for this specific experiment for which Rabobank employees had clear incentives of contributing them for the benefit of the company. Therefore we believe that gamification for this task explain little of the positive effect to the time spent executing them. 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.

Focusing on the results of the IBM experiment, we see that our hypothesis regarding the leaderboards only can be confirmed for the relation extraction tasks and the cell count tasks, where they were also found to be significant. A statistically significant positive effect was found for the combination of collaborative and competitive game mechanics for the relation extraction and the moral decision tasks which contradicts our hypothesis, while task sharing functionality has had a negative effect for all tasks except the moral decision ones, although none of those results were proven to be significant. Compared to the Rabobank experiment it is interesting to denote the difference observed for the relation extraction tasks where in IBM for all groups except group 3 we have a negative effect of gamification on the time spent. We assume that this is explained because of the different incentives of the employees of the two companies

Rabobank		
Relation Extraction Tasks	**Coefficients**	**Significance**
Leaderboard	0.178	0.2
Task Sharing	0.095	0.627
Leaderboard + Task Sharing	0.059	0.750
Moral Decision Tasks	**Coefficients**	**Significance**
Leaderboard	-0.354	0.181
Task Sharing	-0.636	**0.029 ***
Leaderboard + Task Sharing	-0.095	0.763
Cell Count Tasks	**Coefficients**	**Significance**
Leaderboard	-0.412	**0.013 ****
Task Sharing	0.042	0.846
Leaderboard + Task Sharing	-0.229	0.243

Table 5.10: The coefficients and their statistical significance of a Gamma Generalized Linear Regression model fit to describe the effect of different game mechanics to task dwell time for the different task categories in Rabobank (*** : .001 significance, ** : .01 significance, * : .05 significance

on the specific task. Finally there is a discrepancy for the moral decision tasks were we see that in IBM there were a consistent positive effect of gamification to the time executing those tasks as compared to Rabobank.

IBM		
Relation Extraction Task	**Coefficients**	**Significance**
Leaderboard	-0.609	**0.009 ****
Task Sharing	-0.015	0.944
Leaderboard + Task Sharing	0.601	**0.023 ***
Moral Decision Tasks	**Coefficients**	**Significance**
Leaderboard	0.053	0.751
Task Sharing	0.124	0.557
Leaderboard + Task Sharing	0.399	**0.034 ***
Cell Count Tasks	**Coefficients**	**Significance**
Leaderboard	-0.741	**0.004 ****
Task Sharing	-0.302	0.25
Leaderboard + Task Sharing	-0.013	0.973

Table 5.11: The coefficients and their statistical significance of a Gamma Generalized Linear Regression model fit to describe the effect of different game mechanics to task dwell time for the different task categories in IBM (*** : .001 significance, ** : .01 significance, * : .05 significance

Overview of results

We concentrate our results, regarding employees' engagement with our experimental tool in order to answer RQ1 which was posed at the start of this chapter. We look into four metrics that indicate engagement with the mobile crowdsourcing applications to draw our answer.

For the expected number of tasks executed we observed a significant increase when competitive gamification elements such as leaderboards were used in isolation. The

collaborative game mechanic, which was the task sharing functionality, proved beneficial while the result was not that profound compared to only the competitive ones. There is also an indication that the combination of them might be detrimental for crowdsourcing activities within the enterprise. Those results were consistent for both experiments in Rabobank and IBM.

The diminished effect of combining collaboration and competitiveness was also visible in the session time that we measured for the participants for both companies. This difference however was much bolder for Rabobank compared to IBM signifying the importance of the enterprise environment to the effect of those gamification elements. We believe that this result can be attributed 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. To further prove this assumption we looked into the net time spent by the employees interacting with the gamification elements were we found that when those two were combined were not leading to higher levels of interaction compared to when competitive or collaborative game mechanics were used alone.

Interestingly, the time spent to the application when competitive game mechanics are used is not necessarily improved to the baseline despite the fact of increased crowdsourcing output for these groups. These results, in combination with our analysis of the task dwell time signifies higher throughput of tasks from the users get only competitive game mechanics for both enterprise environments. Further analysis on our results of task dwell time shows that the relationship of gamification and task dwell time is heavily influenced by the incentives of the employees and the task types used.

Finally, by analyzing the times that the application was opened we were able to show that competitive game mechanics lead to higher probabilities of reengaging with the app in a specific time interval, both when used in isolation and when used in combination with collaborative ones. The retention levels however, when they are used in isolation are higher. For collaborative ones we observed higher probabilities of reengagement to the baseline but less than the competitive ones alone.

5.4 Worker quality

In this section we focus on measuring the quality of contributions of the employees in the two experiments in order to answer RQ2. We base our analysis on the responses collected from the employees for the task types of cell count and information extraction, since it is not meaningful to suggest quality of responses for the moral decision tasks.

We first focus in analyzing the contributions of the employees on the cell count tasks. The possible input to those tasks was a positive integer. Due to the lack of a golden standard regarding the number of cells in the images that we introduced in our experiment, the quality of the contributions is based on the plurality answer agreement metric which was presented in Section 3.3.3. This metric accounts for the number of

tasks in which a worker was in symphony with the majority divided by his total contributions to limit the metric within the interval from 0 to 1. A quality score of 0 signifies consistent disagreement of an employee to the majority vote in for the set of tasks that she worked on, while a value of 1 signifies a consistent disagreement with the majority vote for the same set.

In order to avoid calculating the quality of employees' contributions based on weak majorities, we also incorporate the labels obtained for the tasks, from the users in the pilot phase of the application and we also merge the labels obtained from 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 us to more robust results regarding experimental users' contributions for these tasks. We also recognize that within the available set of tasks there are some with higher level of difficulty and some with less, thus it is possible that participants of a group would only annotate easy tasks. However, due to the randomness of the task assignment that we designed in our experimental tool, we believe that an equal analogy of difficult and easy tasks were solved in all of our experimental conditions.

Our hypothesis for the quality of contributions is that competitive and collaborative game mechanics will increase the quality of answers while combined they will have stronger effect. This is based on the fact that scores were based on agreement while the collaborative game mechanic that we used is also favoring agreement of contributions between the users sharing tasks. The boxplots in Figure 5.5 depict the distribution of plurality answer agreement scores that we obtained for the two experiments. From these plots there is an indication that competitive and collaborative game mechanics are indeed beneficial for higher quality contributions compared to the baseline for both companies. Their combination in Rabobank yielded even better results on average while for IBM was on par with the group that had only the leaderboard. Revisiting Table 5.5 which informs us of the number of shared tasks for each group that had collaborative game mechanics we cannot definitely conclude that this mechanic was pivotal for the difference observed between Group 3 and and the control. However the difference between Rabobank's Group 3 and IBM's Group 3 is more noticeable in terms of the total shared tasks which might explain also the difference in the quality of the contributions we observe between them. It is also interesting to see that even though competitive and collaborative game mechanics lead to quicker task execution times, as we have seen in our results for task dwell time, that is not done to the expense of lower quality of contributions.

Now we want to focus in analyzing whether the observed differences in our experimental groups are statistically significant. We proceed by using a pairwise Mann-Whitney-Wilcoxon non-parametric test for both the experiments. The results of our tests are summarized in Table 5.12. From the obtained values we conclude that we were unable to prove any statistically significant difference between the experimental conditions for both companies. It is however, worth mentioning that with the exception of the participants in the control group the level of agreement achieved for this task type in promising and it highlights the potential for quality contributions in in-house crowdsourcing campaigns.

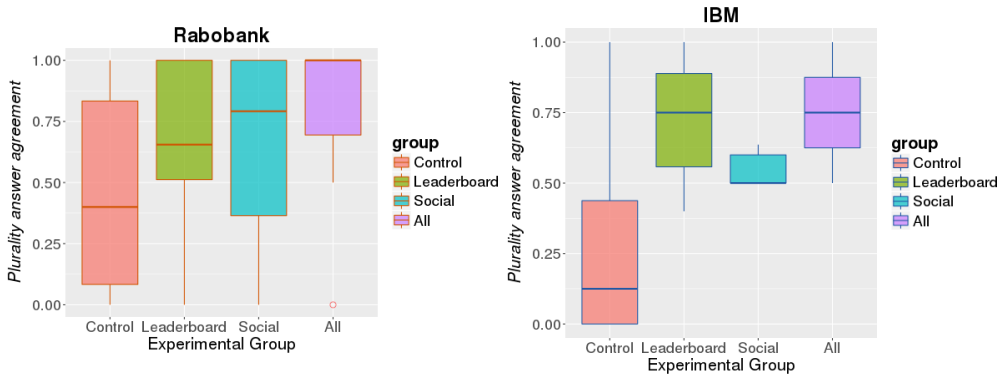

Figure 5.5: Distribution of plurality answer agreement scores of employees for the cell count tasks for Rabobank (left) and IBM (right)

	Rabobank				IBM			
	Control Group	**Group 1**	**Group 2**	**Group 3**	**Control Group**	**Group 1**	**Group 2**	**Group 2**
Control Group	-	.7	.94	.25	-	.36	.66	.86
Group 1	.7	-	1	.94	.36	-	.86	1
Group 2	.94	1	-	.94	.66	.86	-	1
Group 3	.25	.94	.94	-	.86	1	1	-

Table 5.12: Mann-Whitney-Wilcoxon pairwise significance tests (with Holm-Bonferroni adjustment) for plurality answer agreement scores across treatment groups and the two companies

We now turn our interest in measuring the quality of contributions for the relation extraction tasks of our experiments. We are interested in analyzing the quality of the labels provided based on the relations that were found by the employees for the tasks that they worked on. Once again to enrich our corpus of labels, similarly to the practice followed for the cell count tasks we incorporate the labels provided also from the pilot users and also we merge the answers obtained from the two experiments while limiting our analysis to the experimental users.

Although the manual step introduced to generate those tasks, as explained in Section 4.7, provided us with a first indication of the existent relations in the snippets of texts that were used for our tasks, we base our analysis on the agreement that is achieved by a worker on a task unit compared to the answers of all other workers on the same task. For this purpose we use the average worker-worker agreement which was explained in Section 3.3.3 and is suggested in similar studies [3] for relation extraction tasks.

Congruent to our hypothesis for the quality in cell count tasks, we hypothetize that competitive and collaborative gamification mechanics will results in better quality answers from the employees while their combination will have an even better effect. We first visualize the distribution of the quality scores we obtained for the experimental users for both experiments in Figure 5.6. From these plots we observe that contrary to our hypothesis in the Rabobank experiment we obtain slightly less agreement for the leaderboard group compared to the control on average. An almost equal level of agreement for the collaborative group and less when those two are combined. Same

effects are observed for IBM.

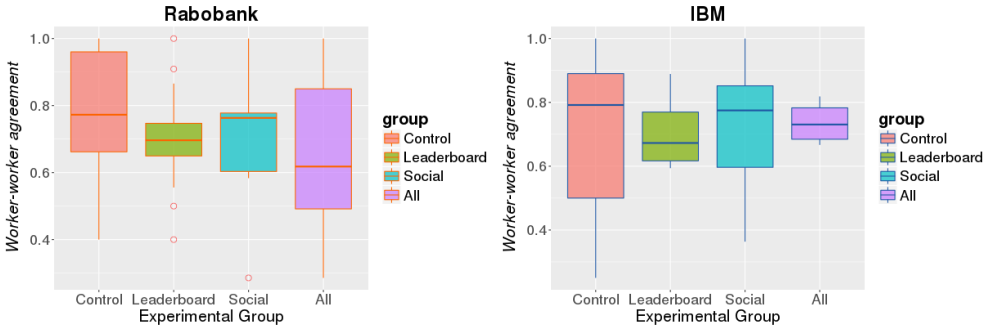

Figure 5.6: Distribution of average worker-worker agreement scores of employees for the relation extraction tasks for Rabobank (left) and IBM (right)

To analyze the differences between the groups we use a Kruskal-Wallis non-parametric test for the two experiment. The results obtained for Rabobank indicate that there is no statistically significant evidence that there is a difference between the distribution of quality scores across our experimental conditions (p=.518). Similarly for IBM we were unable to prove any statistically significant differences for the different treatments (p=.947). What we also noticed is that for the specific task type none of the participants shared it with one of their peers, therefore we believe that little can be explained for the quality of contributions for this specific task for the groups that had this capability.

Overview of results

In our effort to answer RQ2, we observed that both collaborative and competitive game mechanics were beneficial in raising the quality of contributions from the employees, as well as in combination for the cell count tasks, although those results were not found to be statistically significant. Such discrepancy was not visible however with the relation extraction tasks were the obtained results were much more close for the different experimental conditions. Those results we believe indicate the potential benefits of such gamification elements, but in addition unravel their dependencies on the specific task types under which we consider them.

5.5 Gamification and enterprise environment

In this section we focus on answering RQ3. More specifically, we focus on analyzing the effect of gamification mechanics for different enterprise contexts for engagement and also worker quality, in order to gain an understanding on how those might affect crowdsourcing activities. We want to compare the metrics of engagement and worker quality used in our analysis in previous sections between same treatment groups for Rabobank and IBM.

By juxtaposing the results found in the previous sections we were able to identify some similarities and some differences between the two companies. Namely we observed similar patterns of crowdsourcing activities regarding 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 game mechanics compared to the baseline, a small increase when collaborative ones were used and a diminished effect when those were combined. On the other hand comparing the expected number of task contributions between the two companies, we were able to see a slight increase of this metric for Rabobank compared to IBM when competitive game mechanics were only used, while the opposite was observed when collaborative ones where introduced and when they were combined. We were also able to notice differences in the session times calculated, where in IBM 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 were introduced in Rabobank compared to IBM while when only task sharing was used IBM's employees were more eager to open the application.

Based on those results, we hypothetize that enterprise environment plays a role in how gamification is perceived, which results in different patterns of crowdsourcing activities. To test our hypothesis we use the non-parametric Kolmogorov-Smirnov test for the engagement metrics of task contributions, session time and number of times that the application has been opened. We make comparisons between the same experimental conditions as used in the two experiments, to assess our hypothesis. The results of these tests are summmarized in the first three rows of Table 5.13. As it is readily seen from those results we are unable to support our hypothesis.

	Rabobank/IBM			
	Experimental Groups			
Variable	**Control**	**Group 1**	**Group 2**	**Group 3**
#TaskExecutions	.934	.431	.717	.762
Session Time	.929	.464	.085	.076
#OpenApp	.908	.791	.387	.962
Plurality Answer Agreement	.727	.991	.18	.995
Average Worker-Worker Agreement	.935	.99	.976	.259

Table 5.13: Kolmogorov-Smirnov significance tests between experimental conditions across the two experiments for the variables of number of task executions, session time and times where the application was opened.

Regarding the quality of contributions for the cell count task type and the information extraction task type, by revisiting our results in Section 5.4 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. We again hypothetize that the enterprise environment affects the relationship of gamification elements with the observed quality of the contributions. Similarly to our previous analysis we use Kolmogorov-Smirnov tests for which the results are illustrated in the last two rows of Table 5.13. The obtained results signify that we cannot reject or accept our hypothesis regarding differences of quality contributions between the two experiments.

Overview of results

In our effort to answer RQ3, we were able to notice commonalities between the two experiments regarding the usage of competitive and collaborative game mechanics. Looking into our engagement metrics there is evidence of a general preference of employees' to competitive game mechanics when gamification is considered in enterprise crowdsourcing. Although not in an equal magnitude to the competitive ones, collaborative game mechanics proved beneficial as well for some of our engagement metrics. There is also indication of a complex interplay between competitive and collaborative game mechanics that leads to decreasing effects of interest from the side of the employees to the crowdsourcing endeavor. In addition for the quality of contributions both game mechanics have proven to be beneficial either in isolation or in combination, depending on the task type considered.

By studying into the potential effects of enterprise environment in the gamification elements used, we were unable to prove any significant effect of them both for the engagement of the employees to the crowdsourcing application and also for the quality of their contributions for the task introduced.

5.6 Chapter Conclusions

In this Chapter we analyzed the results of our two enterprise crowdsourcing experiments. Although not central to our main research question, we have first analyzed the time in which crowdsourcing activities took place. We saw a general interest into crowdsourcing either at the start of a working day or close to the end with no general preference for times in between.

