Measuring the Effect of "Bottom-Up" Gamification in a Microtask Setting

Full Paper

Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, Antonio Krüger
German Research Center for Artificial Intelligence (DFKI)
Saarland Informatics Campus, Germany
firstName.lastName@dfki.de

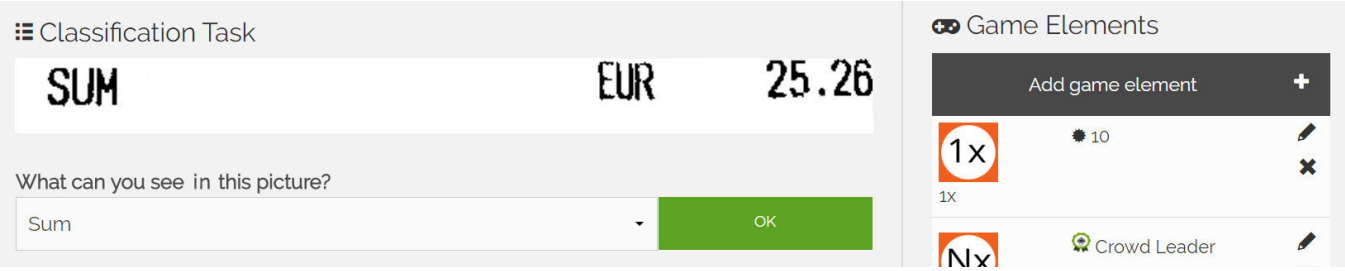

Figure 1: Part of the "bottom-up" microtask interface, showing a classification task and the (at runtime) selected game elements. The interface was originally in German and was translated for the images in this paper.

ABSTRACT

We investigate "bottom-up" gamification: allowing users to define their own gamification setup at runtime, i.e., they can select game elements fitting their needs. The goal is thereby to make gamification more engaging and relevant, by providing users with more autonomy. While it was shown previously that users appreciate "bottom-up"gamification, it was left open which effects "bottom-up" has in comparison to "top-down" gamification. We contribute to this by analyzing the effects within a microtask setup in the context of the text recognition domain. Participants (N=106) had to solve tasks, and were either in one of four gamified conditions (which varied in the amount of offered choices) or had to solve the tasks without gamification. Participants making use of the "bottom-up" choices solved more tasks, providing further support that systems should enable users to define their own gamification at runtime as this can lead to positive effects.

CCS CONCEPTS

• **Human-centered computing** → **User studies**;

KEYWORDS

Gamification, customization, user-led design

AcademicMindtrek'17, September 20–21, 2017, Tampere, Finland

ACM ISBN 978-1-4503-5426-4/17/09…$15.00
https://doi.org/10.1145/3131085.3131086

ACM Reference format:
Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, Antonio Krüger. 2017. Measuring the Effect of "Bottom-Up" Gamification in a Microtask Setting. In *Proceedings of AcademicMindtrek'17, Tampere, Finland, September 20–21, 2017,* 10 pages.
https://doi.org/10.1145/3131085.3131086

1 INTRODUCTION

Gamification, utilizing elements known from games in a non-game context [8], is used as one option in persuasive technologies to change the behavior of users, raise efficiency of solving specific tasks and/or simply provide more fun (cf. [26]). Since 2011 this topic has also become more relevant for academia [22]. Research tries to contribute to various aspects in this domain, among others whether gamification works (e.g.,[12]), how game elements can be manipulated to induce effects (e.g.,[2]) and whether there are individual differences in the effectiveness of gamification (e.g.,[25]).

The latter suggests that a one-size-fits-all approach is not the best possible solution, as various criteria (for example different personalities/player types [9]) demand different gamification setups. A gamified intervention can lead to frustration and pressure, if users' needs, such as their autonomy, are not considered properly [13]. In this paper, we investigate "bottom-up" gamification [16], i.e., users can decide which game elements (or none at all) attract them and can gamify a system at runtime on their own. As a difference from user-centered design approaches where (often only a representative few) users are considered at design time, "bottom-up" gamification follows the idea of customizable gamification setups at runtime that can be altered by every user individually.

In a recent paper, Lessel et al. [16] investigated the acceptance of such a user-led approach at runtime and revealed that users appreciate such an option, but did not compare this approach directly to a "top-down" setup (i.e., the case where the gamification

mechanics are decided by researchers, managers, or developers and not by the individual user (e.g.,[28])). Our contribution is to fill this gap by investigating this in a crowd-sourcing setup: Users were asked to solve simple microtasks (see Figure 1) and were told that this would enhance an optical character recognition algorithm (comparable to [33]) for receipts that would lead to a novel digital household accounting book application. As no definite goal was given on how many tasks needed to be solved, it allowed us to analyze whether different game interventions lead to more solved microtasks. We analyzed different gamified conditions (among them "top-down" and "bottom-up" gamification) and found that conditions that provide more autonomy can lead to a higher number of solved mircrotasks, compared to the "fixed" settings and using no gamification. Although the setup posed some limitations, the results provide further support towards benefits of user-adaptable gamification options at runtime.

The paper is structured as follows: We will first give an overview of related work in the gamification and crowd-sourcing microtasks domain. We then will introduce the study method and the system we used. After presenting and discussing the results made, we will close this paper with recommendations for future work that will lead to further insights into the "bottom-up" gamification idea.

2 RELATED WORK

Crowd-sourcing is used in various domains in which an algorithm cannot solve the problem alone or in a reasonable time frame [14]. The core component of crowd-sourcing is the idea of the wisdom of crowds [29], i.e., a group of people will come to a better decision than an individual alone. This concept is utilized in crowd-sourcing by providing many of the same tasks to different people, and their results are aggregated to come to a decision [1]. Research has been done about task length. Microtasks, i.e., tasks that can be solved within seconds (for example image tagging or simple decision questions), seem to be more beneficial for motivation, if the underlying problem allows the formulation of such tasks [30, 32]. Mason and Watts analyzed a paid setting and found that a higher payment increases the quantity of work performed (without having effects on the quality) and that a particular compensation scheme can have a significant effect on the quality [18]. They also stated, though, that using non-financial rewards and making tasks fun/harnessing intrinsic motivation is helpful for reaching the same or better quality. There is also evidence that money negatively influences the quality of the generated solutions [3, 19]. Gamification is one option that is often used in crowd-sourcing to motivate users to solve tasks without providing a monetary incentive, or with a lower payment. Work on crowd-sourcing and gamification has grown rapidly, according to a literature review conducted by Morschheuser et al. [22].

For example, Feyisetan et al. investigate gamification (points, badges, leaderboards and levels) in crowd-sourced image labeling tasks and found that gamification leads to lower costs and better accuracy and can make paid microtasks more engaging [10]. Altmeyer et al. worked on microtasks in the domain of receipt capturing. In an evaluation they found that gamification dramatically increases the willingness of participants to solve simple microtasks [1]. In their gamification approach, they used points, badges with levels and leaderboards, in a "top-down" fashion. Kobayashi et al. also worked in the crowd-sourcing OCR domain [15] and analyzed different motivational elements, among them also gamification (also in a "top-down" fashion, similar to the aforementioned approach), which also showed positive effects. These approaches show that gamification apparently has positive effects on the performance of unpaid crowd workers. All these approaches have in common that they use a "top-down" defined gamification approach. Work such as [24, 25, 28] hints that a one-size-fits-all approach is not reasonable, that individual differences needed to be considered and that users can decide best what they want. Thus, it remains questionable how many more tasks could have been solved, if instead of a "top-down" approach a tailored approach had been used.

