Do Points, Levels and Leaderboards Harm Intrinsic Motivation? An Empirical Analysis of Common Gamification Elements

Elisa D. Mekler[1], **Florian Brühlmann**[1], **Klaus Opwis**[1], **Alexandre N. Tuch**[1 2]

[1]Center for Cognitive Psychology and Methodology
University of Basel
Switzerland
{elisa.mekler,klaus.opwis}@unibas.ch
florian.bruehlmann@stud.unibas.ch

[2]Dpt. of Computer Science
University of Copenhagen
Denmark
a.tuch@unibas.ch

ABSTRACT
It is heavily debated within the gamification community whether specific game elements may actually undermine users' intrinsic motivation. This online experiment examined the effects of three commonly employed game design elements – points, leaderboard, levels – on users' performance, intrinsic motivation, perceived autonomy and competence in an image annotation task. Implementation of these game elements significantly increased performance, but did not affect perceived autonomy, competence or intrinsic motivation. Our findings suggest that points, levels and leaderboards by themselves neither make nor break users' intrinsic motivation in non-game contexts. Instead, it is assumed that they act as progress indicators, guiding and enhancing user performance. While more research on the contextual factors that may potentially mediate the effects of game elements on intrinsic motivation is required, it seems that the implementation of points, levels, and leaderboards is a viable means to promote specific user behavior in non-game contexts.

Author Keywords
Gamification; gameful design; motivation; game design elements

ACM Classification Keywords
H.5.2 Informations interfaces and presentation: User interfaces

General Terms
Human Factors

INTRODUCTION
Digital games have become increasingly popular over the last few years [1], with many players investing countless hours in gaming [17, 19]. Industry professionals have taken notice of this trend and have attempted to apply games' motivational appeal to various non-gaming contexts to foster user engagement. This practice is nowadays best known under the moniker "gamification", commonly defined as *the use of game design elements in non-game contexts* [8], and has become a heavily debated subject in its own right [7, 20].

Most prominently, gamification has been commonly associated with points, levels and leaderboards [8, 27] – "the things that are least essential to games" [20], – which has irked game designers and psychologists alike. Some have cautioned against the over-reliance on such elements, as they may diminish intrinsic interest in both game- and non-game contexts, ultimately leading users to stop interacting with the application or service altogether [6, 11]. In fact, previous research in psychology provides ample evidence that different forms of rewards, feedback, and other external events can have detrimental effects on intrinsic motivation (for an overview see [4]). However, there is still a lack of empirical evidence on whether and under what circumstances these game elements may actually undermine users' intrinsic motivation [11].

While several studies have already examined how game elements, such as points, levels [9, 23] or badges [5, 10], affect user behavior, to our knowledge, none have looked yet into their impact on intrinsic motivation. This makes it difficult to infer how game elements affect users' intrinsic motivation and behavior in non-game contexts. Yet, this issue is highly relevant to gamification designers and researchers alike, as these elements have been and continue to be applied to a broad spectrum of non-game contexts with varying degrees of success and are still widely considered "[...] the heart of any gaming system" ([27], pp. 36).

The present paper aims to address the aforementioned research gaps by investigating how three of the most commonly employed game elements – points, leaderboards, levels, – affect users' behavior and intrinsic motivation. By no means do we claim that the implementation of these game elements form *good* or *bad* examples of gamification. Rather, we believe that the prevalence of these game elements in many gamified applications warrants a closer examination of their effects, to form a clearer understanding of when their imple-

Gamification'13, October 02 - 04, 2013, Stratford, ON, Canada.

http://dx.doi.org/10.1145/2583008.2583017

mentation may prove beneficial or harmful to user engagement.

RELATED WORK

Points, levels and leaderboards

Points, levels and leaderboards form three of the most basic game patterns [26, 27]. Zagal et al. categorize them as *goal metrics*, as all three are used to keep track of and provide feedback on player performance in games [26]. Due to their apparent connection to digital games and due to them being readily applicable to various non-game contexts, points, levels and leaderboards have become the poster children of gamification (e.g., [27]). Von Ahn and Dabbish, for example, consider them essential to increase enjoyment in human computation tasks [24]. However, research on the effectiveness and potential side effects of goal metrics in non-game contexts is still few and far between.

Farzan et al. notably studied the potential of a point-based incentive system (i.e., points, "status" levels and a leaderboard) to promote user activity in an enterprise social networking site [9]. Indeed, user activity initially increased, compared to the control group that was not presented with such game elements. However shortly after launch, user activity reverted back to baseline, and after the game elements were removed, user activity even dropped below to what it had been before implementation of the incentive system [23]. In fact, some users hinted at feeling driven by the leaderboard to keep up with other users [9]. This further suggests that the incentive system extrinsically motivated users to engage with the social network, as after the removal of the game elements, they were less inclined – in other words, intrinsically motivated, – to do so.