We then focused on analyzing cues of engagement of the employees by looking into how they performed crowdsourcing activities within the observational interval of one month. We showed that there is a higher preference of competitive game mechanics in enterprise contexts which translates to higher levels of expected task contributions and also higher probabilities of reengaging with the application in short time intervals with prolonging effects as well. We also observed that although collaborative mechanics are beneficial to some dimensions of our engagement analysis, they are not necessarily as effective as the competitive ones. Interestingly enough, we also show that when competitive and collaborative game mechanics are combined, their interplay results in lessen interest in crowdsourcing activities within the enterprise.

We also analyzed the quality of collected task contributions for two different task types based on agreement scores. Although for collaborative and competitive game mechanics,depending on the task and the employees' incentives, can lead to less task execution time, their effect is at least as good as the baseline in which they are not used.

Finally we looked into the effects of the enterprise environment for the relation between gamification and engagement and crowdsourcing quality. By analyzing our metrics of engagement and quality between the experimental groups we were unable to support our hypothesis that gamification elements can be perceived differently and

lead to different effects depending on the enterprise environment. We believe however that larger sample sizes are needed to unravel this probable mediating effect.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

The aspriration of this study was to research how gamification can effectively support enterprise crowdsourcing activities, in terms of employee engagement and also quality of their contributions. In order to attain our goal, we organized our research in an exploratory and an explanatory part. We used the former to gain an understanding on which gamification techniques are more suitable in enterprise environments as well as to identify an application for which enterprise crowdsourcing could be used. We rely on the explanatory part to build knowledge on the ramifications of different gamification techniques to crucial parameters of success for enterprise crowdsourcing.

Based on Bartle's theory of player types, our qualitative exploratory analysis affirmed, for the enterprise environment, the suggestion of non mutually exclusive player type characteristics. More importantly, by combining our 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 funneled our explanatory research, for which we deployed a gamified mobile crowdsourcing application which combines competitive and collaborative game mechanics. We used our experimental tool into two large multinational enterprises for an observational interval that lasted two months. Our experimentation involved 75 active participants from Rabobank and 26 from IBM. We disseminated the participants fairly in different treatment groups and we analyzed the effects of collaborative and competitive game mechanics as well as that of the enterprise environment to the engagement and quality of contributions of the employees.

The use of gamification in an enterprise crowdsourcing was viewed positively from the employees who engaged with the application. From informal interviews after the end of the experiment with some of the participants in our experiments we understood that the experimental tool was intuitive and easy to use. For this matter, one employee stated to us : *"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 gamifica-

tion elements were perceived as motivating and retained their interest in contributing tasks. One employee when asked about it, he revealed to us : *"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 relation 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"*.

More importantly our experimental results, suggest a preference of competitive game mechanics over the collaborative ones, both in terms of the expected task contributions as well as time spent to the application. We were also able to support a theory that states the detrimental effects that their combination might yield in users' engagement. 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. 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. We recognize, however that experiments, with higher amounts of participants are necessary in the future to solidify these results.

6.2 Validity threats

Possible internal and external validity threats in our study are related to the history effect, selection and also diffusion of treatment [56].

The history effect is caused by things that change participants' environment and can have detrimental effects in internal validity. This is addressed in our study 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. We recognize however that since we allowed for flexible sign up times for the participants within the observational interval we do not completely control for effects that might arise. What is more, there is also the possibility of a history effect threat for IBM as compared to Rabobank, where in the first experimentation with enterprise crowdsourcing application has already took place [6]. 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.

The selection effect can have negative effects to the internal validity of our study by introducing different number of participants per experimental group. It is also an issue for the external validity if groups with specific characteristics are formed. We coped

with both threats by assigning participants randomly to experimental conditions. Signing up to the application wa permitted by requesting credentials as discussed in Section 3.3.2. Credentials were provided by us from a predetermined list of users which are assigned in experimental groups in a round robin fashion. Looking back at the demographics of our experiments in Section 5.2 we showed that this strategy yielded acceptable results considering the number of our samples.

Diffusion of treatment refers to the potential threat in 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. This is something that is also empirically shown in Section 5.2 were we noticed that a large amount of employees who participated were from diverse departments within the two companies, thus minimizing the specific threat.

6.3 Future work

The current study provides some preliminary steps in untangling the competitive and collaborative game dynamics that game elements foster in enterprise crowdsourcing and their effects. As crowdsourcing is by definition built on collaborative incentives, it is of our strong belief that there is high potential for future research that revolves around developing effective collaborative engagement loops for the employees.

A first step into evolving the current study would be to provide personalization to the crowdsourcing application by logging in, for example, with a social network account. It would be interesting to explore how personalization 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.

From our experiments it was also readily seen that task types and their design play a crucial role in enterprise crowdsourcing when gamification is concerned. For example the moral machines which were related to fun incentives were adopted widely and were interesting for the majority of the employees. It would be interesting to research which task parameters mediate the effect of gamification in enterprise crowdsourcing and whether there are some which possibly negate its merits.

Finally, although we were not focused in analyzing the prolonging effects of gamification in this study due to the duration of our experiments we noticed significant declining effects of engagement within weeks of enrolling to the application. For this reason, tt would be beneficial to conduct longitudinal studies on gamification in enterprise crowdsourcing to assess its potential prolonging effect to crowdsourcing ac-

tivities. Another possibility is to propose and study, possibly dynamic, gamification schemes that could remedy ephemeral effects.

Bibliography

[1] Mohammad Allahbakhsh, Boualem Benatallah, A Ignjatovic, HR Motahari-Nezhad, E Bertino, and S Dustdar. Quality control in crowdsourcing systems. *IEEE Internet Comput*, 17(2):76–81, 2013.

[2] Omar Alonso and Stefano Mizzaro. Can we get rid of trec assessors? using mechanical turk for relevance assessment. In *Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation*, volume 15, page 16, 2009.

[3] Lora Aroyo and Chris Welty. Measuring crowd truth for medical relation extraction. In *2013 AAAI Fall Symposium Series*, 2013.

[4] Lora Aroyo and Chris Welty. The three sides of crowdtruth. *Journal of Human Computation*, 1:31–34, 2014.

[5] Richard Bartle. Hearts, clubs, diamonds, spades: Players who suit muds. *Journal of MUD research*, 1(1):19, 1996.

[6] Sarah Bashirieh. Mobile Crowdsourcing in an Enterprise Environment. Master's thesis, Delft University of Technology, the Netherlands, 2016.

[7] Roland Benabou and Jean Tirole. Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3):489–520, 2003.

[8] Alexander Bogner and Wolfgang Menz. Das theoriegenerierende experteninterview. In *Das Experteninterview*, pages 33–70. Springer, 2002.

[9] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. The social dilemma of autonomous vehicles. *Science*, 352(6293):1573–1576, 2016.

[10] Anne Bowser, Derek Hansen, Yurong He, Carol Boston, Matthew Reid, Logan Gunnell, and Jennifer Preece. Using gamification to inspire new citizen science volunteers. In *Proceedings of the first international conference on gameful design, research, and applications*, pages 18–25. ACM, 2013.

[11] A Colin Cameron and Pravin K Trivedi. Regression-based tests for overdispersion in the poisson model. *Journal of econometrics*, 46(3):347–364, 1990.

[12] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, April 1960.

[13] Mihaly Csikszentmihalyi and Isabella Selega Csikszentmihalyi. *Optimal experience: Psychological studies of flow in consciousness*. Cambridge university press, 1992.

[14] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM, 2011.

[15] Pinar Donmez, Jaime G Carbonell, and Jeff Schneider. Efficiently learning the accuracy of labeling sources for selective sampling. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 259–268. ACM, 2009.

[16] Anca Dumitrache, Lora Aroyo, Chris Welty, Robert-Jan Sips, and Anthony Levas. 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*, pages 16–31. CEUR-WS. org, 2013.

[17] Carsten Eickhoff, Christopher G Harris, Arjen P de Vries, and Padmini Srinivasan. 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*, pages 871–880. ACM, 2012.

[18] Lee Erickson, Irene Petrick, and Eileen Trauth. Hanging with the right crowd: Matching crowdsourcing need to crowd characteristics. 2012.

[19] Enrique Estellés-Arolas and Fernando González-Ladrón-De-Guevara. Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2):189–200, 2012.

[20] Oluwaseyi Feyisetan, Elena Simperl, Max Van Kleek, and Nigel Shadbolt. Improving paid microtasks through gamification and adaptive furtherance incentives. In *Proceedings of the 24th International Conference on World Wide Web*, pages 333–343. ACM, 2015.

[21] David R Flatla, Carl Gutwin, Lennart E Nacke, Scott Bateman, and Regan L Mandryk. Calibration games: making calibration tasks enjoyable by adding motivating game elements. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 403–412. ACM, 2011.

[22] Uwe Flick. *An introduction to qualitative research*. Sage, 2009.

[23] David Geiger and Martin Schader. Personalized task recommendation in crowdsourcing information systemsâĂŤcurrent state of the art. *Decision Support Systems*, 65:3–16, 2014.

[24] Ian Glover. Play as you learn: gamification as a technique for motivating learners. 2013.

[25] Jorge Goncalves, Simo Hosio, Denzil Ferreira, and Vassilis Kostakos. Game of words: tagging places through crowdsourcing on public displays. In *Proceedings of the 2014 conference on Designing interactive systems*, pages 705–714. ACM, 2014.

[26] Ido Guy, Anat Hashavit, and Yaniv Corem. Games for crowds: A crowdsourcing game platform for the enterprise. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1860–1871. ACM, 2015.

[27] Juho Hamari. 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):236–245, 2013.

[28] Juho Hamari and Jonna Koivisto. âĂIJworking out for likesâĂİ: An empirical study on social influence in exercise gamification. *Computers in Human Behavior*, 50:333–347, 2015.

[29] Juho Hamari, Jonna Koivisto, and Harri Sarsa. Does gamification work?–a literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences*, pages 3025–3034. IEEE, 2014.

[30] Lars Hetmank. A synopsis of enterprise crowdsourcing literature. 2014.

[31] Matthias Hirth, Tobias Hoßfeld, and Phuoc Tran-Gia. Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. *Mathematical and Computer Modelling*, 57(11):2918–2932, 2013.

[32] Mokter Hossain. Users' motivation to participate in online crowdsourcing platforms. In *Innovation Management and Technology Research (ICIMTR), 2012 International Conference on*, pages 310–315. IEEE, 2012.

[33] Jeff Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.

[34] Kai Huotari and Juho Hamari. Defining gamification: a service marketing perspective. In *Proceeding of the 16th International Academic MindTrek Conference*, pages 17–22. ACM, 2012.

[35] Panagiotis G Ipeirotis and Evgeniy Gabrilovich. Quizz: targeted crowdsourcing with a billion (potential) users. In *Proceedings of the 23rd international conference on World wide web*, pages 143–154. ACM, 2014.

[36] Ehud D Karnin, Eugene Walach, and Tal Drory. Crowdsourcing in the document processing practice. In *International Conference on Web Engineering*, pages 408–411. Springer, 2010.

[37] Sarath Kumar Kondreddi, Peter Triantafillou, and Gerhard Weikum. Combining information extraction and human computing for crowdsourced knowledge acquisition. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pages 988–999. IEEE, 2014.

[38] Markus Krause, Marc Mogalle, Henning Pohl, and Joseph Jay Williams. A playful game changer: Fostering student retention in online education with social gamification. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 95–102. ACM, 2015.

[39] Janaki Kumar. Gamification at work: Designing engaging business software. In *International Conference of Design, User Experience, and Usability*, pages 528–537. Springer, 2013.

[40] Gioacchino La Vecchia and Antonio Cisternino. Collaborative workforce, business process crowdsourcing as an alternative of bpo. In *International Conference on Web Engineering*, pages 425–430. Springer, 2010.

[41] Edith Law and Luis Von Ahn. Input-agreement: a new mechanism for collecting data using human computation games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1197–1206. ACM, 2009.

[42] Edith LM Law, Luis Von Ahn, Roger B Dannenberg, and Mike Crawford. Tagatune: A game for music and sound annotation. In *ISMIR*, volume 3, page 2, 2007.

[43] Tak Yeon Lee, Casey Dugan, Werner Geyer, Tristan Ratchford, Jamie C Rasmussen, N Sadat Shami, and Stela Lupushor. Experiments on motivational feedback for crowdsourced workers. In *ICWSM*, 2013.

[44] Janette Lehmann, Mounia Lalmas, Elad Yom-Tov, and Georges Dupret. Models of user engagement. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 164–175. Springer, 2012.

[45] Andrew Mao, Ece Kamar, and Eric Horvitz. Why stop now? predicting worker engagement in online crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.

[46] Andrzej Marczewski. User types. *Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design*, 1:65–80, 2015.

[47] Elaine Massung, David Coyle, Kirsten F Cater, Marc Jay, and Chris Preist. Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 371–380. ACM, 2013.

[48] Elisa D Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N Tuch. Disassembling gamification: the effects of points and meaning on user motivation and performance. In *CHI'13 extended abstracts on human factors in computing systems*, pages 1137–1142. ACM, 2013.

[49] Elisa D Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N Tuch. Do points, levels and leaderboards harm intrinsic motivation?: an empirical analysis of common gamification elements. In *Proceedings of the First International Conference on gameful design, research, and applications*, pages 66–73. ACM, 2013.

[50] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1003–1011. Association for Computational Linguistics, 2009.

[51] Benedikt Morschheuser, Juho Hamari, and Jonna Koivisto. Gamification in crowdsourcing: a review. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 4375–4384. IEEE, 2016.

[52] Stefanie Nowak and Stefan Rüger. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the international conference on Multimedia information retrieval*, pages 557–566. ACM, 2010.

[53] Alexander J Quinn and Benjamin B Bederson. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1403–1412. ACM, 2011.

[54] Marta Rauch. Best practices for using enterprise gamification to engage employees and customers. In *International Conference on Human-Computer Interaction*, pages 276–283. Springer, 2013.

[55] Vikas C Raykar, Shipeng Yu, Linda H Zhao, Gerardo Hermosillo Valadez, Charles Florin, Luca Bogoni, and Linda Moy. Learning from crowds. *Journal of Machine Learning Research*, 11(Apr):1297–1322, 2010.

[56] Colin Robson and Kieran McCartan. *Real world research*. John Wiley & Sons, 2016.

[57] Richard M Ryan and Edward L Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68, 2000.

[58] Maik Schacht and Silvia Schacht. Start the game: Increasing user experience of enterprise systems following a gamification mechanism. In *Software for People*, pages 181–199. Springer, 2012.

[59] Katie Seaborn and Deborah I Fels. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74:14–31, 2015.

[60] Victor S Sheng, Foster Provost, and Panagiotis G Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 614–622. ACM, 2008.

[61] Lei Shi, Alexandra I Cristea, Suncica Hadzidedic, and Naida Dervishalidovic. Contextual gamification of social interaction–towards increasing motivation in social e-learning. In *International Conference on Web-Based Learning*, pages 116–122. Springer, 2014.

[62] Kristin Siu, Alexander Zook, and Mark O Riedl. Collaboration versus competition: Design and evaluation of mechanics for games with a purpose. In *FDG*, 2014.

[63] Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Y Ng. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the conference on empirical methods in natural language processing*, pages 254–263. Association for Computational Linguistics, 2008.

[64] Laurentiu Catalin Stanculescu, Alessandro Bozzon, Robert-Jan Sips, and Geert-Jan Houben. Work and play: An experiment in enterprise gamification. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 346–358. ACM, 2016.

[65] Osamuyimen Stewart, Juan M Huerta, and Melissa Sader. Designing crowdsourcing community for the enterprise. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 50–53. ACM, 2009.

[66] Osamuyimen Stewart, David Lubensky, and Juan M Huerta. Crowdsourcing participation inequality: a scout model for the enterprise domain. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 30–33. ACM, 2010.

[67] Scott Thiebes, Sebastian Lins, and Dirk Basten. Gamifying information systems-a synthesis of gamification mechanics and dynamics. 2014.