Several theories exist on how games can create a context that increases the likelihood that people engage to play (see [27] for an overview); one of the most prominent is the Self-Determination Theory [7] (SDT). According to it, there are three psychological needs: autonomy (the feeling of acting under one's own volition), competence (to be able to solve a task at the proper difficulty level but also experience mastery and effectiveness) and relatedness (the feeling of being related to others) and when they are satisfied, the chance to be intrinsically motivated to do a task increases. With the "bottom-up" idea, introduced by Lessel et al. [16], autonomy especially could be reached, as users could define their own gamification at runtime. The authors investigated the acceptance of such an approach with an online questionnaire and a task management application, and their results indicate that such an approach is indeed perceived positively and participants subjectively reported changes in their behavior while confronted with it. Positive effects for a similar approach were also reported by Guy et al.: Within a platform, employees could create simple games and rules on how to interpret crowd answers to harness knowledge inside the company [11]. They found that people voluntarily participated and mostly created games for harnessing knowledge, instead of leisure games. In general, this work shows that users are open to creating their own game setup for crowd-sourcing tasks, which we follow in our experimental setup, with fixed tasks, but more flexibility in the game elements. From this, we hypothesize that "bottom-up" gamification should raise the engagement of unpaid crowd workers. For our comparison between "bottom-up" and "top-down" gamification in terms of which intervention provides a higher effectiveness, it seems reasonable to use microtasks and "top-down" settings as reported in the aforementioned approaches, as all led to more contributions by the crowd workers.

3 STUDY

We conducted a study to investigate effects of "bottom-up" compared to "top-down" gamification. For this, we use a crowd-sourcing setting, in which participants are able to solve as many microtasks as they want. Our focus here was not the actual results of the microtasks, but whether providing more choices towards the game elements led to a higher number of solved microtasks.

3.1 Method

We use the microtask setting as explained in [1]. In this paper, Altmeyer et al. have already analyzed one specific kind of "top-down" gamification and showed that the amount of solved microtasks is

higher in the gamified setting than in the non-gamification baseline condition. We received access to their data set containing pictures of parts of receipts (i.e., single lines from receipts, comparable to the "Sum" area in Figure 1), resulting in more than 6000 different possible microtasks. To mirror their setting, we also used their microtask types:

- *Classification Microtasks (Classification)*: In this task type, participants needed to decide what was shown in the picture: an article, a total sum, specific additional information (e.g., deposit), quantity indications or something else. They could state that the picture was ambiguous or that multiple elements were shown. Participants were presented with a drop-down menu containing these options.
- *Article Correction Microtasks (Correction)*: In this task type, participants had to type the article name shown in the picture. We left it open whether it should be typed exactly as seen or without abbreviations. The ground truth data contained both options.
- *Article Categorization Microtasks (Categorization)*: In this task type, the product seen in the picture should be categorized (e.g., as "food" or "clothing"). Participants were presented with a drop-down menu containing ten different categories overall.

We created a German online platform. On the platform's start page we briefed participants with the cover story that on this platform they will solve microtasks which will improve the recognition performance of an algorithm for automatic digitization of receipts. We also put emphasis on our research direction, i.e., that we want to create recognition algorithms that could later on be used in easy-to-use digital household accounting books. This should lead to a general positive framing effect, as shown in [20]. We did not mention that gamification elements are integrated, to avoid attracting only gaming-affine people or giving the impression that we have a game scenario (cf. [17]). A difference from the setting of [1] was that we only focused on the crowd-sourcing part, i.e., the participants had no access to the digital household accounting elements. We hypothesize that this will reduce the overall number of solved microtasks, as participants did not benefit from this further. The link to the platform was distributed via student mailing lists (consisting of design, computer science and psychology students) and the authors' social networks. We accompanied the link with a short explanation that participation will improve an algorithm to digitize receipts. No hint towards a motivational study was given. The participation was not rewarded in any way, to rule this out as a confounding variable.

When participants decided to take part in the study, they needed to enter a username and a password (but no further data was requested). We clarified that with these credentials they could leave the website and continue anytime later. This was followed by some demographic questions (age, gender and one question on their affinity to games), and a small tutorial in which every task type was explained and three such tasks needed to be solved. Besides the rationale of giving an introduction to the different task types, we also used this to learn how the task types are perceived: After each triple, we let participants answer the statement "Solving these tasks was fun" and "It was easy to solve these tasks". Additionally, we asked how many such tasks they could imagine solving at a time and in a week. We provided drop-down menus in which they could simply select a range (for the first one, in 10 segments, ranging from zero to >40, for the second, 10 segments ranging from zero to >80). All questions that required a response to a statement had to be answered on a 6-point scale to force a decision, labeled from "strongly disagree" to "strongly agree".

We worked with five different conditions: one baseline condition in which no gamification elements are available, and four gamification conditions (all conditions will be explained in more detail in the next section). Up to this point every participant had seen the same content. After the tutorial, participants were allowed to enter the usernames of friends. We stated that certain elements are integrated that might make the solving of tasks more interesting and that some of them can be done with friends. This was necessary as elements such as leaderboards are not comparable across the gamification conditions (which can have an effect on the motivation [2]), and thus were strictly decoupled from each other. When a new participant entered usernames of friends, we ensured that he or she would be assigned to the condition that most of his/her friends were in. If no usernames were provided, we distributed new participants to the conditions equally.

Subsequently, participants received access to the main view, which will also be explained below. A guided tour explained all areas of the interface. We ensured that all tasks were presented in the same sequence for every participant and we also strictly alternated between the three task types. This was done to be able to compare the number of solved tasks across participants, as otherwise ordering effects might be a reason why participants discontinued earlier than others. This seems reasonable for our setup where we wanted to learn specifically how gamification interventions motivate, but it is not advisable in a general crowd-sourcing setting as recent work shows [4, 23]. Participants did not receive any feedback after they had solved a task and we decided against formulating a clear platform goal, such as an upper limit of how many microtasks participants needed to solve. The number of solved tasks, the task correctness rate and the task-solving time represent our dependent variables (with the different conditions being the independent ones). As a consequence, a post-session questionnaire was provided, but was treated as optional. With it, we had the goal to better understand what drove participation. Participants could fill it out anytime they wanted, even though we stated textually that it should be done as soon as they have the feeling they have solved enough tasks. The questionnaire itself consisted of two parts. The first part assessed the intrinsic motivation, by using a German, 12-item 5-point scale version [34] of the Intrinsic Motivation Inventory (IMI) [6] and the second part contained several statements that focus on the perception of the game elements used. The questions in the latter part were adjusted to match the condition the participant was in. Besides two free-text fields, all statements needed to be answered on a 6-point scale, ranging from "strongly disagree" to "strongly agree". Finally, we logged the interactions with the webpage and the tasks in the backend for a quantitative analysis.