Intrinsic Motivation

Ryan and Deci differentiate two forms of motivation [21]: *Intrinsic motivation* denotes the pursuit of an activity, because it is inherently interesting or enjoyable, whereas *extrinsic motivation* is defined as doing something due to a separable outcome, such as money or deadlines. However, such extrinsic incentives have been found to reduce intrinsic motivation in various contexts [4]. In other words, a person is no longer intrinsically drawn towards engaging in an activity, because s/he is pushed to do so through external means. This is unfortunate, as intrinsic motivation is not only associated with improved psychological well-being, but also benefits the extent and quality of effort that people put into a given task, which results in enhanced performance, creativity and learning outcomes in a variety of domains [21].

External events, such as feedback and rewards may impact intrinsic motivation in different ways, depending on whether they are perceived as informational or controlling [4]. Put differently, the effects of external events on intrinsic motivation are mediated by a person's perception of how these events influence the need for competence and autonomy. Competence signifies the perceived extent of one's own actions as the cause of desired consequences in one's environment [21] and thrives when met with direct and positive feedback. But if perceived as controlling, even positive feedback may thwart people's inherent need for autonomy and hence, decrease intrinsic motivation [4], whereas feedback that is perceived as both noncontrolling and informational, supports people's need for competence and subsequently boosts their intrinsic motivation (see Figure 1).

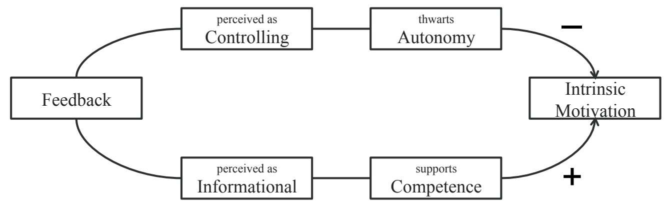

Figure 1. Feedback may be perceived as controlling or informational, thereby affecting need satisfaction and intrinsic motivation in different ways. (Figure adapted from [4, 7])

Aim of the study

Points, levels and leaderboards are commonly considered a form of extrinsic incentive in non-game contexts (e.g., [14]) and may therefore threaten users' intrinsic motivation to engage with a gamified system. However, as they often transmit a form of positive feedback, it is imaginable that they could enhance feelings of competence, and therefore increase intrinsic motivation [7, 12]. While the aforementioned studies point towards a possible detrimental effect of game elements on intrinsic motivation [9, 23], this might have been due to users feeling pressured to engage in the social networking service of their *employer* [3]. The incentive system may have simply exacerbated this, as users' activity or lack thereof was made apparent to their co-workers.

As such situational factors may mediate how points, levels and leaderboards affect intrinsic motivation [6], it is important to examine their effects in different non-game contexts, in order to gain a better understanding of when the implementation of these game elements may or may not harm users' intrinsic motivation. By investigating the effects of goal metrics on both user performance and intrinsic motivation, as well as on autonomy and competence need satisfaction, we wish to learn more about whether they are invariably perceived as controlling.

The present study aims to expand upon existing research by investigating how goal metrics affect user behavior and intrinsic motivation in an image annotation task. This non-game context was deemed suitable for several reasons. First, human computation tasks (e.g., image annotation) are often engaged in voluntarily, for fun or for pastime [2]. Thus, the threat of contextual factors acting as confounding influence (e.g., the workplace being a potentially controlling setting) is minimized. Secondly, von Ahn and Dabbish explicitly state that the aforementioned game elements increase user motivation in human computation contexts [24]. However, to our knowledge, no actual empirical evidence to back this claim exists, which provides an additional motivation for the present study.

In line with previous findings on the effects of game design elements on user behavior [5, 9, 10], we formulate the following hypothesis:

H1: *Points, levels and leaderboards significantly boost performance in the image annotation task, compared to the control condition.*

As there exists no definite previous scientific evidence on the effects of points, levels and leaderboards on intrinsic motivation in non-game contexts, the following hypotheses are based on existing research on the effects of rewards and feedback in educational settings [4, 21]. Thus, if these game elements are to be considered a form of extrinsic reward, we posit that:

H2: *Points, levels and leaderboards significantly decrease autonomy need satisfaction and intrinsic motivation, compared to the control condition.*

METHOD

To test our hypotheses, we conducted a between-subject online experiment. The independent variable were three of the most common game elements: Points vs. leaderboard vs. levels vs. control condition. The dependent variables were user performance (amount of tags, "cheating behavior", time spent on task), intrinsic motivation and satisfaction of autonomy and competence needs.