[68] Gustavo F Tondello, Rina R Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E Nacke. The gamification user types hexad scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, pages 229–243. ACM, 2016.

[69] Dmitry Ustalov. Towards crowdsourcing and cooperation in linguistic resources. In *Russian Summer School in Information Retrieval*, pages 348–358. Springer, 2014.

[70] Jennifer Wortman Vaughan. Adaptive task assignment for crowdsourced classification. 2013.

[71] Petros Venetis and Hector Garcia-Molina. Quality control for comparison microtasks. In *Proceedings of the first international workshop on crowdsourcing and data mining*, pages 15–21. ACM, 2012.

[72] Luis Von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326. ACM, 2004.

[73] Luis Von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.

[74] Luis Von Ahn, Ruoran Liu, and Manuel Blum. Peekaboom: a game for locating objects in images. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 55–64. ACM, 2006.

[75] Maja Vukovic and Claudio Bartolini. Towards a research agenda for enterprise crowdsourcing. In *International Symposium On Leveraging Applications of Formal Methods, Verification and Validation*, pages 425–434. Springer, 2010.

[76] Maja Vukovic, Jim Laredo, and Sriram Rajagopal. Challenges and experiences in deploying enterprise crowdsourcing service. In *International Conference on Web Engineering*, pages 460–467. Springer, 2010.

[77] Peter Welinder, Steve Branson, Pietro Perona, and Serge J Belongie. The multidimensional wisdom of crowds. In *Advances in neural information processing systems*, pages 2424–2432, 2010.

[78] Jacob Whitehill, Ting-fan Wu, Jacob Bergsma, Javier R Movellan, and Paul L Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in neural information processing systems*, pages 2035–2043, 2009.

[79] Man-Ching Yuen, Irwin King, and Kwong-Sak Leung. A survey of crowdsourcing systems. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 766–773. IEEE, 2011.

[80] Haichao Zheng, Dahui Li, and Wenhua Hou. Task design, motivation, and participation in crowdsourcing contests. *International Journal of Electronic Commerce*, 15(4):57–88, 2011.

[81] Gabe Zichermann and Christopher Cunningham. *Gamification by design: Implementing game mechanics in web and mobile apps*. " O'Reilly Media, Inc.", 2011.

[82] Angelina Ziesemer, Luana Müller, and Milene Silveira. Gamification aware: users perception about game elements on non-game context. In *Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems*, pages 276–279. Brazilian Computer Society, 2013.

Appendix A

Privacy Statement

A.1 Privacy Statement

RaboCrowd is an application developed by IBM Center of Advanced Studies `http://www.research.ibm.com/university/cas/benelux`

It facilitates the creation, modification and monitoring of crowdsourcing campaigns within the enterprise and is used as an experimental tool to conduct research in enterprise crowdsourcing.

This page informs you of our information practices. Those include the collection, use and disclosure of Personal Information from users of the RaboCrowd application. We describe the information that are collected and how they will be used, and also include information about the steps we take to ensure and protect your privacy. By using this application, you agree to the collection and use of information in accordance to this policy.

A.2 Information we collect

RaboCrowd receives and collects information when the application is installed and used.

A.2.1 Information you provide

When you register to use the application you use predefined username and password that is provided to you by an administrator of the application. In this way your access to the application is completely anonymized and only the administrator can link your username to personal information.

A.2.2 Automatically collected information

While using the application we log and collect information about your activity. This includes the first time you signed up to the service, the time using the application, the

number of tasks you submitted, the time you submitted a task and the elapsed time for submitting a task.

A.3 How we use your information

The information collected by the application are analyzed anonymously and used exclusively for the following purposes:

- For research purposes
- For the development of Artificial Intelligence applications

The automatically logged information will be used to analyze the activities of the users of the application. The analysis will be conducted in aggregated form and no link to any personal information will be used either by the administrator or the research team. The information provided by the users through task submission will be used as training data for the development of machine learning applications.

A.4 Data Security

RaboCrowd application uses a secure server to store and collect any information, which is located in a secure network. Data stored in this server will be only accessible and used by the team conducting the research. The application also provides an option to opt-out of the system. By opting-out we will immediately take all necessary actions to delete any information you provided.

A.5 Contact Us

If you have any questions about this Privacy Policy, please contact us at one of the following e-mail addresses:

- casbnl@nl.ibm.com
- greg.afentoulidis@nl.ibm.com

Appendix B

Interview guides

B.1 Player types interview guide

Based on the requirements of the research and the theoretical foundations on which it will be based, we identify six separate topics for the interviews. The first four are related to questions that aim to unravel the player types in the enterprise both of the interviewee and also of his/her coworkers. In the fifth topic we are interested in eliciting the dominant player type of the interviewee and also understand which suits the majority of the employees. While the final is related to the available times in which an employee would be able to contribute in enterprise crowdsourcing. For the first four topics we choose to address questions to the interviewee that reflect the main characteristics of each category and ask them to what degree they believe that those describe them.

1. **Topic 1: Socializer player type**

    a) Question 1 : How important is your social environment at work as a motivator for your everyday tasks (e.g. being able to share knowledge, to contribute to the well-being of others and enjoying group activities)?

    b) Question 2 : How important is in your opinion to the efficiency of your work the sense of being part of team and interact with other people in your working environment?

2. **Topic 2: Achiever player type**

    a) Question 1 : Often in a working environment appraisal of performance is shown with certain types of rewards. Should in your opinion rewards always come hand in hand with successful undertaking of work related tasks?

    b) Question 2 : Do you think rewards are one of the main motivators of your work?

3. **Topic 3: Killer player type**

a) Question 1 : Do you think is important to adhere to standardized processes and rules while conducting you work duties?

b) Question 2 : In the case you are restricted by some rules, would you try to adhere to them as much as possible or would you try to find a better way?

4. **Topic 4: Explorer player type**

   a) Question 1 : Do you generally enjoy working independently and being able to follow you own path?

5. **Topic 5: Dominant player type and generalization**
   So far we have discussed several employee qualities. To summarize them we talked about:

   a) Employees who are mainly driven by their social environment and their interaction with it.

   b) Employees who are reward oriented and strive for competence.

   c) Employees who are out of the box thinkers and are intrigued to question the status-quo.

   d) Employees who prefer to work independently and are curious to try new things.

   a) Question 1 : In which one of those categories would you put yourself into? Can you explain aspects of your personality that reflect those qualities?

   b) Question 2 : Which of the aforementioned categories of qualities, in your opinion, describes best an employee of Rabobank?

6. **Topic 6: When and how will employees contribute in enterprise crowdsourcing.**

   a) Question 1 : What time of the day would you ideally want to be able to contribute to enterprise crowdsourcing?

   b) Question 2 : How much time do you think you would be able to devote everytime you use the application?

B.2 Expert interview guide

Based on the requirements of the research and also most common caveats associated with expert interviewing found in the literature we formulate three relevant topics that we want to address in each interview and also the related questions to these topics as seen below:

1. **Topic 1: The experts' processes when working for a FAR report**
    a) Question 1 : Please describe what are the job duties of an expert in the food and agriculture research department.
    b) Question 2 : Describe the process an expert undergoes when compiling a FAR report.
        i. Can you please provide a recent example of your work and provide more details on this example? (PQ)
    c) Question 3 : Can you recognize independent steps in the process of compiling FAR reports? If yes can you briefly describe them?
        i. Can you please provide a concrete example where you used them? If yes can you describe how you applied them and what was the outcome? (PQ)
2. **Topic 2: The data sources that the experts use**
    a) Question 1 : What are some of the data sources you use in your work in writing a FAR report?
    b) Question 2 : Do you use specific data sources for specific agricultural commodities?
        i. What data sources would an expert use for a monthly report on dairy products? (PQ)
        ii. What data sources would an expert use for a monthly report on animal protein? (PQ)
3. **Topic 3: Aspects of their work (related to their interaction with textual sources) that experts perceive as important to be automated by a cognitive solution**
    a) Question 1 : How does an expert handle the available information found in the data sources to use it in writing a FAR report?
        i. Can you describe what information experts are usually experts looking for? (PQ)
        ii. Can you describe how the experts are usually looking into the data sources for the information they need? (PQ)
    b) Question 2 : Can you identify a task of an expert related to his interaction with textual data sources that is common and repetitive?

c) Question 3 : What task in experts' work would be in your opinion more beneficial to be automated and handled by a machine (i.e. something that would make your work easier and more efficient)?

   i. If you were to have a machine assist you to this task, what would you expect ideally to be the outcome? (PQ)

In the guide above abbreviation PQ is used to denote a probe question. Those questions are meant to be used to clarify the answer of the expert and assist to provide more depth in the interview. In that sense they will be used when deemed necessary and conditionally to the answers to the main questions. Along with the interview guide presented above we also recommend the following general structure for the interview:

Structure:

1. Introduce interviewer and build rapport with the interviewee
2. Provide the reason of the interview
3. Describe the main goals of the interview and its usability for the research
4. Determine the duration and provide justification for recording it (i.e. notes, voice recorder). Also allow for interviewee to object in recording it
5. Allow for questions before the start of the interview
6. Proceed with the main body of interview using the proposed guide
7. Express appreciation and allow for questions from the interviewee after the end of the interview

B.3 Employee interview details

	Job Role	Department	Gender
S1	Vendor Manager	Vendor Management	Female
S2	Service Delivery Manager	Cross Channel Service Desk	Male
S3	License Manager	Vendor Management	Male
S4	Innovation Manager	Human Resources	Male
S5	Business Analyst	Data Warehouse Distribution	Female
S6	Senior Buyer	-	Male
S7	Senior Jurist	Legal Retail	Female

Table B.1: Details of employees interviewed for player types in Rabobank

Appendix C

Interview transcriptions

C.1 Employees' interviews transcriptions

Interview Information

Date of the interview : *06-02-2017*
Place of the interview : *Croeselaan 18, 3521 CB Utrecht*
Duration of the interview : *00:15:56*

Participants

Interviewer : *Gregory Afentoulidis (GA)*
Gender : *Male*
Profession : *Graduate Research Intern*

Employee : *S1*
Gender : *Female*
Profession : *Vendor Manager*

Interview Transcript

1. Socializer player type

GA: Okay, so my first questions is...for your every day work how important is your social environment as a motivator for your every day tasks. I mean being able to share knowledge with your colleagues, to contribute to their well being and enjoying any group activities.

S1: I think that this is rather important, I am a contract and vendor manager so I try to maintain a relationship with our biggest vendors, and in this case it is not IBM but two different ones and I have to know what happens around the organization with those two vendors and I have to make sure that my colleagues doing business with them will inform me whether something is relevant there, so it is rather important to stay in touch, to stay connected to everyone. Yeah I was working with them so, I have a lot of different contexts with lot of different people around the organization being more or less in IT, but even outside I have some contacts even there. On the food and agri part, I have been incorporated in some actions around the use of microsoft environment for the portal there, there is a little contact sometimes...

GA: As I understand it, it is an inevitable part of your work in that sense. But what I would like to find out is...let's say that you are now interviewing for some new job that you want, how important would the social factor be inside that company for you.

S1: When I am working together with people you mean?

GA: Yes...

S1: Yes, that is absolutely important, and I enjoy working with people a lot and get my knowledge also from elsewhere and my network is very important part.

GA: So would you say this is a main motivator for your work?

S1: Yes, absolutely yes. I wouldn't be very happy in a single analytic job, doing it from home never contacting anybody.

GA: Ok.

2. Achiever player type

GA: So I would like to move on to our next topic. Often in a working environment appraisal of performance is shown with certain rewards. Either it can be something psychological or salary raise, bonuses etc. So what is your opinion on rewards? Should they come hand in hand with successful undertaking of work related tasks?

S1: Yes, I think so. I am rather result driven and I think it is important. It shouldn't always be money, that is not needed, because you get used to it. It is more important that you get recognition for what you do.

GA: But you can get in a sense an intrinsic motivation, that could be a good work from your manager...

S1: For example a chance to do something new or to develop a new skill in your working environment...can be various thinks.

GA: So would you say that you strive to find paths in your work that you would be able to earn some reward. Is this something that drives your motivation in selecting what to do?

S1: In some ways, it is not the only thing, but in some way it drives me absolutely.

3. Killer player type

GA: Another question is whether you think it is important to adhere to standardized processes and rules while you are conducting your work duties?

S1: More or less, but it is also important to challenge them and improve them and not always follow them. I am not what we call a 'process bunny'. So I do not enjoy always doing things in the same way, but I know it is important and I adhere to them as long as they fit the purpose.

GA: So let's say your are restricted by some rules in your work, would you say for yourself that you would try to adhere to them as much as possible or would you try to find a better way?

S1: Yes the last one most probably, I adhere to them as long as they fit to the goal I have and if they do not then I challenge them.

GA: Can you recall an example of that in your work?

S1: For example we have a very big risk assessment tooling for every movement in the cloud and it takes months and it does not cooperate with our targets to shorten our time to market. So then I try to figure out what is the possibility to don't fall out the excel for example because it has over 20 tabs you have to fill in, but be aware of the risks that might come out and try to make it in more efficient way for example.

GA: So is this an aspect of your character that you want to do this to make thinks work better?

S1: Yes.

GA: But do you think this is tied with your willingness to have good performance in your work or do you think this inherently in you?

S1: Absolutely. I do not accept that something is true because someone says so.

4. Explorer player type

GA: Another question...do you enjoy working alone?

S1: Sometimes it is nice because you can make some progress. I work at home half days a week and this sometimes make it easier to do some things because in the work you have appointments or someone is calling. But I think 2-3 days a week is enough because I think it is also important to interact with people for me.

GA: So if you had a task that you are perfectly capable of doing alone but you could choose to share it with somebody else and do it together, what would you think you would choose? And let's just say that the end result will be the same.

S1: Yeah that makes a difference, it depends on the results. Then I would look at it in an efficient way, can we finish faster together?

GA: Let's say the same amount of time, for some weird reason. So end result is the same and time is the same.

S1: In that case it depends if I like working with that person involved. And if that is true, then I would choose that one and if that is not true I would choose to do it by myself.

5. Dominant player type and generalization

GA: So, in a sense we have discussed characteristics of different types of employee type. They are not necessarily mutually exclusive. To summarize it, we have employees who are driven by their social environment and the interaction with it, we have employees who are reward oriented and strive for competence, we have employees who are out of the box thinkers in a sense and they are intrigued to questions the status quo, and employees who prefer to work independently and are curious to try or discover new things. So in which of these categories would you put yourself, if you had to select one?

S1: Yes, that is difficult. Because I have something from all of them, everyone does probably. So what do I prefer? I really enjoy figuring out new things, so something challenging.

GA: What is that drives you to select this?

S1: I would say that because if there is a new IT technology available I then figure out what it is. This is the reason why I said yes to this question, because I am curious what we can do with it and what value we can have. For example I am the cloud expert in our department for that reason just because I just started figuring it out first. Same for blockchain and stuff like that. But on the other hand I am still not working in the innovation department but as a vendor manager so there is a flip side for everything. But within my profession that is what intrigues me.

GA: And I guess this is one of the main reasons that you selected this profession. It is not something that evolved through your job.

S1: No probably, because I am business trained. I had a business degree and I ended up to this because of that. I changed several jobs, I have been a purchaser. There you have to do the hit and run part and when you finish it and someone is going to work with it and develop it then you are out of the office so this is the reason I changed to this job and be more in touch with the business there and the decisions.

GA: So in order to generalize a little bit, I am shifting the focus from you to the general Rabobank employee. Which one of the aforementioned categories that we described you think best matches an employee of Rabobank.

S1: Yes, that is a difficult part. I am not sure, because I see a lot out of the box thinkers leaving the company also and not only due to the reorganizations, but also because it takes to much time inside and they are willing to go.

GA: So there is a lack of effective communication you would say?

S1: No I think we are too busy with our day to day job to make the time to invest in communicating...If I had to choose the out of the box thinkers and innovation is very important but it does not come easily so If I had to choose I would say the social part. We have a lot of people talking around, we are very focused on our customers and there is a big informal network. So I think that would be the one to choose. But it is rather difficult because it also depends where you ask this. In the IT department we are more innovative, more independent from other parts of the organization. But if you take it as a whole and I have to choose between those four then I would choose the social part.