3.2 Conditions and System Design

We used five different conditions, which will be explained together with the corresponding system and UI design. With "no gamification" being a baseline condition, in which only the microtasks were

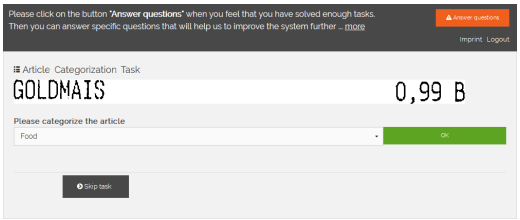

Figure 2: The UI in the no-gamification condition.

presented, the other four conditions modeled different gamification approaches, manipulating the options participants have to modify the game elements. Besides the "bottom-up" condition, in which they could add game elements as they see fit, and a "top-down" condition in which we fixed the set of game elements that were always activated, we also added two conditions that are in the middle of this spectrum: One of the results by Lessel et al. suggested that a pure "bottom-up" approach is potentially not the best option and a mixture of "bottom-up" and "top-down" elements might be worthwhile [16]. Therefore, we added the "selective top-down" condition in which we give participants the option to select which game configuration they want (without giving them the option to edit or add a new configuration) and we added the "selective bottom-up" condition in which game configurations could be selected and edited (without the option to add new elements to configurations).

The microtask area was prominently placed in all conditions and instances of the aforementioned microtask types were presented. In all conditions, participants could skip tasks instead of solving them. Above this area, the button "Answer questions" (leading to the mentioned questionnaires above) was shown with text informing the participants that they could answer specific questions after feeling that they had solved enough tasks. We also stressed that they could continue to solve tasks, even after having answered these questions. The conditions were:

No gamification: In the "no gamification" condition, the participants were only presented with the above-mentioned UI elements (see Figure 2). The number of solved microtasks was not shown. This condition serves as a baseline with no game elements.

"Top-down" gamification: Participants in this condition were confronted with a "top-down" defined game setting. We integrated all elements that were also available in the "gamification" phase of [1]: We awarded 10 points for every solved microtask, a leaderboard showing performance in comparison to others, and two types of badges were awarded (in [1] more badges were possible, but were given for using the household accounting book; we only made use of badges that focus on microtasks). Participants could receive a badge by solving 40 microtasks a day. The other badge type was awarded for solving specific amounts of microtasks, and thereby could reach higher levels ("Crowd Specialist level 1" for solving 60, level 2 for 120, etc.). The leaderboard was initialized with 10 entries (ranging from 0 to 1000 points). To ensure comparability, only participants that were also in the "top-down" condition were shown in this leaderboard. The points were visible at the top and the leaderboard below the microtask area (see Figure 3). The game elements were placed on the right side and the available badges

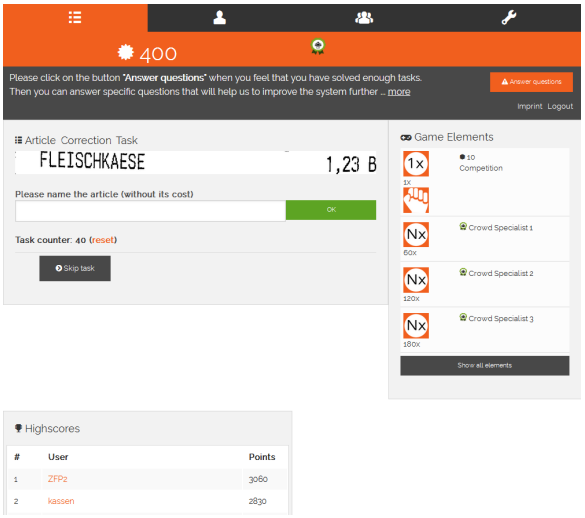

Figure 3: The UI in the "top-down" condition.

could be inspected further. Additionally (similar to all other gamification conditions), participants were shown a header, in which they could inspect their recent badges and points, befriend other people and inspect other profiles. The participants in this condition could not disable, edit or change any of the game elements. All areas were explained by the guided tour after the first login. For the game elements in this condition the tour stated: "*To make the task solving more exciting, you can receive points and achievements. Additionally, below the task area you can see a leaderboard with all users sorted by points. You can also see (after clicking on a user) which badges he/she has already unlocked.*".

"Selective top-down" gamification: In this condition, participants could select which "top-down" defined configuration they wanted to use. The user interface was similar to the one shown in Figure 3, with the only difference being a drop-down menu below "Game Elements" in the right part of the user interface containing ten configuration (see Table 1). After one was selected, we showed a textual explanation of the configuration and the corresponding game elements similarly to the "top-down" condition. Additionally, a help button opens a dialog with an explanation of the game element icons. The initial configuration was selected randomly for every participant, but could be switched anytime. Participants could not decide to not use any game configuration, leading to an always active gamification. The game elements itself were not editable and no new game elements could be added to any configuration in this condition. All configurations that contained leaderboards were treated separately, i.e., points gained in one configuration were not visible on leaderboards in other configurations, as the necessary action to receive the points varies and is thus not comparable. The same is true for badges, which is why we needed to assign them "per configuration" anew (unlocked badges in a configuration remain unlocked if a user decides to come back to this configuration later). All participants in this condition were directly shown with their username on these leaderboards. The configurations were based on commonly used elements [22]. Although discussions about the usage of "Points, Badges and Leaderboards" exist [5], it seemed to

Table 1: Available configurations in the selective conditions.

No.	Name in drop-down	Descriptive text shown on the webpage	Game setting description
01	Points	Receive points for every task solved.	10 points per solved task.
02	Points Leaderboard	Receive points for every task solved and compare yourself with others.	10 points per solved task and performance is shown on a leaderboard.
03	Points Badges	Receive points for every task solved. Additionally, you can unlock badges.	10 points per solved task and access to badges as explained in the "top-down" condition.
04	Points Time pressure	Receive points when you solve a task fast.	10 points per solved task if the solving time was below 5 seconds. A timer is shown in the microtask area and starts at 0 whenever a new task is shown.
05	Points Leaderboard Time pressure	Receive points when you solve a task fast and compare yourself with others.	10 points per solved task if the solving time was below 5 seconds. A timer is shown in the microtask area. Performance is shown on a leaderboard.
06	Points Leaderboard Badges	Receive points for every task solved and compare yourself with others. Additionally, you can unlock badges.	10 points per solved task, performance is shown on a leaderboard and there is access to badges as explained in the "top-down" condition.
07	Points Cooperation	See how many tasks are solved in total and who has contributed how many.	1 point per solved task and a list showing how many tasks were solved by the different participants in this configuration. This was also visible below the microtask area.
08	Points 20 tasks Leaderboard	Receive points whenever you have solved multiple tasks and compare yourself with others.	200 points are assigned for every 20 solved tasks and performance is shown on a leaderboard.
09	Points every 2 minutes	Receive points for every 2 minutes in which you participated.	200 points are assigned for every two minutes the user stays in this condition.
10	Points 20 Tasks Time pressure	Receive points when you solve 20 tasks in under 2 minutes.	200 points are assigned when 20 tasks are solved in under two minutes. A button is shown with which a timer can be started. The timer runs down starting at two minutes.