Materials

Image Tagging Platform

The image annotation task consisted of 15 abstract paintings that were taken from Machajdik and Hanbury's study on affective image classification [15]. In order to control for social factors, a single player image tagging platform was designed, loosely modeled after the one created by Wang and Yu [25]. An image was presented for 5 seconds, before flipping over and revealing the input area, where participants could enter their tags.

In the *control condition*, no game design elements were present and the right-hand side of the screen was left blank.

In the *points* condition, participants earned 100 points for each tag they entered. The current score was displayed in the upper right corner of the screen (see Figure 2 and Figure 3). Points had no further meaning, other than depicting how many tags a participant had generated. After completing all 15 images, participants were presented with their final score.

In the *leaderboard* condition, participants could compare their current score to four fictitious participants in a leaderboard on the right-hand side of the screen (see Figure 2). Participants were deliberately left unaware of the fact that fictional participants occupied the leaderboard. This static leaderboard was implemented so that all participants had the same chance to rise in ranks, as the leaderboard posititioning may have had a confounding effect on motivation otherwise [24]. To reach the lowest position on the leaderboard, participants had to generate at least ten tags. For each subsequent position, participants had to generate even more tags.

Put differently, the four competing players had a score of 1000, 3000, 6000, and 10000 respectively. These step sizes were chosen to allow participants to reach a reasonably high position on the leaderboard, but it was expected to be still reasonably challenging for participants to come up with more than 100 tags.

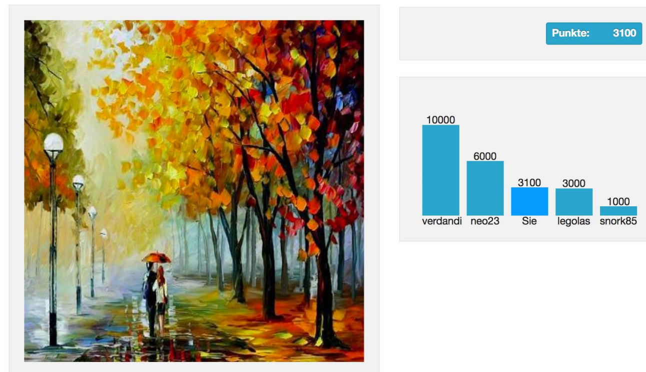

Figure 2. Screenshot of the tagging platform with points and leaderboard

In the *levels* condition, participants were presented with a vertical progress bar labeled with "next level" and the corresponding points necessary to reach the indicated level (see Figure 3). Progression to the next level mirrored the leaderboard condition, albeit without the option for (seemingly) social comparison. Whenever participants reached a score of 1000, 3000, 6000, 10000 and finally 15000, they would gain another level symbolized by an asterisk.

Measurements

Performance was measured by tracking the amount of tags generated per participant. Additionally, the amount of time participants spent on the image annotation task was tracked. According to von Ahn and Dabbish, the *throughput* (i.e., the number of tags generated per human-hour) and the overall time spent on the task determine whether the gamification of a human computation task was successful [24]. *Intrinsic motivation* (α = .95, range .93 - .96) and satisfaction of the *autonomy* (α = .68, range .65 - .74) and *competence* needs (α = .86, range .84 - .88) were assessed with the Intrinsic Motivation Inventory (IMI) [22] (7-point Likert scale, 1 = not at all true, 7 = very true).

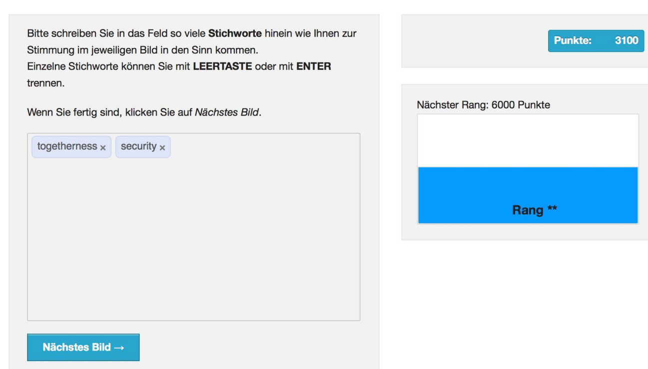

Figure 3. Screenshot of the input area with points and levels

Participants

Participants were recruited by e-mail from the university's own database, where people may sign up, if they wish to participate in studies. A total of 295 participants (93 male, 191 female, 11 not specified; mean age 32.85 years (SD = 12.33), range 17-68 years) completed the online study. Five \$50 gift coupons for an online consumer electronics retailer were raffled among all participants. The raffle was deliberately chosen as incentive, because it was assumed that it would not distort the experimental effects of game elements on intrinsic motivation, due to being a form of unexpected, task-noncontingent reward. In their meta-analysis, Deci, Koestner and Ryan found that task-noncontingent rewards do not affect intrinsic motivation, as these rewards do not require doing or completing the task and hence are not perceived as controlling [4].