GA: And If you had to choose one of them to be necessary for new hires in the company, which one do you think is the most important?

S1: I think for now it is important to focus in the innovative part, discussing about the boundaries is important but we are getting more and more that we have to work with from the dutch national bank to the european cenrtal bank. So challenging them all the time will not be very efficient, we have to cope with them and we should have an efficient to work, or around them in some way. So I think the focus on innovation is important.

GA: I will try to quantify a little bit what we said so far. In terms of scale from 1 to 10 how much of a socializer would you say you are?

S1: Yes it is still important to get your contacts within the organization and keep in touch. I still think that should be a 7 or an 8.

GA: So how much reward oriented would you rate yourself to be?

S1: 6 or 7

GA: Out of the box thinker?

S1: Well I would say 6

GA: And as person who can work independently and try new things for themselves?

S1: I think that would be an 8

6. When and how will employees contribute in Enterprise Crowdsourcing?

GA: Last two questions. Let's say that you were to participate in this research by contributing data with a mobile application. So what time of the day would you ideally want to be able to contribute to this?

S1: Very early in the morning. 7 or 8 in the morning first thing when I get to the office. That is the time that I have available to do some things like this.

GA: So do you usually check out your mobile maybe in some of the breaks that you take duting a work day?

S1: Yes, I do it several times.

GA: Do you usually have one break? The lunch break?

S1: I usually have no breaks. That is by choice, I check my mobile on the go between appointments.

GA: And this the last question. How much time do you think you would be able to devote in this?

S1: A quarter of half an hour...

GA: Great. That was it thank you so much for the discussion.

Interview Information

Date of the interview : *06-02-2017*
Place of the interview : *Croeselaan 18, 3521 CB Utrecht*
Duration of the interview : *00:17:03*

Participants

Interviewer : *Gregory Afentoulidis (GA)*
Gender : *Male*
Profession : *Graduate Research Intern*

Employee : *S2*
Gender : *Male*
Profession : *Service Delivery Manager*

Interview Transcript

1. Socializer player type

GA: So the first question is, how important is your social environment and how much does it work as a motivator for your everyday tasks at work? So for example being able to share knowledge with your colleagues, contribute to their well being and join group activities in general.

S2: For me it is very important that I can share or that I can retrieve knowledge from my colleagues, without this I do not think I can do my job very well. You cannot know everything so you can ask your colleague or you can go to the internet or read a book or whatever. But additional information, additional knowledge is crucial to open up your eyes as wide as possible. And so with all this information, we can make for example a better decision for the customer or whatever.

GA: We see socializing and being able to share knowledge as an inherit part of the work, but let's just say we are in a scenario that you can do something on your own without communicating anything to anybody but you can also do it by communicating it with other people. What would you prefer in this case? And let's just say that the result can be the same but you have those two paths to choose in order to do that.

S2: I don't like to do it on my own, because I like to discuss with you for example, to discuss my ideas and my point of view. I am sure that you have a different

point of view that helps to complete the idea. If i lock myself in a room, I close the windows and I make this idea it will never be the same and I believe it will never be as good as compared to discussing this in a broader way.

GA: So would you say for yourself that being able to socialize within you working environment is a main motivation for your work?

S2: Yes it is a main motivator for my work. For example if I work at home, I miss this interaction so I prefer to be among my colleagues or I prefer to ask a stranger for example a colleague who I do not know but I know that he has expertise. This way he can feed me and we can discuss and it is important to do my work. If I have to do everything on my own then within a month...

2. Achiever player type

GA: So moving on, often in a working environment appraisal of performance is shown with certain types of rewards. It can be bonuses, raise of salary or something psychological like a good word for you or something related. Is it in your opinion rewards tightly connected to successful undertaking of work related tasks? Is it something really important to be there for your work?

S2: No, I believe if there is an intrinsic motivation then you do not need these rewards. Because asking me to think with me together, maybe I can help you also. I think you have to do something from within yourself and I do not need an extra salary or kudos or what their name is. For me it is not a game. It has to do with your intrinsic motivation and your passion and to understand why you just have to do this.

3. Killer player type

GA: Do you think it is important to adhere to standardized processes within your work, or rules that you need to adhere. Is that really important for you for your work? So for example certain processes around your work or protocols. Do you think that those are something that you should always adhere to?

S2: For me it is a guideline. Processes are important of course, a standard way of working etc. But I have to have the opportunity to step out of these processes.

GA: Ok, so let's talk with an example. Let's say that you need to do something for your work and you currently are restricted by some rules and you do not find them efficient enough to follow them. What would your option be in this case? An let us say that the result will be the same talking from the perspective of efficiency or the end quality of the result.

S2: For me it is very important if I change the way of working, so if I don't confirm to the process and I choose another way, does it add as much value to the customer as if I were to work by the process. And If I think my way of thinking is best for the customer, then I make this decision my own. I feel free to do this because I work in such an environment, I have a safe environment and if somebody says that this was not the right decision then okay we can talk about this. Give me

feedback i am alright with that, but at that moment this was what I decided. Of course you cannot do that with every process, but for simple things like I need to talk to 10 people to fill in forms if I can do it this in two steps then I choose for the two steps. The result is the same the effort is less and the customer is satisfied.

GA: So apart from that, would you say that this is something that you seek to do. So when you are thrown into something new when you are studying something new and there is a steep learning curve trying to familiarize yourself with things around you trying to learn new processes and so on. Is this something that you say ok...you haven't really analyzed the efficiency yet, but you are like I might have something in mind that might drive me to choose it but if I follow the processes they are standard if I follow them then the result will be the same. In an essence do you seek the comfort of standardized processes or do you like to always discover something new?

S2: Well it is something from both worlds. I try to work as much as possible within the processes but I try to look always through my lean glasses to see is it possible to skip one step or do something different so the time will be shortened or in general if I can skip a few steps. This is something I look always if there is this possibility. I will work by conforming with the processes but I will also think on my own. Something may be written five years ago, and time changes very fast so we might be able to do it better, and everything better for the customer not for my own purpose but for the customer.

4. Explorer player type

GA: We kind of already covered this I think but allow me to reformulate it. Do you generally enjoy working independently or you always seek the company of somebody else? So posed maybe a little differently. If you work independently usually you can choose your own path but if you are working with a group then maybe you are able to do that again but always as part of the group. So what would you rather choose when you are faced with something you need to do and you have these two options?

S2: I think a combination. Many times I like doing things on my own but most times I like doing them as part of a team. Because I think if you do things as a member of a team you can grow.

5. Dominant player type and generalization

GA: So thus far, you may have noticed it, we discussed several characteristics of employees, not necessarily mutually exclusive. So to summarize them for you we have employees that are mainly driven by their social environment and the interaction with it, then we have employees who are reward oriented and they strive for competence, we have employees who are out of the box thinkers in a sense and they always want to challenge the status quo, and then we have employees who are preferring to work independently and are curious to find new ways to do things. So looking at those categories if you were to decide which

type of employee you would be what would that be? I mean most probably you have the qualities from each category, this is obvious, but if you were to select only one what would that be? Which gathers the most qualities suitable for me?

S2: I am always looking to challenge the status quo.

GA: And in order to generalize a bit. For the same categories, in which of them do you think the average Rabobank employee suits best?

S2: Wow, that is a difficult question. I think that most of the employees of the bank are trying to work within the processes, within predefined processes. This is what I think but also what I see. I do not know why but maybe because we had a big reorganization and I really saw that people were drawing back and waited to see what happens and they were just doing their work. Now this big change has fulfilled and everybody is in his position and everybody is settled down, almost I hope. We like to work with high performance teams, self organizing teams, agile, scrum all these terms you know, I really hope that more people will look for this challenge to operate not only inside specific lines but try to step out of the box. Because I think we can do things much much better, more efficiently by working together but looking closely to every process. Make this part of our DNA, to challenge yourself to find if we can do something in a better way to have a better customer experience or put less effort in this.

GA: So let us put you in a position of a hiring manager. You have to hire people for the company and you have to base your decisions on those criteria mentioned before. What do you think is mostly needed for Rabobank now?

S2: I like a team with employees with combination of those qualities. But I like then to share knowledge in a natural way, to know that this is part of their job and that they understand that they cannot do everything alone and that by sharing knowledge it is not only for the receiver but also for the sender. This makes you grow. The perfect team, there is no perfect team so you have always a combination of number one,two, three...

6. When and how will employees contribute in Enterprise Crowdsourcing?

GA: Now let us say that you use the mobile crowdsourcing application and you contribute data, what do you reckon would be the ideal time of the day in which you could contribute?

S2: Early in the morning, between 6-8AM because this is when I am sitting in the train. Or after 4:30PM because in the meantime it is difficult.

GA: And also how much time do you think you would be able to devote for that?

S2: If I can contribute something to something bigger and I can do it by reading something every day that is 5 or 10 minutes maximum then no problem.

GA: Ok, this was my last question. Thank you so much.

Interview Information

Date of the interview : *08-02-2017*
Place of the interview : *Croeselaan 18, 3521 CB Utrecht*
Duration of the interview : *00:24:10*

Participants

Interviewer : *Gregory Afentoulidis (GA)*
Gender : *Male*
Profession : *Graduate Research Intern*

Employee : *S3*
Gender : *Male*
Profession : *License Manager*

Interview Transcript

1. Introduction

GA: Before we start can you please introduce yourself? What is your job role within Rabobank?

S3: My job is license manage within the department of vendor and contract management. Our department is responsible for the contracts with all the big and small vendors. Big clients such as IBM or Oracle and Microsoft. So my position is license manager and I am responsible to...so we have contracts all over the world like in Rabobank Netherlands and Rabobank Australia or Brazil for example and everywhere we use IBM software. There are many people involved there especially for IBM software which is my specialty. So the big thing is that we have a lot of licenses and we must measure if we are compliant to our vendor and we measure all the time in our data centers that the license consumption is in balance with the licenses that we have and that weâĂę for example IBM has contracts with many rules and we must adhere to that and we must check it within the organization...there is an audit and compliance and we must check that we are compliant to IBM and to Oracle and to Microsoft. The big thing in the software license world is audits from the vendor, that is an easy way to make money.

GA: I guess from governments also?

S3: No, well less the most audits come from the vendors who think they can earn money from offering more. Oracle is famous, well famous in not the right word...they say let me see what do you use...aha may I propose something else?

2. Socializer player type

GA: So I am now moving on to the first part of this discussion. I would like to discuss with you some personal characteristics of you in the work environment. So how important is your social environment in your work as a motivator for your everyday tasks? So being able to share knowledge, to contribute to the well being of your peers or colleagues or enjoy in general group activities?

S3: This is important I think. You mean my peers are the clients and the people that I have contact within the organization right?

GA: Yes.

S3: Well this is important for me.

GA: Ok I would like not to tie this too much to your work rather I am interest if this is coming as a result of your personal interest. Let us say for example you are applying for a new job how important would the social environment be to you to select to take this job role?

S3: Well, it is very important. I have a lot of contacts with delivery managers with service owners, the social network and contacts is really important, to know the way to measure something to speak with someone who know something about it. For me it is important to know my way into the organization and to know people.

GA: So it is also something it is inherit in your character you would say? It is not only a necessary part of your work by it is something that you seek for?

S3: Yes. I work for 25 years, or more than 25 years for this organization and I know a lot of people I have worked in differrent departments. The network I have created the past 25 years is really important to me. So if I need for example to know something about a data center I know people and I know who to ask.

GA: So how important would you rate is for your efficiency in your work to be part of a team and work together and interact?

S3: Can you please repeat the question?

GA: So it is more or less questioning whether you are trying to do something on your own or is it always that you like being in the presence of your colleagues trying to have your everyday tasks worked out like that?

S3: I mix of both I think. Part of my job I can do alone. It is a dedicated environment and I can do a lot of things on my own I do not need my colleagues to do my primary job. But for a lot of things a need other people and I need colleagues to understand for example why do we buy this and why talk to this way to a vendor

and what happened there. For this kind of things I need my colleagues and I need interaction with them.

GA: Let us work with an example. Let's say you have to work on a specific task and you have the option of doing it by yourself or with a company of a peer and the result will be the same regarding the quality of the output. In this scenario what would you choose between the two options?

S3: It is not possible to do my work without other people. I depend on the input of other people. Purely alone it is not possible to do my work completely isolated. A big part of my work depends on the input and interaction with other people.

GA: So working alone it is not even an option for you.

S3: True it is not even an option.

3. Achiever player type

GA: It is usual in a working environment that appraisal of performance is shown with certain types of rewards either it can be bonuses in your salary or some good words from your managers or your peers. Should in your opinion rewards come always hand in hand with successful undertaking of work related tasks.

S3: This is an interesting question. For me the most important rewards are...and I like it and I feel good when they say this a good relation, i think you contribute to the subject we are busy on and you are doing your job good. That is the most important reward for me. When you want something for IBM licenses you need to go to him [he means himself]. I think this is the most important reward that you can get from other persons.

GA: So would you say that rewards are not one of the main motivators for your work, and I mean extrinsic rewards. They do not drive your work in a sense that you do not always strive to find the rewards that you could get in your working environment?

S3: Sorry can you please repeat this?

GA: Do you think rewards are one of the main motivators for your work, as we discussed them so far?

S3: Yes I think for every person and a professional it is important to get rewards. It is not necessary to be friends with everybody. Sometimes you do something and it is wrong and you need to do it in another way. You are not always friends.Not problem this is part of the job. If you have a relation, and you do something wrong but we are talking in a good way...Okay I am not happy with it, but if you are right and we solve it then this is a kind of reward.

GA: So there are intrinsic rewards, which are related to your intrinsic motivation for example someone saying a good word for you appraising your work and extrinsic rewards such as raise to salary, bonuses or a better position. What is more important for you between those two?

S3: I would say the colleagues and the bonuses...well I have a good salary from Rabobank and that is ok. Well if my boss says you get more money that is also ok. But that is not my primary concern.

4. Killer player type

GA: Do you think it is important to adhere to standardized processes and rules while conducting your work duties? So I guess that in most parts of your work there are certain processes and predefined steps that you need to follow, do you think that it is always important to adhere to them?

S3: Yes, I think it is important but in my role we need more policies and standard way of working. The problem is that often compliance is a result of not making good agreements and not following procedures. When you say hey I have a new server I set them there and you do not think of our licenses then you have a problem. For us it is important to make good policies and good rules. When you buy a new server it is important to buy licenses and you must think of licenses. For us it is important to follow patterns and there is a lack of that. Important part of our job is to make it transparent and make it easy and make the procedures okay to do what you want.

GA: In the case where you are restricted to your work by certain rules. Would you say that you would try to adhere to them as much as possible? Would you try to find a solution within a standardized process or would you go about finding a better way that might alter the whole process.

S3: Tricky one. Part of my work is we have a contract and we have the rules and we have to fulfill the contract. But also part of my job is to do that in a good way to try to do it in a way that is good for Rabobank and also good for the vendor. So in a way it is my job to do it in a creative way. I try, if for example there is a rule from the vendor that it is not good for us then I try to fulfill the rule and this is good for Rabobank.

GA: So to summarize this topic if I may. As much as possible adhere to the processes and the rules because those are the driving guidelines but be able also to see through them and see any hiccups that might arise and maybe make it better within your capabilities.

S3: Yes.

5. Explorer player type

GA: This is something we might have already discussed, but I want to reformulate the question. Do you generally enjoy working independently and being able to follow your own path?

S3: It is necessary and I do it with pleasure to work with the people in the IT. But I also like it, to say I must make this project and I can do it on my own.

GA: So there is the main path of your work and maybe within this main path certain opportunities to collaborate with somebody else. So do something that is not immediately related to your work. Is this something that you would do eagerly?

S3: Sure, yes sure.

GA: So when you are given the opportunity to diverge from your standard working path you would do it, if that does not distract you from your main work right?

S3: Yes, that is true.