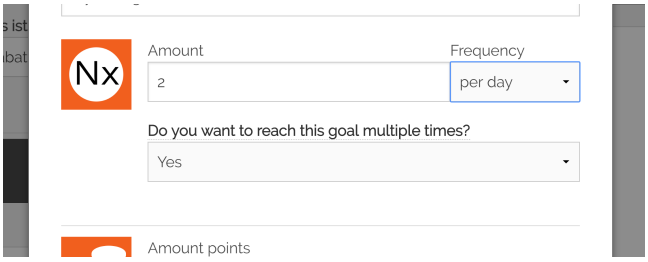

Figure 4: Selected game elements can be modified.

be a good starting point for our investigation. Sets of configurations are comparable (for example 01 and 02 could in theory be compared to see which effect the leaderboard introduces in the combination). All other parts were the same as in the "top-down" condition; the guided tour explained the game elements as follows: "*Here, you can see suggestions for game configurations, with one being always active. Which one depends on you, as you can switch the configurations in the drop-down menu. More information on the different game elements can be found by clicking on the question mark.*".

"Selective bottom-up" gamification: In this condition, participants had the same interface and options as in the "selective top-down" condition. The difference here was that all configurations could be edited further. For configuration 05 (see Table 1) a user could adjust the points from 10 to, for example, 123 and the time from 5 seconds to 11 seconds. The only limitation here was that participants were not able to add further elements to configurations or to create new ones. A small edit icon (being the only difference in the user interface compared to the "selective top-down" condition) was always shown next the active configuration and led to a dialog in which these adjustments can be made (see Figure 4). As consequence of this freedom, leaderboards/cooperations were treated differently to "selective top-down" as points are not necessarily comparable anymore. After changing to a configuration with one of these elements, users could invite friends and only these were shown on the leaderboard/cooperation board. Only the "host" could further customize such configurations, i.e., it is not possible for friends to alter the configuration where they take part by invitation, but they could simply create a new one anytime and invite others by themselves. An important difference in this condition was that participants started without an active game configuration and could be switched to this state whenever they wanted. Thus, participants had the freedom to not use any gamification in this condition. The tour explained the game elements with "*Activate goals, define rewards and play with friends. Here, you can see suggestions for game configurations, with one (or none) being active. You can edit them freely. At the beginning no game elements are active, but if you want to try out configurations, you can select them in the drop-down menu. More information on the different game elements can be found by clicking on the question mark.*".

"Bottom-up" gamification: This condition offers the most freedom for its participants, as game elements could be combined freely. We followed Lessel et al. [16] for implementing the "bottom-up" condition and the UI (see Figure 5) and made the following game elements available: points, badges, leaderboard, cooperative element, task assignments to friends and conditions that unlock the

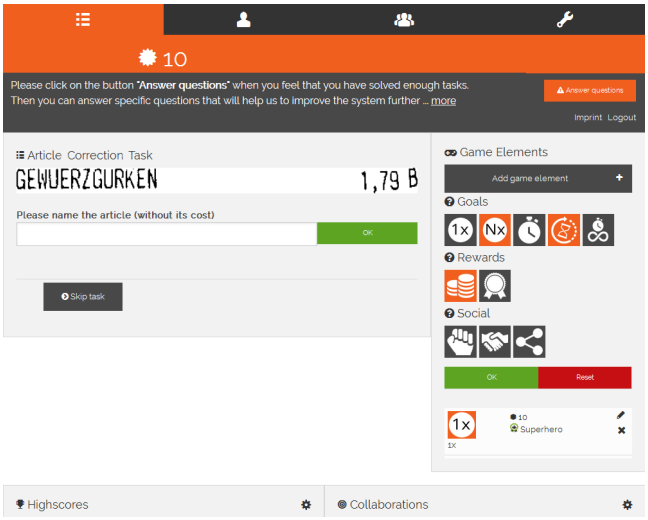

Figure 5: The UI in the "bottom-up" condition.

aforementioned rewards: reaching a specific amount of tasks solved, solving tasks below a certain time, solving tasks for a certain time and receiving the reward periodically (e.g., every 5 minutes). We allowed multiple configurations to be active in parallel and all elements could be edited at any time. For comparability, every configuration that was offered in any of the other gamified conditions could be created here. As explained in the "selective bottom-up" condition, the leaderboards/cooperations also needed to be friends-specific in this condition as well. Thus, for every configuration in which such an element was used, a separate leaderboard/cooperation was created to allow comparisons. Friends could be invited to participate in this configuration (with the same restrictions as already explained above). As several configurations with leaderboards and cooperations could be active in parallel, we also allowed participants to customize which of these are directly visible below the task area (see lower part of Figure 3). Users remain in such leaderboards (whether visible or not) as long as they want, i.e., they could solve tasks and also improve their score for currently-not-shown leaderboards as well. Initially no game elements were configured; thus, participants again had the freedom to not use any gamification at all in this condition. The tour explained the game elements as "*Activate goals, define rewards and play with friends. You can add as many configurations as you want. More information to the different game elements can be found by clicking on the question mark.*".

3.3 Hypotheses

We have the following hypotheses that we wanted to investigate with the presented method and the system:

H1 Participants in the gamified conditions will solve more tasks than participants in the "no gamification" condition.

H2 Participants in the conditions offering customizable gamification ("selective top-down", "selective bottom-up", "bottom-up"), will solve more tasks than in the "top-down" condition.

H1 is supported by the related work in this area, for example [1, 10, 22], showing that gamified interventions in crowd-sourcing are helpful. **H2** is derived from works such as [16, 21, 24] which stress the importance of autonomy in gamified systems. Thus, we would expect that the "selective" and "bottom-up" conditions motivate the participants to solve more tasks.

Table 2: Configuration usage patterns. I: How often the configuration was automatically selected (only in "selective top-down"), V: How often it was selected, but no task solved, A: How often it was selected with tasks solved, T: Number of solved tasks while selected (without V).

No.	Name in dropdown	Selective top-down					Selective bottom-up			
		I	V	A	T_{Avg}	T_{Mdn}	V	A	T_{Avg}	T_{Mdn}
00	No game config.	-	-	-	-	-	4	1	9	9
01	Points	3	5	1	31	32.5	7	3	26.3	24
02	Points Leaderboard	0	7	1	45	45	5	2	6.5	10
03	Points Badges	3	1	1	16	10.5	3	2	163	163
04	Points Time pressure	3	2	2	35.4	42	4	1	8	8
05	Points Leaderboard Time pressure	1	2	1	75.5	75.5	3	1	15	15
06	Points Leaderboard Badges	3	1	2	17	14	1	0	-	-
07	Points Cooperation	1	2	0	5	5	1	2	1.5	1.5
08	Points 20 tasks Leaderboard	2	3	1	58	15	2	0	-	-
09	Points every 2 minutes	1	2	1	6.5	6.5	1	3	1.3	1
10	Points 20 Tasks Time pressure	1	5	2	26.3	16	1	1	24	24

3.4 Results

The study was open for six weeks and 129 participants finished the "warm-up" phase (demographic questions, 3 tutorials, friends), but only 106 participants also solved at least one microtask. Only these will be considered in the results (48 male, 50 female, 8 not reported; age: <21: 13; 21-30: 68, 31-40: 13, >40: 12). 82 characterized themselves as gaming-affine on a 6-point scale (answering with 4 or more). Only two participants befriended each other and only four re-visited the page and continued with solving microtasks.