Procedure

Upon clicking the invitation link to the study, participants were randomly assigned to one of the four experimental conditions. Following a brief demographic questionnaire, they were then introduced to the image annotation task and informed that their tags would help improve affective image categorization. In order to isolate the effect of game elements on intrinsic motivation, special care was taken to ensure that the study description did not contain any wording that might be perceived as controlling (e.g., "you must", "you should") [22]. A test trial consisting of three images, which was the same for every condition with no game elements displayed, preceded the actual experiment.

Before starting the actual experiment, participants' attention was drawn towards the game elements, except for the control condition. Again, because the focus of this experiment was to examine the effects of points, levels and leaderboards and not the task context per se, we made sure that task instructions were worded as noncontrolling as possible, in order to avoid detrimental effects on intrinsic motivation [4]. In the points and level conditions, participants were informed that their score and level would help them estimate their contribution to the study. In the leaderboard condition, participants were told that they had the option to compare themselves to other participants.

Images were presented in random order. After completing the image annotation task, participants in the game element conditions were presented their final score, level or position on the leaderboard. Additionally, participants in the leaderboard condition had the option to enter a nickname on the leaderboard. Afterwards, all participants filled in the IMI [22] and had the option to comment on the study. Overall, participants took on average around 22 minutes to complete the study.

RESULTS

In order to investigate the effects of points, levels and leaderboard on user performance, intrinsic motivation and need satisfaction, analyses of variance (ANOVA) were calculated, unless otherwise noted. To assure homogeneity of variance, data were square-root transformed. For all statistical tests an alpha level of .05 was used.

Performance

Tag quality was determined by matching all generated tags with a German dictionary consisting of over 1.3 million entries (http://germandict.sourceforge.net/). All nonsensical tags and articles (e.g., the) were discarded from subsequent analyses. As participants could receive points even for nonsensical tags, we first checked whether conditions differed in cheating behavior (the use of nonsensical tags). A chi-square test showed that the amount of nonsensical tags differed significantly among conditions ($\chi^2 = 37.71$, $p < .001$). Descriptive analysis indicated that "cheating behavior" was less common in the level condition (4.9% nonsensical tags) than in the leaderboard (6.2%), points (7.4%) and control conditions (8.2%).

As illustrated in Figure 4 and supporting H1, participants in the game element conditions generated significantly more tags than participants in the control condition ($F(3, 291) = 11.109$, $p < .001$, $\eta_p^2 = .102$). Planned contrasts showed that participants in the points condition significantly outperformed participants in the control condition ($F(1, 153) = 10.523$, $p = .001$, $\eta_p^2 = .064$), and were in turn significantly outperformed by participants in the leaderboard ($F(1, 154) = 5.23$, $p = .024$ $\eta_p^2 = .033$) and level conditions ($F(1, 151) = 3.91$, $p = .050$ $\eta_p^2 = .026$). Performance did not differ between the leaderboard and levels conditions.

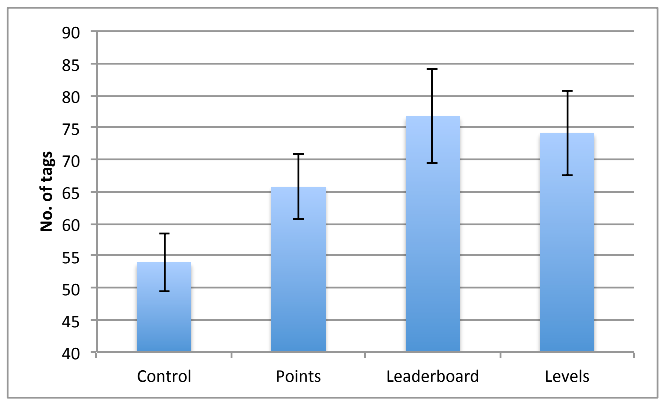

Figure 4. Average number of user-generated tags per condition. Error bars indicate 95% confidence intervals.