6. Dominant player type and generalization

GA: So so far we have discussed so far several employee qualities, as you may have noticed. Those are not necessarily mutually exclusive, meaning that qualities that you might find in a certain type you can also find in another type as well. To summarize them for you, we talked about employees who are mainly driven by their social environment and the interaction with it, we talked about employees who are reward oriented and who strive for competence. We also talked about employees who are out of the box thinkers and who want to challenge the status quo and finally we talked about employees who prefer to work independently and are curious to try new things. So if you were to choose one of these categories to put yourself into it, which one of those would it be? Again, you might recognize characteristics in yourself from all of these categories but let us say that you select one that represents you for example 80

S3: Yes, that is difficult.

GA: We can also rate those to make it easier if you prefer? For each of those categories give a number from 1 to 10.

S3: Yes.

GA: So who much would you rate yourself being dependent on your social environment and the interaction with it?

S3: 8

GA: How much would you say that you are reward oriented, so that you strive fir work that entails some reward for you?

S3: Less...5-6..Maybe 6

GA: How would you rate yourself as somebody who is intrigued to challenge the status quo? So as much as possible try not to adhere to standard processes and try to make them better?

S3: Yes I think this is also a 7-8 I think

GA: And finally how much would you say you prefer to try new things, to work independently and deviate from your main course of your work?

S3: Less I think. 6 I think.

GA: Now I would like to generalize this to also other employees of Rabobank. Which of the aforementioned categories do you think best applies to the average Rabobank employee, If there is such an employee?

S3: I think there is not.

GA: I understand that, but I want to know mainly from your experience working all those years and interacting with so many people, which one do you reckon is the most characteristic?

S3: I think those who are challenging the status quo...but this is difficult. There is no average, it depends on the function level, it depends whether commercial or technical and it depends if he is an architect or an engineer or in a call center. It is really wide.

GA: What if we were to put you in a hiring manager position which of those qualities is the most important you think?

S3: The socializing I think.

7. When and how will employees contribute in Enterprise Crowdsourcing?

GA: Let's say that you wanted to participate in this research and you could annotate data using this mobile application. What time of the day you think ideally you would be able to contribute to this?

S3: I am an early starter and my day ends at 4 o'clock. My best time is early in the morning. And I start at 7:30am.

GA: And also how much time you think you would be able to devote each time you decide to use the application?

S3: That is difficult to say. I cannot imagine how it works.

GA: You can answer this by taking into account how much time you devote into interacting with your mobile device and then calculating a portion of that for our purposes if you are interested.

S3: 10 minutes then or a quarter of an hour, but I cannot really say since I cannot imagine how this works.

GA: Ok that concludes our discussion. Thank you very much.

Interview Information

Date of the interview : *08-02-2017*
Place of the interview : *Croeselaan 18, 3521 CB Utrecht*
Duration of the interview : *00:34:18*

Participants

Interviewer : *Gregory Afentoulidis (GA)*
Gender : *Male*
Profession : *Graduate Research Intern*

Employee : *S4*
Gender : *Male*
Profession : *Innovation Manager*

Interview Transcript

1. Introduction

GA: Let us start by telling me something about yourself. A small introduction.

S4: My name [he states his full name].I have been working for Rabobank for 7 years now. So I am a bit of novice in terms of Rabobank perspective, people work here for many many years. I have a background in educational technology and have been working as a consultant, trainer and project manager for many years in different companies and joined Rabobank 7 years ago as an innovation manager for the development area. Since a couple of months now I am more involved with organizational development. So subjects like how Artificial Intelligence might help the organization to get to higher performances, the use of Big Data for next level search engines and all those types of possibilities are also subjects that I am familiar with.

2. Socializer player type

GA: So as a first topic in our discussion, I would like to know how important is for you your social environment for you as a motivator in doing your everyday tasks? Like being able to share knowledge, to contribute to the well being of others and also being part of a team.

S4: It is a huge factor in your motivation I think. Based on what we know this is one of the main areas...no, just let me give my personal view on that. For me personally it is very important.

GA: I am more interest on how important this is for you outside of the context of your work duties. Obviously in the working environment you have to communicate with somebody, but the question is mainly focused to understand if this Is something that you seek for. For example you work in environments that fulfill your ambitions but they are different in terms of the social environment in which one would you rather be.

S4: For me, I am a bit of a typical worker for Rabobank. My work is very much knowledge work, it is a type of R&D type of role. So it is my explicit role to find knowledge with others that might help us get some new viewpoints in what we are doing here. So my network is a typical for an HR person. I have many many contacts in many different departments ranging from Rabobank researrch to innovation to local banks. So for me I could not do my work if I wasn't networking with different teams and different social networks to actually get that done. So in terms of my job it is important and in terms of my motivation is very important. I have been doing a lot of work almost semisoloistically and this is not good for your motivation, you need to see faces. So an interesting example is for instance I have been working on a couple of reports and some time you really need to focus on the reports. So I went home to start writing and is like start writing a book. And after two or three days I was like ok now I want to go back to the office again just to get updated with the input from others. This is not necessarily that I was seeking it, but being able to be physically in this environment and being able to overhear colleagues talking on certain subjects...the element of surprise if a very important element. Things come to you because you overhear something and you are like ah this was the missing part of my puzzle now I can finish it.

GA: So in a sense in a social environment you can have two forms of interaction with your colleagues. It can be collaborative or competitive. Which one do you think is the most important?

S4: Collaborative absolutely. In my field there is hardly any kind of competition. There might be only a few competition. So it depends, from an HR perspective if we work for the people of the company we work for clients. So for me personally is very much collaboration.

GA: In a sense you already answered also my follow up question which is how important is for the efficiency of your work to be part of the team. So how effective do you find yourself be as part of a team rather working alone.

S4: I think there are two aspects there. There is the social network that you operate in which is very important, but also the team. It might not be exactly the same right? So my social network is vast inside and outside of the organization but sometimes it is really really fun and good to have a small compact team and pick up a task and go for it. So somewhere there is a balance between doing that for some part of the week and being in another social context.

3. Achiever player type

GA: Moving on to the second topic. Often in a working environment appraisal of performance is shown with certain types of rewards. Either it can be that you get a good word from your manager or a colleague or it can be something extrinsic like a bonus to your salary. So should In your opinion rewards come always hand in hand with performance in your work?

S4: How do I look at that? One of the things that we have found in the organizational health index that we had just a couple of months ago and which I also feel myself as being true is that we don't reward people that do exemplary good work and we do not tell people who do mediocre work to step up. So whatever you do it is okay, and we need to change that. We do not have a reward system in place. I am a firm non believer in external rewards but you know just giving the stage to somebody because we believe that they have done a great job and tell us about it is important.

GA: So recommend him in a way, put a good word for someone.

S4: Yes this is it.

GA: So you really think this should be attached with your performance in your work?

S4: Yes, yes certainly. And I believe the whole redesign of our performance management is based around this principle. That you need short samples of feedback, being able to address things that happened last week or the week before while they are still fresh and you can talk about that and the things that you really do good you make bigger. And another thing, based on the things that you are good at you are trying to compensate for the things you are not so good. So, I do not know how much you know about are performance management system, but it has changed 180 degrees from the 1st of January compared to what we did before and we still need to work on that.

GA: So taking some pieces of your answer, I think, and correct me if my wrong, you put this in the social context again or not? That is because you said I am firmly opposite to the extrinsic rewards, so the rewarding system should be related to your social context.

S4: Yes absolutely. But there is also a cultural thing that is going on here as well that you need to be aware of. There was a research about a project that was called finding without searching and one of the ideas was that we could use big data, our internal big data to actually trigger information for people that might be able to actually use it in their work. So for instance is somebody was preparing a case for a client there would be a trigger, that would say okay this might be helpful for you and people really enjoyed that. They said you know that is great it will save so much time it would allow be to have a quick chat with somebody who really knows or had the same case just last week and it would save me a lot of time and you know my performance would go up. And the moment they say it, they are like ok wait a second, who is that person that pops up in my screen? This is somebody who did a case similar to yours [he means they reply to the one posing this question]. So it could be a colleague? Sure. So it could be me? Yes,

sure. So you are saying to me that if I am really good at something and I have expertise then people will see my face and they will start asking me questions. Yes it could be. And then their responses were like, and this is the think that I was talking about, so I do not have time to do my work any more. Question of course is, what is your work? Where is your main performance located? Is that in your task description that you have? And you know your manager says you have to have so many talks with clients and you have to have a conversion rate like that or allowing your expertise to flow within your organization and help others to reach high performance? This is a cultural thing that we need to work with.

GA: So is this something that in your opinion missing? This collaboration in a sense?

S4: I think deep down people would love to do that but the system is rewarding them for other things than that. So it is a systemic thing within our organization, that we drive for certain performance measurements but we do not drive for enough health ((incomprehensible)) . Now we understood that this is the case and we are trying to change that, but this is one of the areas where we need to make a shift in.

4. Killer player type

GA: So moving on to our next topic. Do you think it is important to adhere to standardized processes and rules in your work? So I guess in your work your are sometimes faced with such processes that you have to follow and that you have to comply. Is that something that you think it is really important to adhere to?

S4: For me no not at all. But that is just me, of course there are other colleagues here in the HR function who do work in a more procedural working environment. But I think if you look at it from a distance you will see that the work that is being done here in the central HR can be characterized as non-routine analytic work. We do have a set of routine analytic work but this is done in a different part of the organization which is HR support. So in terms of replicating the same procedure continuously I think that HR support is a more interesting environment. So that is also an environment where AI solution might be more appropriate.

GA: In the case you are restricted by some rules in your working environment, would you rather try to follow the standardized processes as much as possible, to your best of your capability? Or would you be the one breaking the loop in a sense trying to find an alternative?

S4: For me it would be breaking the loop. But it depends on whether my analysis of the future situation and where we are heading is congruent with the direction that is being given.

GA: Let us say that the results are the same.

S4: I don't have any problem in working inside a specific context that says this is our vision, this is our strategy this is what we are going to do and do your work. But it is my job to understand if that is a valid course and if it is not then it is my job

to start giving ((incomprehensible)) , and try to get other people to understand that there are other ways. So I am not only the legitimized but I am also asked to look into things in a different way.

5. Explorer player type

GA: Do you generally enjoy working independently and being able to have your own path? So to formulate It a little better, I guess there I a main path that you follow while conducting your work, and in this path you might get the opportunity to diverge a bit and it might not be closely related to your work. Would you take those opportunities?

S4: Yes sure. For me that is a no-brainer...But on the other side of the story it is sometimes very helpful to have a baseline set of agreements that you work with. For instance I think we are still hindered by the fact that we still the old RI and RN type of IT environments. It is just not that helpful to have two of them so if we integrated that it would be really good. Because now part of my job is to disseminate knowledge and disseminate information to other people and there are just so many buckets that I have to fill to actually get my message across and that is not very helpful. So I think we are making a good progress there but I also think that there are more to be done there. So there is a balance there again if those lower level things that people need to adhere to are okay and they work then it is easier to diverge to other areas. Now sometimes you need to diverge even for the basic things and that is not good.

6. Dominant player type and generalization

GA: So far, as you may have noticed, we kind of discussed several employee categories and their characteristics. So to summarize it for you, we talked about employees who are mainly driven by their social environment and their interaction with it. We talked about employees who are reward oriented and strive for competence. We also talked about employees who are out of the box thinkers in a sense and they like to question the status quo and finally employees who like to work independently and follow their own path. So based on these categories, and keep in mind that those are not mutually exclusive and that you can find that you fit to either of them, which one would you select to put yourself into? Also if that is difficult to answer then we can also rate them by providing a number of 1 to 10 for each of them.

S4: Yes let us do that, it is interesting.

GA: Okay so from a scale from 1 to 10, how would you rate yourself being mainly motivated by your social environment and the interaction with it?

S4: I would say 6.

GA: Reward oriented and strive for competence?

S4: 4...well strive for competence I would say 8.It depends on what you define by reward, because my reward is to strive for competence.

GA: Out of the box thinker, and as an employee who is disruptive to the formal procedures?

S4: 9

GA: And employees who want to work independently and want to learn new things?

S4: 7 or 8

GA: Now to generalize this, if we could say that there is an average Rabobank employee, in which category would you think you would put him in.

S4: Why are you asking this question? What will you find out with this question? How the status now is or what we should strive for?

GA: That would be the follow up question actually. Now we are discussing about the current status and then I would like to ask you how it should be.

S4: Right now I think, I think that there is a huge difference if you work for a local bank or the central organization. So in terms of local bank, I think many people are driven by reward and by working in a set frame of reference, but I think that this is much less the case here.

GA: So they work to get as much out of it as possible?

S4: Yes. Do not look to much out of your boundaries.

GA: So if you can do your work and get the appraisal for that then it will be ok.

S4: Yes, then you are really good. So I think for big part of our population that this is the case. Too many people here in the central are in the other side of the...you know the flip side. I think we have too many disruptors or free minds in the organization that we have the tendency to reinvent things that have already been invented. We should take time to discuss things that have been already decided. So I think we could do with a little less of that. But again here is that the thing is not a personal trait of people rather than a systemic one.

7. When and how will employees contribute in Enterprise Crowdsourcing?

GA: Okay now I would like to focus on questions specific for this research. Let us say that you want to contribute in this research by using your mobile device. What time of the day would you be able to do that ideally?

S4: I think travel time would be a good time.

GA: So this is early in the morning I guess andâĂę

S4: Yes early in the morning and later in the afternoon or the evening. Also I think it is a typical Friday thing. I do not know it feels like a Friday thing. Well Friday is my day off, but it is one of those things that if I could get some questions, something that I could do in my mobile phone even in my free day then it

would be fine. You know I will not have the idea that my boss wants something from me or something like that. Yes I think that would be probably preferable, maybe even weekends. And if it was some type of game or gamification type of environment then that might help.

GA: Finally how much time would you say you would be able to devote for that. Usually in crowdsourcing the tasks that you have to do are some minutes. You can try to answer this by relating also to the time that you in generally spend using your mobile phone.

S4: Again it depends on how much you are immersed in the application. I think that there are some cues just like in Flipboard that keep me going back to them to get even more news.

GA: What are those cues?

S4: I think it is because I am content driven. I think because there is individualization of the content and things that are sent out to me are relevant to me. They are also small enough so I can flip through them and go into a little bit more depth and read them. This pattern keeps you surprised about new things that are coming and I think there is the hook in Flipboard. I am not sure if this Is relevant but maybe that element of gamification is important. Something like you score another 10 points.

GA: And if we were to apply gamification what would be in your opinion the ideal one to choose for?

S4: I think it would be into two levels. The first level would be to understand the net worth that I would have given back to the system with some kind of a score. The second one I think, personally, I would really find it interesting to understand that I am part of a bigger picture and that bigger picture counts for something.

GA: So in a sense an application that works in collaborative way would be ideal?

S4: Yes. It is like run keeper. You have your own scores but sometimes you get a message that says hey you are one of the 5.380 people who actually done 10 km run this day. Sort of gives you the sense that you are part of something bigger and is actually going to contribute for what we are doing for Rabobank or our clients. So it is also my accomplishment but also the accomplishment of the group and keeps you coming back because you feel part of that movement.

GA: Thank you for a really nice discussion.

Interview Information

Date of the interview : *08-02-2017*
Place of the interview : *Croeselaan 18, 3521 CB Utrecht*
Duration of the interview : *00:20:24*

Participants

Interviewer : *Gregory Afentoulidis (GA)*
Gender : *Male*
Profession : *Graduate Research Intern*

Employee : *S5*
Gender : *Female*
Profession : *Business Analyst*

Interview Transcript

1. Introduction

GA: Hi, so before we delve into any specific questions, can you please introduce yourself a little bit. So what are your job duties in Rabobank, what is that you do daily?

S5: I am...I am a data analyst at the data warehouse of distribution of Rabobank. I work a lot with data so I find this very interesting. I am 27, I live in...eh do you want that kind of information?

GA: No, no that is okay.

S5: Okay then that is about it, I guess.

2. Socializer player type

GA: Okay, so let us move to the first question. So how important is the social environment of your work as a motivator for your everyday tasks? So for example how important is for you to be able to share knowledge with your colleagues, to contribute to their well-being or in general enjoy group activities?

S5: Oh yes, for me this is really important, I have a lot of colleagues and I am at the analyst side of the job. It is been a while since I have spoken English so sorry for that.