Game element usages per condition

We analyzed the different conditions in terms of the game configurations or game elements used:

- **No gamification and top-down**: The "no gamification" condition ($n = 23$, tasks solved: $Avg = 47.8$, $SD = 43.5$, $Mdn = 40$) did not offer any game components and the "top-down" condition ($n = 22$, $Avg = 61.6$, $SD = 76.8$, $Mdn = 39.5$) provided no options to change the game elements. For the latter, we have no indication whether the game elements were noticed in general.

Table 3: Tasks solved in the different conditions. Correctness percentages for the different task types relate to the amount of solved tasks in this type. Fixed, Adaptable and Without are revised conditions as discussed in the text.

Condition	Tasks solved									
	Total	*Avg*	*SD*	*Mdn*	*Min*	*Max*	$Correct_{All}$	$Correct_{Classif.}$	$Correct_{Correct.}$	$Correct_{Catego.}$
Top-down (n=22)	1355	61.6	76.8	39.5	2	306	978 (72.1%)	393 (85.6%)	161 (35.6%)	424 (95.5%)
Bottom-up (n=21)	1158	55.1	69.0	32.0	1	256	803 (69.3%)	318 (81.1%)	127 (32.8%)	358 (94.5%)
Selective top-down (n=20)	917	45.9	56.5	26.0	2	230	620 (67.6%)	210 (69.9%)	113 (36.9%)	297 (97.1%)
Selective bottom-up (n=20)	889	44.5	72.2	23.5	2	338	612 (68.8%)	245 (79.5%)	85 (29.6%)	282 (95.9%)
No gamification (n=23)	1100	47.8	43.5	40.0	1	181	801 (72.8%)	311 (82.3%)	148 (40.9%)	342 (95.0%)
Fixed (n=36)	1779	49.4	66.1	25.0	2	306	1244 (69.9%)	479 (80.1%)	201 (33.8%)	564 (96.0%)
Adaptable (n=23)	1892	82.2	87.0	50.0	5	338	1326 (70.1%)	525 (80.6%)	210 (33.9%)	591 (95.0%)
Without (n=47)	1748	37.2	40.0	25.0	1	181	1244 (71.2%)	473 (79.8%)	223 (38.4%)	548 (95.5%)

- **Selective top-down**: Of the 20 participants in this condition, 6 switched the initial configurations at least once. Overall, 42 switches were done by them ($Avg = 7$, $SD = 5.7$, $Mdn = 6$). We analyzed whether people that switched their initial configuration solved more tasks ($Avg = 82.2$, $SD = 75.5$, $Mdn = 60$) than those that stuck with their initial configuration ($Avg = 30.3$, $SD = 40$, $Mdn = 15.5$). A Mann-Whitney U test revealed that the amount of solved tasks is significantly different in these two groups ($U = 72$, $z = 2.48$, $p =< .05$, $r = .55$). Not switching the configuration could hint that participants were not interested in gamification, did not notice these elements or that the initial configuration was already motivating for them. The latter seems unlikely, as the number of solved tasks is lower than in "no gamification" and "top-down" gamification. Concerning the different configurations (cf. Table 1), Table 2 shows usage statistics.
- **Selective bottom-up**: Of the 20 participants in this condition, 8 switched to a game element configuration at least once. Overall, 48 configuration switches were done by them ($Avg = 6, SD = 4.5$, $Mdn = 5.5$). As editing was possible in this condition, we also checked how often an edit was actually done, but only 3 edits happened. Table 2 shows how many tasks were solved while a game configuration was active (the 408 tasks done without a switch from the initial "no game config." were excluded for position 00). We compared the number of tasks solved between participants who activated at least one game configuration ($Avg = 77.8$, $SD = 106.9$, $Mdn = 45$) and those who did not use any ($Avg = 22.3$, $SD = 19.7$, $Mdn = 15.5$) and found a significant difference ($U = 75.5$, $z = 2.13$, $p =< .05$, $r = .48$).
- **Bottom-up**: Of the 21 participants in this condition, 9 used at least one game element. We compared the number of tasks solved between participants in this condition using at least one game element ($Avg = 86.3$, $SD = 85$, $Mdn = 61$) with those who did not ($Avg = 31.8$, $SD = 44.9$, $Mdn = 20$) and found that this differs significantly ($U = 84.0$, $z = 2.14$, $p =< .05$, $r = .47$). Overall, 4 participants only set up their game elements once, 4 used 2 game configurations (i.e., set up game elements, then changed them later on) and one three. Only the cooperative element and the minimum duration time element (i.e., you need to spend a certain time on a task) were not used. Overall, 10 different configurations were created by these 9 participants, indicating that users are quite diverse in what they think motivates them, but once this is selected, users stick with it; otherwise, more configuration switches might have been seen. This is in line with [16]. Overall, the social elements (cooperation, competition, sharing) were not used often, as only two participants were friends in the system.

Table 4: Answers to the statements in the tutorial.

Microtask type	Easy to solve 6-point scale			Fun solving these tasks 6-point scale		
	Avg	*SD*	*Mdn*	*Avg*	*SD*	*Mdn*
Classification	4.8	1.3	5	2.9	1.4	3
Correction	5.3	1.1	6	2.8	1.4	3
Categorization	4.8	1.3	5	3.1	1.5	3

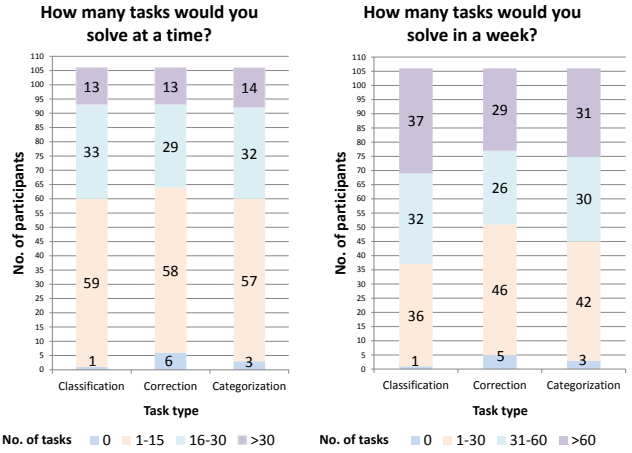

Figure 6: How many tasks participants could imagine (after the tutorial, but before having solved them in the main UI) solving at a time and in a week.