To check whether tagging performance changed over time for the different conditions, we calculated a 3x4 repeated measures ANOVA with time (3 blocks with 5 images each) as within-subject factor and condition as between-subject factor. A significant, significant time x condition interaction ($F(6, 582) = 2.462$, $p = .023$, $\eta_p^2 = .025$), as well as a significant main effect of time on performance ($F(6, 582) = 17.447$, $p < .001$, $\eta_p^2 = .057$) was found. Although performance over all experimental conditions decreased over time, participants' performance in the leaderboard and levels conditions declined more slowly than in the other two conditions (see Figure 5).

Conditions did not differ in the overall time participants spent on the image annotation task ($F(3, 291) = 2.015$, $p < .112$, $\eta_p^2 = .02$). But game elements significantly impacted time spent per tag ($F(3, 291) = 2.956$, $p = .033$, $\eta_p^2 = .03$).). Participants in the game element conditions spent less time per tag than

	Control ($N = 71$)		*Points* ($N = 83$)		*Leaderboard* ($N = 73$)		*Levels* ($N = 68$)	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Tags	54.24	19.120	66.01	23.527	76.86	31.234	74.31	27.175
Competence	3.81	1.161	3.85	1.124	4.13	1.179	3.91	1.017
Autonomy	5.42	1.004	5.20	1.147	5.28	1.156	5.39	.818
IM[a]	4.87	1.454	4.54	1.554	4.79	1.609	4.90	1.445

Table 1. Means and standard deviations of dependent variables for all conditions. [a] Intrinsic motivation.

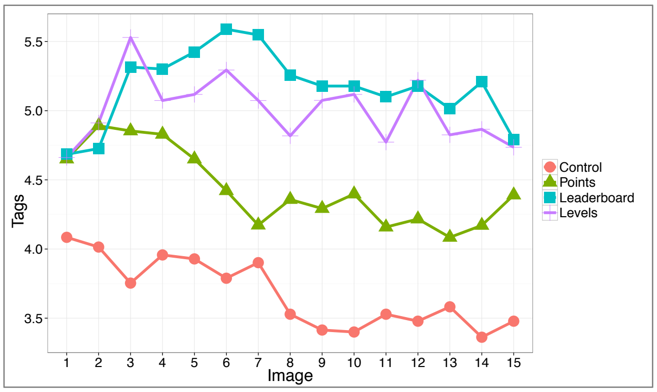

Figure 5. Average number of user-generated tags per condition over the course of 15 images.

participants in the control condition. In other words, participants in the points, leaderboard, and levels conditions generated more tags in the same amount of time as participants in the control condition.

The distribution of participants' number of tags showed some interesting differences (see Figure 6). In the control and points conditions, two (2.8%) resp. five (6.9%) participants came up with more than 100 tags, yet in the leaderboard and level conditions 17 (23.3 %) resp. 14 (20.6%) participants generated more than 100 tags each.

Intrinsic motivation & need satisfaction

Against our expectations, no significant effect of game elements on intrinsic motivation was found (p = .499). Participants reported similar levels of task enjoyment and interest, regardless of whether they received feedback in form of points, leaderboard, levels, or none at all (see Table 1). Also, no significant effects on either autonomy (p = .570) or competence (p = .340) need satisfaction were found. H2 could thus not be confirmed. Overall, participants were rather engaged in the image annotation task (see Table 1) and several commented that they enjoyed coming up with suitable tags for the paintings.

DISCUSSION

Our motivation for the present study was to experimentally assess whether and how points, leaderboards and levels affect participants' performance and intrinsic motivation in an image annotation task. In line with existing research on the potential of game elements to promote user behavior [5, 9, 10, 23], points, and especially levels and the leaderboard prompted participants to generate significantly more tags in less amount of time. Moreover, implementation of the leaderboard and leveling system inspired participants to maintain their performance for longer, compared to the points and control conditions. Arguably, the addition of game elements enhanced the quantity of tags, which confirms von Ahn and Dabbish's recommendations for increasing the throughput [24]. Against expectations, none of the game elements affected intrinsic motivation or need satisfaction in any way. All participants were motivated to similar degrees and reported intrinsic motivation did not mirror their performance in the image annotation task.

The observed performance gains may be due to the employed game elements being all forms of goal metrics [26]. By communicating how many tags have been generated, points likely formed a clear connection between participants' effort and their performance in the image annotation task [12, 24]. The levels and leaderboard perhaps further reinforced this behavior by setting explicit goals for participants to aspire to [9, 10, 12, 24]. In fact, previous research on goal setting in information systems found that users' performance increased when given a clear goal, as opposed to users who were simply asked "to do their best", even if the latter were aware of their performance [12]. Thus, it seems plausible that in the present study, participants in the points condition "did their best", but had no point of reference to judge their performance. In contrast, participants in the level and leaderboard conditions were always shown how many points they needed to reach the next level/next rank on the leaderboard, which may have prompted them to generate more tags than the other conditions. Still, even though the level and leaderboard conditions generated the most tags, their throughput eventually declined with time, which suggests that these elements did not really add to the task's "interestingness".