GA: No, problem.

S5: And we have ETL processes and that is the flow of the data and also the people that put the stuff in production. So we get the data from ((incomprehensible)) , which is one of our main sources amd after that we create RTV models and then we make a ((incomprehensible)) for the people who use that data.

GA: So would you say that in this pipeline, being able to communicate is really important?

S5: Yes that is really important, it is far better to question someone rather than go to your computer and search by yourself in a very long documentation or e-mails.

GA: Ok, so I understand that big part of your job is to be able to communicate with people and being able to exchange ideas and maybe discuss the workflow that you have to do. Now to give an example in which you are facing two job opportunities, and regarding your personal ambitions those are the same, but there is a difference between the social environment between the two. Let us say in the first you could more easily mingle with the people and match with them and in the other one it is not like it is much different but it is offering a more personalized path of work and you just do your daily work. What would you select between them and why would you select it?

S5: The first one I guess, I am more of a social person and this is why I am a business analyst, I need to talk with the IT and also the business, because I am the person in between and this who I am and what I love doing. And why? Well it makes the days go faster.

GA: Ok, so would you say that this is something that really affects your efficiency in your work? So for example being able to work in a team makes a difference on your work efficiency rather than working alone?

S5: Yes, sometimes I like being alone so that I can focus on one thing without getting distracted. But you also need some guidance. You cannot focus for 8 hours straight, you have to have some distraction to generate some new ideas.

3. Achiever player type

GA: Ok, so moving on to our next topic. Often in a working environment appraisal of the performance of the employees is shown by some kind of rewards. This can be from a bonus to your salary or a raise, to just a good word from a colleague or your manager. Do you think, in your opinion, that rewards should always come hand in hand with successful undertaking of work related tasks? So do you think that your job should be always connected with some kind of reward at the end?

S5: Yes, I think so, there are different kind of rewards. Some people build a product and just to see that product working can be a big reward, to see the product be successful or seeing people being positive about the product. There is also another reward more social, people like working with you are enthusiastic about you.

GA: So do you think that rewards is really important to be tightly connected to your performance?

S5: Sometimes, but I also believe in a very positive way to motivate people, to keep their energy and keep them enthusiastic and passionate about what they are doing. So sometimes you see people losing their energy and I think it is really important to highlight to them the positive things and motivate them to climb up again and make great products again and get inspiration.

GA: So if you were to select between two kinds of rewards, one being extrinsic, like bonuses or a raise in the salary and the one getting acknowledge by your peers which one would you regard as the most important?

S5: I think what people think and say about you is most important than money.

4. Killer player type

GA: Also do you think it is important to adhere to standardized processes and rules while conducting your work?

S5: Again, please?

GA: Yes, so I guess while you are working there are sometimes in which you are faced with some standardized processes that you might need to follow or some kind of rules that you need to adhere to. Do you think it is always important to follow these rules as much possible?

S5: Of course it is important to follow those rules as much as possible but there is always and exception and this is why we are human and some task can be done by machines. So you need to use your mind, and your feelings and experience to make the right decision.

GA: So if I understand correctly you think you need to adhere to standardized processes as much as possible but when you also deem this is necessary you need to put your mind into work and maybe choose something alternative?

S5: Yes, I do no think I understand this correctly, can you please repeat it one more time?

GA: So we are talking about potential situations in your work where you might need to follow some predefined steps in order to do something. If now you are faced with something like that it is always important for you to follow this steps. So for example if your manager requests you to do something and proposes certain steps but you are also faced with an alternative to do that what would you rather do? And let us say that the result would be the same.So the quality of your work does not depend on that, but you are given a closed form way to do that and you are also given an opportunity to do it somewhat else, in a different way. Would you go for the different way? This is sort of the question in a sense.

S5: I would go for the alternative, if this is way to do things faster.

GA: It might not be necessarily faster, it is just a new way of doing things. So the efficiency of doing the specific work remains intact and the result is the same. You are just given in one hand the opportunity to follow something while on the other hand you can improvise or break the loop in a sense.

S5: Ah okay, yeah I would go for breaking the loop I guess. What I am thinking now is that for example we have power designer and we have a certain way of working with the power designer and I am not always following that and I am just clicking and seeing what works instead of following the manual. The result is the same in the end, we have the same model. Well, I am just referring to this now to give an example.

GA: So you just do this to learn for yourself to try something new.

S5: Yeah exactly. To exactly understand what is happening, in order to have this understanding you need to explore.

5. Explorer player type

GA: So do you also in general enjoy working independently and being able to follow your own path? So let us say in your career you have your everyday working path. You do what is it you need to do and you deliver what you need to deliver, but in this working path you might be given the opportunity to do something else that is not quite your main concern, or the opportunity to work into something else. So do you take those opportunities to diverge a little bit or are you like this is my job I have to do it and I am perfectly fine of just doing this?

S5: Well, for now because I have only been working for Rabobank for one year, I would say that I am focusing on my job and maybe on the side I do some other stuff, but It is not that I am finding my own way, I just do my job. And when I get the sense of trust with my environment and everything that I work with, then I tend to go a little bit more outside.

GA: So you are mainly focused on what you are doing and if you have the luxury to check on something else then it is not necessarily that you would reject it, but you have to first be ok with yourself and your primary responsibilities.

S5: Yes, exactly. I do not have a lot of working experience so this is why I am focusing on this one first, to be more specialized to what I am doing.

6. Dominant player type and generalization

GA: So we have discussed in a sense so far different categories of employees and different qualities that they may have. It is not that they are necessarily mutually exclusive and the are characteristics in one category that you might be able to find also in another one. So I will reiterate what we have been talking thus far. So we have been talking about employees who are mainly driven by their social environment and the interaction with it. We talked about employees who are reward oriented and strive for competence. Employees who are out of the box thinkers and are intrigued to question the status quo in a sense and finally employees who rather work independently and have their own ways of doing stuff and they like finding new ways. So I guess who may have qualities from all these groups. If you had to choose one of those groups as the most representantive of yourself which one would it be?

S5: I think I am more of a social person I would say.

GA: We can do it in another way if this is preferred. I can repeat the groups and you could rate them from 1 to 10 to denote how much you think they express yourself.

S5: Yeah let us do that.

GA: Okay so from 1 to 10 how much would you rate yourself as being driven by your social environment and the interaction with it?

S5: I think 8.

GA: How much would you say that you are reward oriented and strive for competence?

S5: I think 7

GA: How would you rate yourself as being out of the box thinker and intrigued to question the status quo?

S5: I think 7 or 8.

GA: And how much would you rate yourself as preferring to work independently and follow your own path?

S5: Yes, for now 6

GA: And now I would like to generalize this to all the employees of Rabobank based on your experience. If there is the average Rabobank employee, I which of the previously mentioned categories would you put him in? Which one expresses better the Rabobank population?

S5: I think the first.

GA: Socializers that is?

S5: Yes.

GA: And why would you select this one?

S5: Because there are a lot of meetings going on, always. People are very social here in general.

GA: So is it from your experience so far that people here are really collaborative?

S5: Yes, exactly. If I have a question they always have time and they are very friendly. If I want to drink a coffee and discuss something they always have time.

7. When and how will employees contribute in Enterprise Crowdsourcing?

GA: So now if you were to participate in this research by using the mobile application to annotate tasks so we can train our algorithms. What time of the day would it be for you ideal to contribute for this cause?

S5: So you mean when I have time to open the app and use it?

GA: Yes.

S5: Well at lunch, I think. I think that is the best time.

GA: And do you usually have one break during your day?

S5: Yes, mostly yes

GA: And finally how much time you think you would be able to devote in this?

S5: I would say 15 minutes a day, or something like that. Maybe not only at the lunch time.

GA: Okay, great! Thank you so much for this discussion.

Interview Information

Date of the interview : *15-02-2017*
Place of the interview : *Croeselaan 18, 3521 CB Utrecht*
Duration of the interview : *00:20:09*

Participants

Interviewer : *Gregory Afentoulidis (GA)*
Gender : *Male*
Profession : *Graduate Research Intern*

Employee : *S6*
Gender : *Male*
Profession : *Senior Buyer*

Interview Transcript

1. Introduction

GA: Before we begin, I would like some things about yourself, so a small introduction for you.

S6: I am responsible for the procurement part that we do with IT with our biggest vendors like IBM, Oracle and HP, SAP and a large number of suppliers. And everything is based on infrastructure as I said so everything is underneath the application layer it is my responsibility to come up with strategies, to make a certain negotiation plan but also legal perspectives that have to be arranged in contracts etc.What I do mainly is I am in projects and I control a whole team with people within Rabobank, and we work with specific timelines and measures and milestones. And I have a team responsible of informing and advising the senior management team within Rabobank, who have to make decisions. Once the negotiations have ended and the contracts are signed, then that is when usually my contribution ends for these projects and then I get some new projects.

2. Socializer player type

GA: Okay, to kick-start our discussion I want to introduce our first topic. How important is for you the social environment at work as a motivator for your everyday tasks? So for example to be able to share knowledge, to contribute to the well-being of others and generally enjoying group activities. And if possible try to see this outside of what is necessary for your work.

S6: Oh, that is a difficult question to answer. Of course you want to be successful and I like doing this work because you can put great effort in it and you can always come back with a great result and that is what drives me. To do the best things for the bank or what we call risk mitigation the best price performance and also motivate other people to support your plan etc. And for the rest I keep private and work pretty much separated some times I try to make people understand what I do for my living but they do not really understand a lot about it, and that works find with me.

GA: So how important is for the efficiency of your work to be able to communicate with others?

S6: Vital.

GA: In what sense? How would you imagine yourself without being able to communicate with so many people?

S6: I could not do anything at all. Because you need people in order to get requirements and discuss the risk factors and for long term strategies you need to involve the senior manager for the governance and safety of your project so that you are sure that you project is going to succeed. You need to know what the actual goal of your project is and if there is anything that changes this goal you need to have these discussions in time to see whether changes have to be put into the targets of your projects or anything else. It is also vital to have the right proper discussions with your suppliers. You need to make them understand, in KPIs as you know, what to really expect from the supplier. But you also have to give to the supplier the opportunity, and this is something that sometimes misses in these project, where we do not give the suppliers enough time and room to give their vision in certain topics. And I would say this is one of the best practices you always need to do this in those kind of projects. And there are also different roles in a team. There is someone who is called the service owner, this one has to deliver to the service or the customer, organization within the bank once we are done with our job. So everything needs to be in the right time in the right place in the right specifications. You have a project leader who is responsible for timelines and milestones as I said and some piece of governance with all the technical related teams around it. So there are a lot of things that you need to discuss and to check with other people who are directly or indirectly involved in the success of your project.

GA: And now let us say that you are faced with a new job opportunity and you have two options to consider. Ambition wise they are the same for you progress of your career. The difference is the one would be more social, In a sense that you recognize that you will be more involved with more people and the one Is more lonely job. What would you rather select between those two and why?

S6: Honestly I would not care. Because I am really good in working In both environments. I am really good at working towards goals, targets and timelines which is ((incomprehensible)) and I like to do this of course in an environment where you can socialize and team up with other people. But I can say that the smaller

the team in which you are involved in then the easier is, most of the times, to make progress. Because you have less discussion and different kind of views on several topics. But that does not necessarily mean that I have a preference in one or the other.

3. Achiever player type

GA: So moving into our next topic. Often in a working environment appraisal of performance is shown with certain types of rewards. This can be something extrinsic that can come in the form of money, such as bonuses or raise or it can be something intrinsic like a good word from a co-worker or a manager. Do you think In your opinion that rewards should always come hand in hand with successful undertaking of work related tasks?

S6: Yes.

GA: Why do you think so?

S6: Because I think like you already said that is where you have an incentive for the people to have the best performance or the best result and do things in the most efficient way. If you would put it in another way around, I have been part of this process within the bank where a few years ago we have been rewarded based on results which has been stopped a few years ago. So today it does not make any difference any more how do I perform in my job, I still get the same salary and that is strange. Because if you see the people from the other side of the table sales people who I have to work with every day they do get pay by results or revenue or whatever. And you can say whatever you want but it does help sometimes to get people to get the best result, it is true.

GA: So between those two types of rewards, so something that is extrinsic and it might be connected to money and something that is intrinsic like an appraisal from your manager or a recognition from your coworkers for your work, which one would you select? Which one do you think it is the most motivating for your work?

S6: I would say the money. Yes.

4. Killer player type

GA: So moving on to the next topic. Do you think it is important to adhere to standardized processes or rules while you are conducting your work duties? So I guess in your working path you have to adhere to certain rules or processes that are specified by the organization. Do you think for those that you should always strive to adhere to them or do you think you should try to bent them sometimes?

S6: I think it is good and practical for most people that you do need some guidelines. But I only use those guidelines if they are necessary, if there are no other resources I would say you can use to have the proper way that you work with other people. For example if a manager says to me that he wants to do this business for this amount of money with this vendor, then I would say could you

please give me the arguments why we would not have a different number of competitors involved so we can have more things to compare with and to benchmark for instance and then get the best price performance which is very usual and common thing to do in the procurement business. Usually you can always motivate someone that there is nothing to lose there and that there is always a win win situation both for him and me. Only if this guy would persist and not give me the arguments of why he would not just go with me into the same direction with me and say this is the best thing to do then I always will fall back and still use the arguments that we have from protocols and procedures within the bank. But this is not something you typically would like to put in the table when you have a first discussion with someone because that would logically put some tension between you and the other person. So I think my skills are, that if you can call yourself a senior consultant in this business you never need to get off these procedures to tell someone how he or she should do it.

GA: So you always have the luxury of improvisation in your work?

S6: Yes I can always can talk to people in terms of trying to think of similarities and mutual goals, and that usually works a lot better than trying to emphasize the differences that you have and then in the end if there is nothing else that you can use to make the other person to do what you want by really forcing him by a set of rules...So this is my experience, I have never had to use them before until now.

GA: But let us say that you are in front of an end goal that you want to achieve and there are two paths. And let us also say that you have the means to analyze them beforehand and see that if you follow some standard processes and if you follow some improvisation of your own or something that you've thought of then the end result will be the same and the quality of the work would be the same. So would you try to improvise in this scenario or you would be like okay let's get on with that let's follow the standard process. I mean do you enjoy to do some things in a different way even if it is not completely necessary?

S6: Yes, yes definitely.

5. Explorer player type

GA: Do you generally enjoy working independently and being able to follow your own path? So I guess you have some standard working path, which consist of your projects, deadlines etc. In this main working path you may be given the opportunity to do something else and get the chance to get out of the loop in a sense and then return. Is this something you would like to do?

S6: Yes, my preferences is what I like to look at the human side of the project and just have all these discussions and see if you can reach an understanding with people and just motivate them to get them on your side, rather than like I said fall back in all kind of procedures. Which is something that you always can't do but this is what I think is the most attractive factor of my job to put it in this way. Otherwise it would be a standard job with a repetitive work that you always do

it in the same way. If I had to do a job like this then I would quit just right away, because this is something I would not feel comfortable with.

GA: So you like to have an overview of the project and work on different aspects of it?

S6: Yes, yes.

6. Dominant player type and generalization

GA: So you may have noticed that so far we have talked about four categories of qualities of employees. They are not mutually exclusive, so it is not necessarily that a person cannot have characteristics from each group. To recapitulate what we have said, we talked about employees who are mainly driven by their social environment and the interaction with it, we talked about employees who are rewards oriented and they strive for competence, we talked about employees who are out of the box thinkers in a sense and they are intrigued to question the status quo and finally we talked about employees prefer to work independently and are curious to try to do new things. In which one of those categories would you position yourself if you had to select only one of them and why?

GA: If you think it is better we can rate each one of them from 1 to 10, just to quantify how much each of those is representative for you.

S6: Let's do that way, yes.

GA: Ok, so how much would you rate yourself as being an employee who is mainly driven by their social environment and the interaction with it?

S6: 6

GA: Same question on how much reward oriented you think you are.

S6: 9

GA: Same thing about how much out of the box thinker you are and how intrigued you are to question the status quo.

S6: I think that should also be a 6.