Tasks across conditions

Table 3 shows that participants made the most errors for *Correction* tasks. In contrast, the task was perceived as significantly easier than the other two task types (see left side of Table 4), as a Friedman ANOVA with pairwise comparison with Bonferroni-corrected p-values showed ($\chi^2(2) = 17.06, p < .01$, *Classification* and *Correction*: $p < .05$; *Categorization* and *Correction*: $p < .05$). One explanation for this was found in the free-text answers the participants could provide in the closing questionnaire: 10 participants mentioned that it was hard to solve correction tasks, as it was not totally clear what the correct solution would have been. Concerning "fun" (see right side of Table 4), no task type is particularly rewarding for most

of the participants (and no significant difference could be found with a Friedman ANOVA, $p = .53$). 8 participants also reported in their free-text answers that the tasks themselves were boring and variety was lacking (interestingly, 5 of these 8 participants were in the "top-down" condition). Nonetheless, Figure 6 shows that the participants stated they were willing to solve microtasks in general by solving a fair amount of them. Table 3 shows how many tasks were solved by the participants in every condition. We conducted a Kruskal-Wallis ANOVA, but were not able to find a significant difference by the amount of solved tasks and condition ($p = .921$). In the previous section, we showed that many participants actually were in a gamification condition, but did not use game elements. Even though framing effects might have motivated them [17], the low number of solved tasks (as reported before) indicated that this was unlikely. Therefore, we clustered the participants into revised conditions (see Table 3 for the amount of tasks solved in these groups), according to the following schema:

- **Fixed gamification (Fixed)**: As we have no indication whether participants in the "top-down" condition actually noticed game elements or not, they remain in their own group. We additionally add the participants that were in "selective top-down" but did not switch their initial group (see above for potential reasons why they might not have switched) and thereby were also in a "top-down" setting ($n = 36$).
- **Adaptable gamification (Adaptable)**: All participants of the "bottom-up" condition who set up game elements and solved tasks, participants in "selective top-down" who switched their configuration at least once and participants in "selective bottom-up" who used at least one configuration were clustered into this group. This group represents "bottom-up" concepts, with different degrees of freedom, but overall, all these participants were able to adapt the gamification at runtime and did so, i.e., they made use of their choice ($n = 23$).
- **Without gamification (Without)**: Participants in "bottom-up" who did not set up any game elements, participants in "selective bottom-up" who did not switch to any game configuration and participants that were already in the "no gamification" condition were clustered into this group ($n = 47$). The participants solved tasks without any gamification being active.

A Kruskal-Wallis test showed that the amount of solved tasks differs significantly between these new groups, $H(3) = 10.51$, $p < .05$. A pairwise comparison with Bonferroni-corrected p-values showed that the groups "without" and "adaptable" ($p < .05$, $r = -.34$) and "fixed" and "adaptable" ($p < .05$, $r = .46$) differ significantly. This hints that in conditions in which a user has a choice how he or she wants to use gamification, and actually uses this choice, more tasks are solved. We also compared the ratio of correct to incorrect tasks across the (initial and revised) conditions, but were not able to find a significant difference with a Kruskal-Wallis test ($p = .66$ and respectively $p = .65$). Analyzing the time spent to solve single tasks in the different conditions and task types with the same test reveals no significant differences in the initial conditions. The revised conditions, on the other hand, revealed significant differences for the overall timings ("fixed": M=10.2s, Mdn=7.1s, "adaptable": M=7.8s, Mdn=5.1s, "without": M=9.7s, Mdn=6.8s) at the $p < .05$ level. A pairwise comparison with Bonferroni-corrected p-values showed that participants in the "adaptable" condition spent less time per task than in the "fixed" and the "without" condition. The adaptable gamification approaches seem to have motivated the participants to solve tasks faster, without a significant loss of correctness.

Table 5: Aggregated "a-priori" estimations of how many tasks participants want to solve separated by condition.

Condition	How many tasks do you want to solve... ...at a stretch			...in a week		
	Avg	*SD*	*Mdn*	*Avg*	*SD*	*Mdn*
Top-down (n=22)	59.2	44.6	42	143.0	95.8	150
Bottom-up (n=21)	46.6	32.6	39	126.0	83.7	105
Sel. top-down (n=20)	49.3	37.6	34	129.8	80.1	112
Sel. bottom-up (n=20)	30.9	20.5	27	99.8	79.9	90
No gamification (n=23)	46.6	22.7	49	141.1	68.9	145
Fixed (n=36)	53.4	42.1	34	132.5	89.0	113
Adaptable (n=23)	48.6	33.8	39	130.7	84.0	115
Without (n=47)	40.5	24.2	39	124.5	76.9	125

Table 6: Results of the short Intrinsic Motivation Inventory (sub-scales were based on 5-point scales).

Condition	Intrinsic Motivation Inventory							
	Interest		Comp.		Choice		Pressure	
	Avg	*SD*	*Avg*	*SD*	*Avg*	*SD*	*Avg*	*SD*
Top-down (n=12)	2.3	1.0	3.6	0.7	2.6	1.0	1.5	0.5
Bottom-up (n=15)	2.3	0.9	3.8	0.8	3.2	1.0	2.1	0.8
Sel. top-down (n=12)	2.7	1.0	3.6	0.7	3.3	0.9	2.3	0.9
Sel. bottom-up (n=13)	2.7	1.1	3.8	0.8	3.4	1.1	1.6	1.0
No gamification (n=19)	2.5	0.9	3.8	0.8	3.4	1.0	1.6	0.8
Fixed (n=18)	2.4	0.9	3.5	0.6	2.7	0.9	1.8	0.8
Adaptable (n=21)	2.6	1.0	3.9	0.7	3.4	0.9	2.0	0.8
Without (n=32)	2.6	1.0	3.7	0.8	3.3	1.0	1.9	0.9

The significant differences found in the "adaptable" condition might also be explainable by having users that were in general more open to participate *and thus* also tried out the gamification as options *instead of* the gamification being the driving factor for solving more tasks. To rule this causality effect out, we analyzed the "a priori" reported amount of tasks participants could imagine to solve (cf. Figure 6) by calculating the "a-priori" task solving count per participant and considered them condition-wise. Table 5 shows the results: even though the median values in "top-down" and "no gamification" are higher (i.e., participants that were later assigned to these conditions reported to want to solve more tasks), Kruskal-Wallis ANOVAs found no significant difference in the initial ($p = .18$ for "at a time" and $p = .42$ for "in a week") or revised conditions ($p = .67$ and $p = .93$). We also asked how *relevant* the topic of digital recordings of receipts is in the closing questionnaire. From the 70 participants that answered it, no significant difference could be found in the initial ($p = .14$) or the revised conditions ($p = .26$), but a trend was visible: The lowest values (i.e., finding it less relevant) can be found in "bottom-up" ($Avg = 2.9$, $Mdn = 2$) and "adaptable" ($Avg = 3.3$, $Mdn = 3$); the highest in "top-down" ($Avg = 4.5$, $Mdn = 5$) and "fixed" ($Avg = 4.2$, $Mdn = 5$). This hints that participants in "bottom-up"/"adaptable" were not more motivated by the topic.