Yet against our expectations, participants' reported intrinsic motivation did not reflect their performance, nor was it negatively affected by either points, levels or the leaderboard. Apparently, participants did not perceive the game elements as particularly controlling, compared to the control condition. Perhaps, points, levels and leaderboards by themselves did not affect intrinsic motivation negatively, because they were not linked to other potentially pressuring external events, such as cash prizes for the best performance [4]. In contrast to the results on game elements in enterprise social network sites [9, 23], participants in the present study found themselves in a

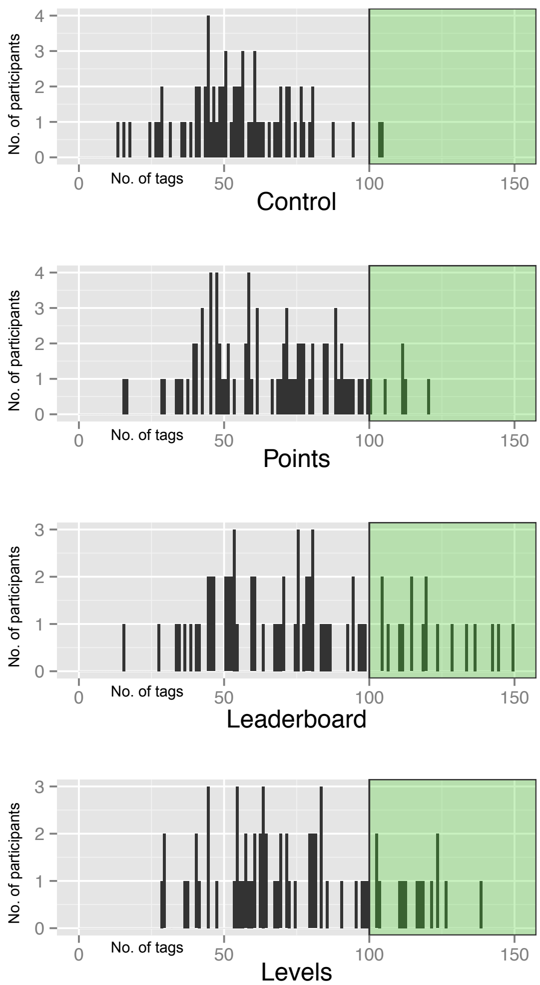

Figure 6. Distribution of the number of tags per participant for each condition.

relatively laid-back setting, as they chose to participate in the study voluntarily and, in case of the leaderboard condition, competing against people whom they did not know. Thus, we assume that due to the relative absence of controlling factors, participants' need for autonomy was not threatened.

On the other hand, game elements also did not increase feelings of competence. Hence, it may be assumed that points, levels and the leaderboard were not considered informational feedback [4]. This might be due to the goal metrics not offering enough meaningful, informational feedback to help participants judge their performance [19], such as whether a tag was fitting an image or not. On the other hand, this may not be as much due to the game elements themselves, but rather due to the nature of the task. While the image annotation task was deemed pleasant, it could hardly be considered a challenge, as participants were free to create any and as many tags as they wanted. For instance, it has been argued that the motivational appeal of games lies in their ability to provide players with new challenges to master [16], hence allowing them to experience feelings of competence [17, 19]. Seeing how many popular games rely on goal metrics [26] (e.g., *Tetris*), these game elements may perhaps only facilitate intrinsic motivation for tasks that demand learning and skill mastery.

Similarly, achievement goal theory differentiates between two types of goals, namely, mastery and performance goals [18]. While mastery goals refer to skill development and task mastery, performance goals focus on the demonstration of competence relative to normative standards. A meta-analysis on the effects of performance and mastery goals on intrinsic motivation found that informational feedback only increased intrinsic motivation for mastery goals, whereas performance goals were left unaffected [18]. Indeed, the image annotation task used in the present study bears more resemblance to a performance than a mastery goal, as participants simply had to "demonstrate" their competence in tagging paintings, relative to the norm set by levels or the leaderboard.

While the findings of the present study raise many questions on intrinsic motivation and gamification, they yield a few practical implications for the design of gamified applications. Designers may consider implementing points, levels and leaderboards in their projects as a quick and easy way to boost user performance for simple tasks, as these game elements may set clear goals for users to strive towards. While intrinsic motivation should remain unaffected by the addition of goal metrics, situational factors should still be kept in mind, as they may determine whether game elements are perceived as controlling and hence, damage intrinsic motivation. However, if the aim of gamification is to facilitate intrinsic motivation, then the mere addition of points, levels and leaderboards is not sufficient to make non-game contexts more engaging.