GA: And finally how much do you enjoy working independently and you are curious to try new things.

S6: Yes I always work as a simple point of ((incomprehensible)) as my boss calls it. So I am always working by myself as a consultant. So I would say 8.

7. When and how will employees contribute in Enterprise Crowdsourcing?

GA: With that we concluded the main part of our discussion. Now I would like to ask you two more questions that are related to the project. So if you were to join this research and contribute some data by using this mobile application, what time of the time would you ideally want to be able to contribute, during a working day?

S6: I presume in the afternoon.

GA: And that would be which hours?

S6: That would be after 12pm.

GA: And how much time do you think you would be able to devote every time for the application? So how much time do you in general spend everyday using your phone and maybe what portion of this time would you like to devote to this?

S6: An hour.

GA: An hour is the total that you check your mobile phone?

S6: Per day.

GA: Okay, so how much time out of this you could devote from this for crowdsourcing?

S6: Half hour.

GA: Okay. That would be it, thank you so much for your time.

S6: Okay, great I hope I contributed something to you and you can do something with that.

GA: Certainly.

Interview Information

Date of the interview : *22-02-2017*
Place of the interview : *Croeselaan 18, 3521 CB Utrecht*
Duration of the interview : *00:20:08*

Participants

Interviewer : *Gregory Afentoulidis (GA)*
Gender : *Male*
Profession : *Graduate Research Intern*

Employee : *S7*
Gender : *Female*
Profession : *Senior Jurist*

Interview Transcript

1. Introduction

GA: So before we continue with any actual questions I would you to make a small introduction of yourself. So what is that you are working on in Rabobank etc.

S7: I am...I work within legal department and I work with insurances like product approval processes and also within credits and I advice our business on risks and also how to make the products correct, so it is within our laws and the international laws.

1. Socializer player type

GA: So moving on to the first topic of our discussion. As part of your work how important is for you the social environment of your work as a motivator for your everyday tasks? So for instance like being able to share knowledge with your colleagues, to contribute to their well-being and in general enjoy group activities.

S7: Very important, because I think that the social aspects improves the work relationships with others. Also If I can bring my knowledge to other people also within the intranet then they can use it for their products. So I think that the social aspect Is really important.

GA: So how important would you say for the efficiency of your work is this, so in the way that you conduct your everyday tasks? Not necessarily strictly tied with your core work. So to be more specific let us say that you have a working environment that is really social, but it is not like that this can help you with your core work but people are there and you are able to discuss with them and then you have another working environment that you have to do your work and there is limited access to other human assets. So how do you regard in that sense your social environment as being important for your work?

S7: Very, very because I think in my work I work with people from risk, from compliance, from legal with the business and if we are together we can get along and discuss the problems that everybody has and we can come up with solutions. Sometimes we do not do that, and we do it by email and we send each other emails and that is not effective. So the social aspect that brings people into a room is very very important to the efficiency of our work.

1. Achiever player type

GA: Okay, so moving on to our next topic. So often in a working environment appraisal of an employee's performance is shown with certain types of rewards, either it can be a bonus or a raise to your salary or just a good word from your manager or your peer. So should in your opinion rewards always come hand in hand with successful undertaking of work related tasks?

S7: Sorry, what do you mean?

GA: So do you think that rewards should be strictly attached to the performance of your work? So do you reckon that the end result of a good work that you do should always come with a certain reward?

S7: Haha, this is a hard question. Yes, I think so. I think if your work hard and you do your work very well and you get rewarded that is an acknowledgement of your good work, and then the next year you want to work even more and even better. And if you do not get this acknowledgement then you get dissatisfied, but I do not thing that the reward should always be money, it could be something like courses that might need money, or hey that was a good job. You know different rewards.

GA: So as I understand rewarding is an important factor but it shouldn't be there every time that you do something.

S7: No,no. Yes, not every task should be rewarded you know.

GA: And if we were to categorize the reward types. So extrinsic would be something like salary raise,bonus and intrinsic would be something like getting an appraisal from your manager or your coworkers and for example the fact that your are known in the company because your are good at something. What would you choose between those two categories?

S7: Haha, those are really hard questions. I think all of them. It is not a good answer but what I mean by that is that if every year I just only get extra money of bonus it is not satisfying because it happens every year and then again if you get only course or good words is also not good. So I guess sometimes the bonuses, sometimes a course...So I cannot choose clearly between those.

1. Killer player type

GA: Okay, so do you think it is important for you to adhere to standardized processes and rules while conducting your work duties?

S7: Well if you ask my opinion as a legal consultant I say yes. Because we are obligated to.

GA: Let us say out of your working obligations. So as a character.

S7: Yes, not too much. I think that if people are experts then people can perform well. I think that this is a little bit the same question as the guidance. You do not need like really a lot of pages and rules. But you I think you need a little bit, some rules a vision or a strategy and within those rules you just have to let people do their job. Give each other feedback, because I think this Is really important because you have people within every organization who do not perform that well and so you need to have some structure how to recognize those people.

GA: So also if you find some processes in your work that do not allow you to work as agile. Do you think you are the kind of person that tries to pinpoint them and tries to make them better?

S7: Yes, yes absolutely.

GA: So you would not say that you are the kind of person that sees a process that we can trust, maybe it is not that agile but we can trust so you say let it be for now and you move on?

S7: Well it depends.

GA: Are you intrigued to change things?

S7: Yes I am really intrigued. I do that all the time. It is outside of my profession, but I do this where I can, I challenge the processes and I ask people what do they think and challenge them for an improvement.

1. Explorer player type

GA: Ok. So for our next topic I would like to ask you if you generally enjoy working independently and choosing you own path? So to make more clear I guess that you have a main working path but you maybe you are given options to swerve a little bit and maybe do things In a slightly different way, would you actually choose to do that?

S7: Yes, I do that already. Of course I have a path and I have to do my work, but I do not thing that I am the regular Rabo type. I like doing a little bit of this and then a little bit of that and I am all over the place.

GA: And do you enjoy having some kind of guidance in your work?

S7: A little bit guidance but also the trust of your manager that you can do it ok. So a little bit of guidance is ok, but too much guidance you know you have the feeling that you are trapped and you feel you cannot go beyond the lines. I do not thing that this is good.

GA: So let us say that you are given a project, which is very open-endedâĂę

S7: I love that.

GA: Ah, do you enjoy that?

S7: I do a lot. But most legal people are not like that but I like that.

GA: So you like finding your own way?

S7: Yes, yes.

1. Dominant player type and generalization

GA: So let me know summarize a little bit what we have told so far since we have discussed some employee qualities in a sense. It is not that those are mutually exclusive you can perfectly have qualities from several categories. In order to summarize them for you we talked about employees who are mainly driven by their social environment and the interaction with it, we talked about employees who are reward oriented and they strive for competence, we also talked about those who are out of the box thinkers and intrigued to question the status quo and finally employees who like to work independently and are also curious to try new things. So if you were to choose only one of the above and put yourself into, which one would it be?

GA: We can also rate each category from 1 to 10 to make this a little bit easier.

S7: Okay, yes let's start with the first one.

GA: So for the first category we have the employees who are mainly driven by their social environment and the interaction with it.

S7: I think that this is important. I would say 8.

GA: Employees who are reward oriented and who strive for competence.

S7: 7.

GA: Employees who are out of the box thinkers and are intrigued to question the status quo?

S7: 9

GA: Employees who prefer to work independently and they are curious to try new things?

S7: 8

GA: Ok. So let us now assume that there is an average Rabobank employee, out of your experience in which category would you put him in?

S7: None. Well not the out of the box thinkers I think this is not the typical Rabobank employee. I think the last one.

GA: So those who are preferring to work independently and are curious for new things?

S7: No, no.

GA: So there are also the socializers and the...

S7: I think the socializers. I think that is the average Rabobank employee.

GA: Not let's modify the question a little bit. Let's say that you are in a hiring position, and you are hiring new employees. Which of those qualities you think are the most important?

S7: I think the most important at this moment are those people who are out of the box thinkers. Because I think that Rabobank needs a lot more innovation. We are already trying, so we have for example the fintech innovation but I think that is a really small part of the Rabobank. I think that there are a lot of people here who are working a very long time on the same work, so some people are doing 20 years in the same team with the same people the same work. And those people are really specialized in one thing and they have really narrow vision. If I were to work in a same position for 20 years I would have had that same narrow vision also, it is inevitable. But I think a lot of people here are sitting too long in the same position.

1. When and how will employees contribute in Enterprise Crowdsourcing?

GA: Okay, so with this we wrapped up the main discussion. Now I have the final two questions which are related to the research. So let us say that you would be able to contribute to this endeavor. What time of the day would you be able to contribute?

S7: Well I want to do it everytime in a day. But if you are asking me when do I have time Thursdays, Tuesdays and sometimes Mondays I am the whole day in meetings. Wednesdays and Fridays are the most quiet days because most of the people are free.

GA: But more specifically what time of the day is the most convenient?

S7: 5PM or 8:30AM in the morning. Monday morning 8:30. But in general I think it is better in the morning, because generally at the end of the day you have to wrap up stuff. It does not take a lot of time,so you know you start your computer you check your telephone and then you do some work there.

GA: And a last question is how much time do you think you would be able to devote every time you open the application?

S7: Me personally or generally?

GA: Yes you personally but you can also answer both, so also what do you thing in general?

S7: For me it does not matter, If I help Rabobank I can do this everyday for half an hour or an hour. I do not think this is a reality...haha

GA: Well that would be too much. It is more related to the overall time that you are using you phone every day.

S7: If you want to let a lot of people participate then I would say a minute. Has to be really quick you know. I think what is also really important Is that managers who are really enthusiastic about it should encourage their employees to participate. But I think that this is not in the question.

GA: Great, thank you so much it was a delight.

S7: Yes, sure thank you.

C.2 Expert interview transcription

Experts' Interview Information

Date of the interview : *07-12-2016*
Place of the interview : *Graadt van Roggenweg 400, 3531 AH Utrecht*
Duration of the interview : *00:41:08*

Participants

Interviewer : *Gregory Afentoulidis (GA)*
Gender : *Male*
Profession : *Graduate Research Intern*

Expert 1 : *E1*
Gender : *Male*
Profession : *Global Strategist - Dairy*

Expert 2 : *E2*
Gender : *Female*
Profession : *Dairy Analyst*

Expert 3 : *E3*
Gender : *Male*
Profession : *Industry Analyst Animal Protein*

Interview Transcript

1. Introduction

GA: So a little bit to recapitulate what we told in the kick-off meeting and also give some information about my research. So I want to research how we can actually collect data from employees of Rabobank, in order to train an artificial model. In general artificial models help the work of the employees in the company, so this is their target. In this project specifically we are targeting the food and agri experts so we can find a task that is manageable for an Artificial model to do automatically. So the main requirements of this meeting, and this is where I want your contribution, are for me to learn how you work, how you actually compile FAR reports, which data sources you use and also identify autonomous tasks, independent tasks in this pipeline and thus identify parts of your work we can automate with an AI model. So I have some questions written down that I want to make to you, but in general please feel free to contribute in this discussion in any way you find suitable so we can have those outcomes. Before I go on and make any questions do you have any question that you want me to answer? Regarding the research, about this meeting?

E3: No, not for me.

E2: No.

2. Experts' work

GA: Ok, so first of all because I do not have any insight on how your work is done the first question I would like you to answer is, what are the job duties of a food and agri expert. Like how do you work, and what are the steps you follow to compile a FAR report?

E1: Ok, so we have a number of outputs that we create. Most analysts will create 2 reports a year but on top of that we will write some web articles, blog articles, we will also spend some time communicating the results of that research externally. So we will write press articles or contribute to press interviews. Another part of our job is to engage with clients, often talking about the research we created but also often talking about the general sector marketplace. So we will spend time visiting the clients, we will prepare a powerpoint slide deck which will take to the client, we will present to the client on a slide by slide basis and that's normally done alongside a banker who then suggests a commercial solution to some of the things we would highlight. So an example would be ... prices are rising in the marketplace at the moment maybe you should hedge your purchases for a particular product to ensure that you can buy them cheaply over the next six months. Or lock in your prices or sell early or create stores. So something that we apply the research to what it means to a company, and the bank will make the proposition from a commercial point of view.

3. On composing FAR reports

GA: So I would like to know more details on how you...what are the tasks you undergo when compiling a FAR report. So you said for example you want to provide insights about prices, let's say that you are now interested in a specific product, agricultural product, how do you go about compiling a FAR report about the markets for this product, the prices and the key insights that you want to offer to clients?

E1: Well it depends on the type of report, because if we were trying to compose a report on a particular marketplace or a particular product or a trade agreement or just the market in total, obviously the data sources would differ. But the way I would go about it is I would normally start with the general, so I would perhaps compile a list...well normally E2 compiles a list of the press about this subject all of the press reports about this subject. I might look at all of the âĂŞ

GA: Press...can you be more specific? Press reports ... like you use something specific or...

E3: Well say you want a FAR report about milk or meat in Russia, then you go find it on the press or things about meat in Russia.

E1: So we have some generic search tools for media, we use a tool called the Factiva [https://global.factiva.com/factivalogin/login.asp?productname=global] which is a Wall Street journal product, which searches all of media and searches web entries, blogs and printed media to compile a list of the general information about the sector. I would spend some time reading through them, and recognizing what the key subject matter would be what are the key points that that media is pointing towards and would probably spend some time looking for the key companies that are operating in that sector...((incomprehensible))

E2: In between you would perhaps also try trade magazines, something that is not always in Factiva and also specific subscription for specific sectors that we use...(coughing)((incomprehensible)) . For dairy for example we would use dairy industry newsletter as well.

E1: Yeah...so there is...I guess media search would be wider than one database across the number of databases and then identify from those media searches the key companies, I might look at their annual reports, their published literature, websites, their equity broker reports, there could be broker reports, collect information about the commercial activity on the subject depending on what the subject was âĂŞ

E3: Yeah...

E1: Once I have done that, I might identify some particular issues about a market, perhaps a particular product seems to be in more demand or less demand, and that might lead me to want me to do some more particular research on one particular aspect, at which point I would try to obtain some data. It might be from one of the databases which we subscribe to. Things like Euromonitor

[http://www.euromonitor.com/] it might be bespoke data that we buy or it might be some data we choose to collect as a team from public sources or from clients. And so that might lead me to some statistical analysis of those data and probably my final stage would probably be, having found what seems to be of interesting point, that is maybe novel I would probably go back to those companies that they are involved in the sector, many of whom would be clients of ours, and I would discuss with contacts within those companies whether these points are relevant and then I would compile the results of this analysis into either a powerpoint presentation or a written report or some kind of graphic or website article or whatever it happens to be.

GA: Okay...so I will just recapitulate what you said and correct me if I am in something mistaken. As I understand it there are two steps in high level. The first is an exploratory step, so you go about using the tool for media... the media tool...getting information from papers,blogs etc. And then you try to identify in this step key points, key companies operating in the domain, depending on what the product you want to do research is, and then when you find something novel some keys insights that you want to further investigate then you move to the higher level step of analysis...

E1: Mm-hm...

GA: ...so you go back at your resources that you have as a department and try to find something relevant about that and as a final step, just to make the picture whole you also...

E1: Communicate...

GA: ...communicate with the companies to be sure that the key insights that you have are those that...they are valid and then you compile a FAR report-

E2: Do you think the same Albert?

E3: Yeah, mostly. Some times you know already from (other) discussions on other subjects with companies that there is a subject that let's say, that might be interesting or you might want to know more about and then you have already a target.

GA: Yeah, obviously from experience you draw insights and then you are able to steer more to something rather than being fuzzy about it.

E3: Yep , yeah. But when the bank wants to know something about a specific country we do not know anything about it then that is the way forward of course.

GA: Ok.

E1: And it depends on the particular subject matter because for some subjects we might jump straight into the analysis because we already have those data.

E3: But still the media search tool can be quite fairly useful.