Additionally, we considered the answers to the IMI, which 71 participants answered (cf. Table 6). We conducted Kruskal-Wallis ANOVAs with the initial and revised conditions and the respective

values the IMI reports. None of these values differs significantly across the initial/revised conditions (interest: $p = .77/p = .80$, perceived competence: $p = .63/p = .13$, perceived choice: $p = .24/p = .05$, pressure: $p = .16/p = .35$). Not finding a significant difference for *perceived competence* is expected as the task (and the task selections) do not change during the experiment and also could not be adjusted with the game elements. The non-significant difference in *perceived choice* is surprising. Considering the average values reported in Table 6 (as $p < .05$ is almost reached), we see that in the revised conditions "fixed" seems most restrictive. The result for *pressure* could be explained by the absence of clear goals such as "you need to solve at least X microtasks"; thus, no source of pressure was available. As no group excels in the values for *interest*, the tasks seem to be perceived as equally uninteresting.

Taking these aspects together, we would reject the assumption that participants in the "adaptable" condition simply solved more tasks because they were more open to the task or the topic itself. Thus, the significantly higher amount of solved microtasks in this condition provides evidence to support **H2**. The statements "the game elements led to more fun" ("adaptable": $Avg = 3.8$, $Mdn = 4$; "fixed": $Avg = 3.0$, $Mdn = 3$; "without": $Avg = 2.7$, $Mdn = 2$; Kruskal-Wallis test: $p = .134$) and "the game elements motivated me to solve tasks more efficiently" ("adaptable": $Avg = 3.6$, $Mdn = 4$; "fixed": $Avg = 2.8$, $Mdn = 3$; "without": $Avg = 2.5$, $Mdn = 2$; Kruskal-Wallis test: $p = .068$) provided further support for this as the "adaptable" condition had the highest value here.

The "top-down" condition, as well as the gamification configuration initially selected automatically and unchanged in "selective top-down", did not motivate the participants to solve more tasks in comparison to participants in the "no gamification" baseline. This stands in contrast to the related work, in which gamification led to an increased performance (see related work section). Statements regarding the meaningfulness of gamification in this setting, and that game elements motivated them, were subjectively disagreed to (with a median of 3 or 4). In general, this hints that for this population and this microtask setup the game elements were potentially not suitable, which might explain why the performance in the "top-down"/"fixed" conditions was not better than in the "no gamification"/"without" condition. We can thus only partially (as "adaptable" indeed was better then "without") confirm **H1**.

3.5 Discussion

The goal of this study was to investigate whether "bottom-up" gamification is able to compete with or exceed "top-down" gamification: After re-grouping conditions the participants were in, we were able to identify differences (as hypothesized) in favor of "bottom-up" gamification. Participants that had a choice in how they wanted to use gamification (the selective conditions and "bottom-up") and used it subsequently performed significantly better as they solved more tasks faster without a decrease in correctness. We also found indications that they were not simply *more engaged overall* and therefore also tried out different game elements, but instead that "bottom-up" gamification was most likely the explanation for why they performed better. One disadvantage of the study design used is that it is currently unclear whether this difference can be accounted for by the choices offered or by the fact that participants "simply" selected a motivating and fitting game configuration. The latter would mean that if participants were provided with their selected configuration in a "top-down" manner, this would have also led to positive results. We do not see this as critical, as in both cases it provides arguments in favor of "bottom-up" gamification: If the choice is what motivates participants, then providing choices for selecting which game elements should be used in gamified systems is reasonable. If in contrast the self-selected gamification configuration mattered, then again, offering participants a choice to select from a set of configurations or build their own configuration (as can both be done in "bottom-up" gamification) seems reasonable as well. In both cases, the question of "can users select the best fitting game elements for themselves" is relevant and will be studied further.

An important question is why 62% of the participants did not use the offered choices. Explanations for this could be that participants were not motivated by game elements in general, did not understand or comprehend their offered options or did not see why they should use game elements at all. For this, the framing of the study (assisting researchers by doing simple tasks without any reward) might have been an explanation: As participants expected to remain on this page for only a short time, they may not have seen why they should also spend time on a peripheral feature, and instead simply focused on the task solving. Another explanation might be that certain users are not attracted by the capability to create their own gamification setup (e.g., as they might dislike the additional effort) and simply do not want to try it. Testing "bottom-up" gamification in different scenarios will provide more insights into these issues. It was also unexpected to find no difference between the "top-down" gamification setting and the setting without gamification, as the same intervention had led to more task solving in the work of Altmeyer et al. [1]. But there were differences in the setup: Participants of Altmeyer et al. had a direct benefit as the microtasks directly improved the algorithm for their digital household accounting book and they could also see how their receipt parts got corrected over time. Second, as stated before, the study frame might again be an explanation. Participants did not resume the work later. The household accounting book, on the other hand, was used several times during a week[1] and differences in the game-element perception might become visible only after a longer time.

3.6 Limitations

Our study had limitations: First, the number of participants, especially in the presence of five conditions, can be seen as a limiting factor. Second, the amount of options in the "selective" and "bottom-up" conditions made the study design rather complex and resulted in several small "bins" (see for example Table 1). [16] showed, though, that systems need a broad range of different game elements to offer a choice for every participant. Reducing the game elements to only a few might result in a situation in which some participants, even when having a choice, could not find a setup that was beneficial for them, which then would again confound the results. To reduce this complexity, follow-up studies could for example pre-select game elements fitting every player type [31] and present only these to the corresponding participants to reduce the complexity. Another

[1] It was app-based and tasks could be done whenever participants were idle. Our system was browser-only and not optimized for mobile phones.

option would then be to do a within-subject design in which all combinations of the pre-selected game elements are tested and in one further condition providing participants with a choice to set up the combination on their own. This would allow us to narrow down whether choice is what matters. Third, the study frame and microtasks context might have reduced the overall effect of "bottom-up". The task themselves were perceived as boring and lacking much variety. This might be more demotivating than the gamification can compensate for. Thus, it will also be interesting to investigate a setting in which the task is already interesting for the participants. Fourth, the lack of social connections (by having no friends in the system) is an issue for people that are motivated by, for example, competitive elements. Fifth, although discussed above, it should again be highlighted that the number of participants that did not use their choices and the missing effect of the "top-down" setting hint that the study setup was not perfect. Thus, even though our results provide a good starting point, more research is necessary.

4 CONCLUSION & FUTURE WORK

This paper presented a study to investigate the effectiveness of "bottom-up" gamification where users can set up their own gamification at runtime in a microtask setting. We found that participants who actually had a choice and used it solved significantly more tasks than participants with a fixed gamification setting or no gamification at all. To our knowledge, we are the first that explored this question. As we discussed certain shortcomings of our study, our results should not be overestimated, but they are a first indication that "bottom-up" gamification can lead to a higher motivational impact than the commonly used fixed gamification.

Several connecting factors for future work, besides the ones mentioned in the paper, are: First of all, conducting further studies to find out whether the actual choice or the self-selection of a fitting game configuration is the motivating factor seems a reasonable next step. Although both support the "bottom-up" idea, understanding this is important towards this concept. The effect of "bottom-up" in different contexts is another research perspective that helps to further understand the relevance of this concept. Furthermore, as this paper did not investigate this, exploring the ramifications of self-selected elements (such as effort to set up such configurations) and how "bottom-up" performs long-term will help to understand the drawbacks that might be introduced by following such a concept.