Limitations and further research

We provide evidence that points, levels and leaderboards are an effective means to increase short-term performance in an image annotation task. Against expectations, goal metrics did not significantly affect participants' intrinsic motivation. However, the present study featured several shortcomings that have to be addressed.

First, we only examined the short-term effects of points, levels and leaderboards. While research in psychology has found that rewards affect intrinsic motivation even for simple, short tasks [4], previous findings on the effects of game elements on user engagement have shown that it is important to also study the long-term effects of game elements, in order to better assess whether and under what circumstances they shape user behavior in the long run (e.g., [5, 10]). Also, while previous research [10, 12] and the findings of the present study suggest that gamification improves performance by means of goal-setting, it still has to be seen whether the implementation of goals without game elements (e.g., by asking participants to generate a set number of tags) affects user motivation and performance differently than when goal metrics, such as points, levels or leaderboards are employed.

Secondly, we only measured participants' self-reported intrinsic motivation. While self-reported and "free choice" measures of intrinsic motivation yielded comparable results in previous studies [4], it would have been interesting to employ a behavioral "free choice" measure of intrinsic motivation by letting participants choose whether they want to continue engaging with the image annotation task, even after the conclusion of the experiment. Future studies should consider combining self-reported and behavioral measures of intrinsic motivation for additional methodological robustness. Also, because participants from the university's database usually engage voluntarily in studies, it is possible that they already had a minimum level of intrinsic motivation from the get-go, which might have affected the results of the present study. More research is required to investigate how users' initial eagerness to engage in a gamified application affects their subsequent motivation.

Thirdly, while von Ahn and Dabbish consider throughput and overall time spent on the main indicators of success of gamified human computation tasks [24], it would be interesting to examine more closely how goal metrics affect the *quality* of tags (i.e., how well tags fit the image). Future studies should look into how game elements may not only be implemented to increase the quantity of a certain behavior, but also how to enhance the quality of a certain behavior. This would not only yield better tags, but turn the image annotation task into a mastery goal, thereby making it potentially more challenging and interesting for users.

Finally, the results of the present study are specific to the image annotation context and the task itself scored only somewhat above average on intrinsic motivation. As rewards only threaten intrinsic motivation for activities that people find *interesting* [4], goal metrics may not have affected intrinsic motivation, due to the task not being interesting enough. Hence, our findings should only cautiously be applied to other gamified applications. It still has to be seen whether these results can be replicated for other gamified tasks. However, we believe that only by studying the effects of game elements in different non-game contexts, can we gain a more comprehensive understanding of how and when points, levels and leaderboards should be implemented to promote user behavior and intrinsic motivation.

More research is required to further investigate the role of contextual, social and situational aspects, as they at least partially determine how game elements affect intrinsic motivation and behavior [6]. While it has already been shown that goal metrics may harm intrinsic motivation in situations that may inherently be perceived as controlling [9, 23], it still has to be seen whether they also drive user behavior for other, potentially more voluntary contexts, other than human computation. For instance, it would be interesting to study the effects of goal metrics for applications focusing on sustainable living, as people usually freely choose to pursue such goals.

Also, Hamari found that badges had only a limited effect on user engagement in a peer-to-peer sharing service [10]. Thus, he argued that game elements might affect user engagement differently, depending on whether they are implemented in utilitarian or hedonic services, and stresses that use scenarios should always be kept in mind when designing gamified applications. Similarly, previous research in psychology suggests that rewards may either undermine or enhance intrinsic motivation depending on whether they are endogenous or exogenous to a given task [13]. Lastly, as previous research suggests that games are motivating due to them providing players with the possibility of expressing their choices and skills [17, 19], future studies should also compare whether "mastery" and "performance" tasks [18] are impacted differently by game elements, in order to examine whether certain non-game contexts may benefit more from goal metrics than others.

CONCLUSION

Points, levels and leaderboards are not only some of the most basic, but also three of the most commonly employed game elements in game and non-game contexts. While it has been argued that they may negatively impact users' intrinsic motivation, no actual empirical evidence exists to back this claim. The findings of the present study suggest that gamification by means of implementing points, levels and leaderboards may be an easy, viable and effective way to drive user behavior – at least in the short term. Perhaps by establishing a clear connection between user effort and performance, and by providing explicit performance goals, these game elements significantly enhanced participants' performance in an image annotation task. While significant performance gains were achieved, intrinsic motivation remained unaffected by the mere presence of points, levels and leaderboards. However, designers of gamified services should still be wary of potential social or contextual factors that may determine whether these game elements diminish intrinsic motivation. Also, as these game elements did not increase intrinsic motivation, they should not solely be relied upon to sustain long-term user engagement.