E1: But in general we start with the why ... we may start with a process before that, as Albert quite roughly said, we identify among our clients what might be interesting or might be at service. Our role here is twofold. One is to increase the profile of the bank within food and agriculture, so part of our role is to be ((incomprehensible)) and show Rabobank's expertise in the sector, being ahead of other banks, and the second part of our role is to provide insights into how our clients and ourselves might make more money out of a particular sector because we understand it better, so how do we leverage that knowledge into commercial activity. So, in doing that, we spend a little bit of time talking to people what might be interesting to clients or suggesting to people what might be interesting if we have a look at that particular aspect next year because things are likely to change. One thing that might be at the moment, are things like trade relation were we see a lot of assumptions in TPP Trans Pacific Partnership and North American free trade and CTA they are all beginning to be questioned, so maybe that is kinda of an agenda for the coming years across the sectors.

4. Experts' data sources

GA: Okay, so let us make an overview of the data sources that you use. You already told me about the media search tool that you have...What is the name again? Sorry...

E2: Factiva, ((incomprehensible)) , that is a very broad use...it is a commercial paid subscription and there are a lot of newspapers in all kinds of languages. So trade magazines...but it is mostly news.

GA: And the other data sources are the internal data of the department?

E1: No, we also have access to a number of journals that we subscribe to which will be particular journals...

E3: Subject related or industry related.

E1: ...Yeah, that's right.

E3: For the Netherlands, ((incomprehensible)) , farmers magazines and industry magazines.

E1: And often they have search facilities on their websites, so we can count on that.

E3: And then of course the general sources from European Commission, USDA, country statistics, ((incomprehensible)) databases.

E1: And then things like currency traders would provide us with daily currency exchanges, so we use things like that, UN population bureau...

E2: We might use Bloomberg...

E3: Yeah, Bloomberg, yeah...

E2: ...for prices...

E3: ...and also some different sources as well, some general and some specific...

E2: ...and also maintain ourselves price data, production data, trade data, that we use often in our research.

GA: So let's say you do not have any insight beforehand about a product that you want to make a FAR report. Would that be different depending on the product itself? So let's say you have animal protein in one hand and dairy products on the other. Would you go and look at all these data sources that we just discussed or you have different data sources for different products.

E1: No, I don't think we would employ you as an analyst if you did not have a good understanding of your sector. So most people we employ ... well that is not true. Many people we employ have been around these sectors for long enough to know where to look and where to find things. We of course employ graduates, and we have to explain that to them, but is something that you gain experience pretty quickly. If you want to know about dairy you probably look at the dairy industry newsletter...

GA: So this is the âĂIJarsenalâĂİ of data sources that you have and...

E1: Then we subscribe into things like Euromonitor ... that's not Euromonitor is that?

E2: Hm?

E1: Euromonitor...

E2: What?

E3: Euromonitor...

E1: Yeah, Euromonitor...

E3: There is also a database about consumer...

E2: That is if you want to know something about consumer trends, or brands, market shares of companies then we use Euromonitor.

E1: And then depending on the sector we would subscribe to particular datasets that we buy. From dairy point of view we buy from ZNB which I guess ... is that a journal or a dataset?

E2: Journals, pdfs...

E1: Yeah so those commercial consultancies would supply us with data as well...

E2: And so often sources are, yeah, paper based, pdfs and other ones have advanced online search tools... that differs.

E1: And then where we can we get access to, without running afoul of legal issues, we get hold of broker reports. There are some internal issues over compliance that limit their use, but if we can get hold of the company broker reports on the key companies then we use those. And then company annual reports and publications are the key area of...

5. Information extraction

GA: Okay, another question is so how do you go about ... so you filtered out the data sources that you use , how do you go about looking for what you want to extract? For example you said before that when you have a subset of the data that you want to use, you usually go through them and read them. So you read the whole reports or is there a technique for doing that or out of experience you know where to look or what to look to extract the information that you want?

E3: Say you want to look for meat in Ukraine and you tap meat Ukraine and then you get a long list and of course you get the titles and then when you see meat highlighted and then you can find out relatively easy which might be interesting or not. That is the way it goes. It is like having a newspaper you look at what is interesting and what is not interesting.

E1: I think there is also the relevance of the author, so if I was looking for something on dairy products in America would I read something by the US department of agriculture? Probably! Because they are the official recording body for the country. Would I read something in ah::: milky stuff journal or...

E3: The newspaper of Minessota...

E1: Yeah, yeah haha. You can pick out the ones that are relevant and read those and then maybe that can give you a view of a subject matter that you can refine your search a little bit and the number of titles gets shorter and more relevant towards what you actually are looking for and then you can drill down accordingly. Another way of doing that is of course to talk to people, so we talk to people about what the issues really are and that allows us to narrow the search before we start. Frequently these research projects get off the ground because people are talking about a subject so you talk to key CEO in the sector or key finance director and he says : âĂIJOh, I heard some research is done in the University of Dublin...so I was looking at thisâĂİ, so we start off by reading some of their papers or journals about it.

GA: Okay, so you get offers and you have some key insights about which parts actually would be more plausible to write something. So you have like a two or three steps of narrowing down to what is important and then you read them and you extract the key insights.

E1: It is probably not as disciplined as the literature review you do for a scientific paper, because our research rigor is not that standard we are at the end of the day drawing attention to trends and providing insights we are not trying to move the body of science forward as in a peer review journal. So our research methods

are perhaps more flexible than that but it is a similar approach to writing a peer reviewed paper.

6. Candidate tasks for cognitive solutions

GA: Okay so do you identify in this pipeline that we discussed so far a task that is common regardless of the domain that you are researching and is also ... it takes some time, it is repetitive and it would be of use to be automated? Can you recognize a task that has this characteristics? And ideally has a relation to analyzing textual data sources.

E3: The search for...the newsletter search...

E1: Yeah...

E3: ...to find the most relevant articles and papers. It is time consuming...relatively

GA: Anything else? So because I also do that when I do research...

E2: Well there is no other way...haha.

GA: ...nothing of that magnitude of course but I always have a problem of making a taxonomy of the papers that talk about this and the papers that talk about that and then somehow compile them and then I get lost in this process. So I guess this indeed is something of use but is there also something else that you find in your interaction with the data sources?

E1: Yeah, I have got one. I spend a whole lot of time working with banking teams who are often looking for companies. And companies that do a certain thing or approach in certain market or produce a certain product. And you think companies are fairly easy to find because they have websites and have...sell products that would not be difficult to find, but we do spend a lot of time looking for companies in particular markets or that they do a particular thing. E2 has one at the desk at the moment,that is...we are looking for a company that produces prebiotics...

GA: Produces sorry?

E1: ...prebiotics these are normally things like sugars that cannot be digested by humans but feed the bacteria in you gut and provide the good bacteria. And we have a client who is already operating in that field but wants to collaborate with others who are operating in the field and we are looking for those. And we spend a lot of time...

GA: Trying to figure out which companies...

E1: ...trying to figure out which companies are there. And normally there are one or two parameters around it, so we want a company that does this in these markets and we want it to be...have a turnover between this size and this size.

E3: That is always a challenge. We have done this also last week,searching for companies in certain markets which produce that specific product...

E1: And it is a pain...

E3: You know that they produce all these ((incomprehensible))

E1: ...it would be great if companies put their products in the database and you could search but...

GA: They don't.

E1: ...they don't. It is a very difficult thing to do. It is actually ... most of the time you are looking at their websites or you are looking at their annual reports...

E3: Trying to figure out if they have the parameters you are actually looking for...

E1: ...yeah...

E3: ... so it is a typical Google search, or search I mean in the browser. And link the companies that are in the domain ...

E1: Yeah ... Google search is good at that haha ... it is a good starting point. Have you found any things?

E2: I mostly ...

E1: Mostly E2 does that.

E2: ... mostly I use conference web sites where you have the exhibitors or the participants or the ... or you can try the ((incomprehensible))

E1: E2 is an expert at this, and spends a lot of time doing that and just takes it ...

E2: Haha, if you can automate me ... haha.

E1: And she pointed out last night : 'Just because I am good at this I do not have to do it a lot', haha.

E2: haha.

... ((inaudible conversations)) ...

E1: When I think my research in a whole, because there is an element of, you know, the expertise short-circuits a lot of it when you are looking for something particular I find it difficult ... I have thought a lot over this the last week if ... the one that really comes in on this is finding the companies that are in that space that I want to actually talk to or who are doing something and then I want to look at it in more depth ...

E3: Especially get a certain size or a certain volume because when companies want to buy company they want to know a certain value or volume, production volume because they have some ideas about it. If it is too big they cannot afford it if it is too small they would not want it and they ((incomprehensible))

E1: The parameters are often quite obvious and they are not difficult but simple, which makes one of the criteria that you guys talk about, but it is a lot of work and it is like searching a needle in a haystack. So you have to look at a lot of companies to find the one or two that fit your criteria ...

E3: And the problem is you cannot call someone because then they know : âĂIJ Hey they are looking for something and there might be something onâĂİ, so the rumor is starting and you must be very aware of this.

E1: ... so that is the best I can come up with, it is difficult to think of any more actually ...

E3: Only the search for the ... the text search.

E2: That could be eh.. what I was saying if you go to company then you also need the news search. If then the computer can select the most relevant ones and we do not have to put ((incomprehensible))

E1: Yeah ... yes. Well I do not need that E2 does that for me haha.

E2: Haha.

GA: Okay, so one task is to find relevant information to facilitate the search for a specific information need, so this is also what we talked about in the kick-off meeting and the second one is to find information about companies relevant to a domain, given some specific parameters such as revenue that they have ...

E1: Yeah, that's right. So yeah, so it is bad looking for relatively few criteria, so normally it would be between these sizes of turnover, it has all ((incomprehensible)) in these markets and has these products. It might be as simple as that.

GA: Yeah.

E1: But then you have to run through a lot of sources of companies that might fit those criteria to find them.

E3: Especially time consuming. It always comes when you are very busy.

E1: Yeah.

GA: Nice, so I am very delighted we found two tasks that could be of use. So just a final question, let's say we have this system now at hand and let's take the two examples that we already talked about. What would be ideally the outcome you would expect from a tool like that? Let's go and talk about the relevant news. What would you expect as an outcome of the system? So let's say you feed it with an initial big set of data and you want it to come out with the relevant ones. One example would be just to point out this is relevant and this is not relevant, another example would be to highlight paragraphs that it finds specifically relevant to you information need. So at what granularity would you want to get results about your search?

E1: I quite like to rank the outcome, so if I am going to a client typically my expert assistant sitting to my right here [he means E2] would send me along a folder full of 20 or 30 pages of press cuttings which I would normally read on the plane on my way to the client just to be familiar with all the things that have happened with the company or the sector or ... I think if we could rank those so that by reading the top 3 ...

GA: You get a very good quick overview of what is going on and ...

E1: ... exactly. So I am going to read the most relevant first and the same one I am doing research, if I don't have to read 20 articles I can make my life much more efficient by perhaps reading 3 articles and getting 90% of what I need to know about.

E3: Especially you do it normally at home, in the evening when you are already tired.

E1: If you can streamline that sorting through and sifting of data, whether is ahead of a client visit or whether is part of a bigger research project so you are up to speed more quickly and then you can refine what you are looking, that is the process.

GA: And same question about the second problem that we have discussed. So let's say you want to find relevant companies in the domain, you need just the names or the size ...

E1: We need It in a format that we can sell to our M&A colleagues because we can make a fortune of doing this-

E2: Haha.

GA: So what would that format be?

E3, E2, E1: Haha.

E2: Slide format, with the logos and the products and the financials. [in a jocund tone]

E1: This is going in a format that is going to probably construct a set of company profiles that is going to a slide deck somewhere. So we identify companies that fit criteria and we present them to the colleagues internally or to our clients.

E2: Perhaps we should better talk to the M&A

E3: No just the list of the names which are meeting that criteria, say between a hundred and one million sales and these three products that they produce.

E1: Yeah, but if you are in search for this data you are also picking up the data that somebody else needs to go back and collect ((incomprehensible)) . So maybe some standard data, like turnover, like EBITDA [Earnings Before Interest Tax Deprecation and Amortization], like number of employees, locations ...

E2: And perhaps volume...

E3: Yeah.

E1: ... company history, or CEO name, ownership. Some ... it might be cooperative it might be private ownership it might be public ownership, so ownership status would be one of them. We could supply a list of a dozen data items which if they are included in the review material it would be useful to collect rather tell us what the name is and having some going to collect that same piece of information and pick up the data.

GA: So just to be a little more detailed than that. Let's say I am an expert and I do not know anything about it and I just came across a company that might be relevant for a specific need that you have. You gave me a task to find that. Is it easy for me to find the information that you just said just by browsing the internet. So would you expect a company in this domain to have a Wikipedia page or ...

E1: Could be, could be. Not for all companies you would be able to collect the data, there might be some blanks in the database. But if you are visiting the page of a public company say, and they have their accounts then you can pick up all that data from that page.

E2: But mostly yeah, US companies do not publish financials, and private companies ...

E1: A list of names would be sufficient, because it is always possible to go back and review ...

E3: Yeah. But especially the names because you know they are between that criteria and then you can dive into it ((incomprehensible)) .

E1: ... but even if you are looking at the criteria, so if you say companies with turnover between five hundred million and billion, then you find a company that has got a turnover of seven hundred million, which is clearly in the list it would be useful to collect the seven hundred million [the turnover data found] and report that at the same time rather than just the company.

E3: Yeah.

GA: Yeah, the more the better.

E1: : Yeah sure. You do not want to overcomplicate the task if can look for that piece of data anyway.

GA: Great, I think I have got all the necessary information that I wanted. Thank you, you were very specific to whatever I asked, thank you for letting me record it. Do you have any other questions that I could actually help you with?

E3: The next step.

E2: Yeah what is the next step?

GA: So the next step is that now I have some key information about what is of value for you to automate in your work. I will discuss it with my supervisors, we will try to turn to a computer science problem and also translate it for my research to tasks. So you gave me now the requirements of the AI model, that the model has to do this and we as computer scientists will need to find out what are the tasks that we to feed to the model to be trained and be able to do the task that you want. So I will maybe need to ... depending on the task that we will select and the feasibility of it I will maybe need to have data sources that you use so I can extract tasks that I will deliver through a mobile application to the employees.

E1: Okay, sounds good.

GA: There is also something else. So the part of modeling and training the application will be done by a second student that will come from January on. My part is to collect the data for the model to be trained. But in order to do that you need to know what the model will do, so this is why the interview was of high importance for me and thank you very much for it.

E1: Okay, there is one other thing that while you were talking came to mind. We collect data from various sources and compile our own databases with regularly used data. And of course some that will be scraped from internet sites, not that we are very good at it but we will get better. But then often ... if we automate that, the validity of that data can often be later questioned. So government sites often get data wrong, we found, and of course you can do that with range checking, you can check your validity of data, but there is something around collecting of data from various sites that we regularly use and where we can actually automate that.

GA: So, you have the data sources that are unstructured information and you want to turn them into structured information?

E1: No this is structured information. We will for example ((incomprehensible)) out of the European Commission data on dairy productivity or dairy prices across Europe. Quite frequently are Italian colleagues or Swedish colleagues or our French colleagues will retrospectively change the data and not tell us about it. So this is kind of keeping our database in line with others, is often a repetitive task which is quite difficult and part of that is range checking. So if the production in Spain goes up 300% next month ((incomprehensible)) with the cows who actually have given 300% more milk it is more likely to be that someone's hit the null button too many times when they import the data and nobody has picked it up. And then we risk putting that into our models.

GA: How do you validate it?

E1: How do we validate it at the moment?

E3: To make a graph of it and say : 'Oh wait, why there is that outlier?'

E1: I mean partially the reason why we do not outsource our data for maintenance, is because we have various attempts of doing that to offsite, to India. But when

you do that you get a non expert system, we get somebody who is quite capable of preparing your dataset but is not capable of saying production in Spain did not go up 300% last month because cows do not do that. So you then spend time retrospectively changing data.

GA: But for this you need expert background and this is an obstacle for us. Because when you see the data you have the background knowledge to understand something wrong is going on here. So this is difficult for us in a sense, because we will use crowdsourcing for that so the inexpert crowd will not be able to assess it.

E1: Fair enough, fair enough.

GA: So thank you very much for this conversation.

E2: You are welcome.