REFERENCES

[1] Maximilian Altmeyer, Pascal Lessel, and Antonio Krüger. 2016. Expense Control: A Gamified, Semi-Automated, Crowd-Based Approach For Receipt Capturing. In *Proc. IUI 2016*. ACM, 31–42. https://doi.org/10.1145/2856767.2856790

[2] Jason T. Bowey, Max V. Birk, and Regan L. Mandryk. 2015. Manipulating Leaderboards to Induce Player Experience. In *Proc. CHI PLAY 2015*. ACM, 115–120. https://doi.org/10.1145/2793107.2793138

[3] Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's Mechanical Turk a New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6, 1 (2011), 3–5.

[4] Carrie J. Cai, Shamsi T. Iqbal, and Jaime Teevan. 2016. Chain Reactions: The Impact of Order on Microtask Chains. In *Proc. CHI 2016*. 3143–3154.

[5] Yu-kai Chou. 2014. Points, Badges, and Leaderboards: The Gamification Fallacy. Internet. (2014). Retrieved August 21, 2017 from http://yukaichou.com/gamification-study/points-badges-and-leaderboards-the-gamification-fallacy.

[6] Edward L. Deci and Richard M. Ryan. 2003. Intrinsic Motivation Inventory. Internet. (2003). Retrieved August 21, 2017 from http://selfdeterminationtheory.org/intrinsic-motivation-inventory.

[7] Edward L Deci and Richard M Ryan. 2011. Self-determination theory. *Handbook of Theories of Social Psychology* 1 (2011), 416–433.

[8] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proc. MindTrek 2011*. ACM, 9–15. https://doi.org/10.1145/2181037.2181040

[9] Lauren S. Ferro, Steffen P. Walz, and Stefan Greuter. 2013. Towards Personalised, Gamified Systems: An Investigation into Game Design, Personality and Player Typologies. In *Proc. IE 2013*. ACM, 7:1–7:6. https://doi.org/10.1145/2513002.2513024

[10] Oluwaseyi Feyisetan, Elena Simperl, Max Van Kleek, and Nigel Shadbolt. 2015. Improving Paid Microtasks Through Gamification and Adaptive Furtherance Incentives. In *Proc. WWW 2015*. ACM, 333–343.

[11] Ido Guy, Anat Hashavit, and Yaniv Corem. 2015. Games for Crowds: A Crowdsourcing Game Platform for the Enterprise. In *Proc. CSCW 2015*. ACM, 1860–1871. https://doi.org/10.1145/2675133.2675189

[12] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *Proc. HICSS '14*. IEEE Computer Society, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

[13] Carrie Heeter, Brian Magerko, Ben Medler, and Yu-Hao Lee. 2011. Impacts of Forced Serious Game Play on Vulnerable Subgroups. *International Journal of Gaming and Computer-Mediated Simulations* 3, 3 (July 2011), 34–53.

[14] Shih-Wen Huang, Pei-Fen Tu, Wai-Tat Fu, and Mohammad Amanzadeh. 2013. Leveraging the Crowd to Improve Feature-Sentiment Analysis of User Reviews. In *Proc. IUI 2013*. ACM, 3–14. https://doi.org/10.1145/2449396.2449400

[15] Masatomo Kobayashi, Shoma Arita, Toshinari Itoko, Shin Saito, and Hironobu Takagi. 2015. Motivating Multi-Generational Crowd Workers in Social-Purpose Work. In *Proc. CSCW 2015*. ACM, 1813–1824.

[16] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2016. "Don't Whip Me With Your Games": Investigating "Bottom-Up" Gamification. In *Proc. CHI 2016*. ACM, 2026–2037.

[17] Andreas Lieberoth. 2015. Shallow Gamification: Testing Psychological Effects of Framing an Activity as a Game. *Games and Culture* 10, 3 (2015), 229–248.

[18] Winter Mason and Duncan J. Watts. 2009. Financial Incentives and the "Performance of Crowds". In *Proc. HCOMP '09 SIGKDD Workshop on Human Computation*. ACM, 77–85. https://doi.org/10.1145/1600150.1600175

[19] Elaine Massung, David Coyle, Kirsten F. Cater, Marc Jay, and Chris Preist. 2013. Using Crowdsourcing to Support Pro-Environmental Community Activism. In *Proc. CHI 2013*. ACM, 371–380. https://doi.org/10.1145/2470654.2470708

[20] Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Disassembling Gamification: The Effects of Points and Meaning on User Motivation and Performance. In *Proc. CHI EA 2013*. ACM, 1137–1142. https://doi.org/10.1145/2468356.2468559

[21] Ethan R. Mollick and Nancy Rothbard. 2013. Mandatory Fun: Gamification and the Impact of Games at Work. *The Wharton School Research Paper Series* (2013).

[22] Benedikt Morschheuser, Juho Hamari, and Jonna Koivisto. 2016. Gamification in Crowdsourcing: A Review. In *Proc. HICSS '16*. IEEE Computer Society, 4375–4384.

[23] Edward Newell and Derek Ruths. 2016. How One Microtask Affects Another. In *Proc. CHI 2016*. 3155–3166. https://doi.org/10.1145/2858036.2858490

[24] Scott Nicholson. 2012. A User-Centered Theoretical Framework for Meaningful Gamification. *Games+Learning+Society* 8, 1 (2012).

[25] Rita Orji, Julita Vassileva, and Regan L. Mandryk. 2014. Modeling the Efficacy of Persuasive Strategies for Different Gamer Types in Serious Games for Health. *User Modeling and User-Adapted Interaction* 24, 5 (2014), 453–498.

[26] Chad Richards, Craig W. Thompson, and Nicholas Graham. 2014. Beyond Designing for Motivation: The Importance of Context in Gamification. In *Proc. CHI Play 2014*. ACM, 217–226. https://doi.org/10.1145/2658537.2658683

[27] Ganit Richter, Daphne R. Raban, and Sheizaf Rafaeli. 2015. *Studying Gamification: The Effect of Rewards and Incentives on Motivation*. Springer International Publishing, Cham, 21–46.

[28] Katie Seaborn and Deborah I. Fels. 2015. Gamification in Theory and Action: A Survey. *International Journal of Human-Computer Studies* 74 (2015), 14–31.

[29] James Surowiecki. 2005. *The Wisdom of Crowds*. Anchor.

[30] Jaime Teevan, Daniel J. Liebling, and Walter S. Lasecki. 2014. Selfsourcing Personal Tasks. In *Proc. CHI EA 2014*. ACM, 2527–2532.

[31] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. The Gamification User Types Hexad Scale. In *Proc. CHI PLAY 2016*. 229–243. https://doi.org/10.1145/2967934.2968082

[32] Khai N. Truong, Thariq Shihipar, and Daniel J. Wigdor. 2014. Slide to X: Unlocking the Potential of Smartphone Unlocking. In *Proc. CHI 2014*. ACM, 3635–3644.

[33] Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. 2008. reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science* 321, 5895 (2008), 1465–1468.

[34] Matthias Wilde, Katrin Bätz, Anastassiya Kovaleva, and Detleff Urhahne. 2009. Testing a Short Scale of Intrinsic Motivation. *Zeitschrift für Didaktik der Naturwissenschaften* 15 (2009), (31–35).