ACKNOWLEDGEMENTS

We thank Roland Hübscher and the reviewers for their very helpful suggestions on improving this paper. Alexandre N. Tuch was supported by the Swiss National Science Foundation under fellowship number PBBSP1 144196.

REFERENCES

1. Essential facts about the computer and video game industry. Tech. rep., Entertainment Software Association, 2013. Retrieved June 1, 2013 from `http://www.theesa.com/facts/pdfs/ESA_EF_2013.pdf`.
2. Antin, J., and Shaw, A. Social desirability bias and self-reports of motivation: a study of amazon mechanical turk in the us and india. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, ACM (2012), 2925–2934.
3. Deci, E., Eghrari, H., Patrick, B., and Leone, D. Facilitating internalization: The self-determination theory perspective. *Journal of personality 62*, 1 (1994), 119–142.
4. Deci, E., Koestner, R., and Ryan, R. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin 125*, 6 (1999), 627–668.

5. Denny, P. The effect of virtual achievements on student engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 763–772.
6. Deterding, S. Situated motivational affordances of game elements: A conceptual model. In *Gamification: Using Game Design Elements in Non-Gaming Contexts, a workshop at CHI* (2011).
7. Deterding, S. Coding conduct: Games, play, and human conduct between technical artifacts and social framing, 2012. Retrieved June 1, 2013 from `http://www.slideshare.net/dings/coding-conduct-games-play-and-human-conduct-between-technical-code-and-social-framing`.
8. Deterding, S., Dixon, D., Khaled, R., and Nacke, L. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, ACM (2011), 9–15.
9. Farzan, R., DiMicco, J. M., Millen, D. R., Dugan, C., Geyer, W., and Brownholtz, E. A. Results from deploying a participation incentive mechanism within the enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2008), 563–572.
10. Hamari, J. Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications* (2013). `http://dx.doi.org/10.1016/j.elerap.2013.01.004`.
11. Hecker, C. Achievements considered harmful, 2010. Retrieved June 1, 2013 from `http://chrishecker.com/Achievements_Considered_Harmful`.
12. Jung, J., Schneider, C., and Valacich, J. Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science 56*, 4 (2010), 724–742.
13. Kruglanski, A. W. The endogenous-exogenous partition in attribution theory. *Psychological Review 82*, 6 (1975), 387.
14. Kumar, J. M., and Herger, M. *Gamification at Work: Designing Engaging Business Software*. The Interaction Design Foundation, Aarhus, Denmark, 2013.
15. Machajdik, J., and Hanbury, A. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, ACM (2010), 83–92.
16. Malone, T. Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 conference on Human factors in computing systems*, ACM (1982), 63–68.
17. Przybylski, A., Rigby, C., and Ryan, R. A motivational model of video game engagement. *Review of General Psychology 14*, 2 (2010), 154–166.
18. Rawsthorne, L. J., and Elliot, A. J. Achievement goals and intrinsic motivation: A meta-analytic review. *Personality and Social Psychology Review 3*, 4 (1999), 326–344.
19. Rigby, S., and Ryan, R. *Glued to games: How video games draw us in and hold us spellbound*. ABC-CLIO, 2011.
20. Robertson, M. Can't play, won't play, 2010. Retrieved June 1, 2013 from `http://hideandseek.net/2010/10/06/cant-play-wont-play/`.
21. Ryan, R., and Deci, E. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology 25*, 1 (2000), 54–67.
22. Ryan, R., Mims, V., and Koestner, R. Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology 45*, 4 (1983), 736–750.
23. Thom, J., Millen, D., and DiMicco, J. Removing gamification from an enterprise sns. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM (2012), 1067–1070.
24. Von Ahn, L., and Dabbish, L. Designing games with a purpose. *Communications of the ACM 51*, 8 (2008), 58–67.
25. Wang, J., and Yu, B. Labeling images with queries: A recall-based image retrieval game approach. In *Proceedings of the ACM SIGIR Workshop on Crowdsourcing for Information Retrieval* (2011).
26. Zagal, J. P., Mateas, M., Fernández-Vara, C., Hochhalter, B., and Lichti, N. Towards an ontological language for game analysis. In *Proceedings of International DiGRA Conference: Changing Views – Worlds in Play.* (2005), 3–14.
27. Zichermann, G., and Cunningham, C. *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media, 2011.