Disassembling Gamification: The Effects of Points and Meaning on User Motivation and Performance

Elisa D. Mekler
Department of Cognitive Psychology and Methodology
University of Basel
4055 Basel, Switzerland
elisa.mekler@unibas.ch

Alexandre N. Tuch
Department of Computer Science
University of Copenhagen
2300 Copenhagen, Denmark
a.tuch@unibas.ch

Florian Brühlmann
Department of Cognitive Psychology and Methodology
University of Basel
4055 Basel, Switzerland
florian.bruehlmann@stud.unibas.ch

Klaus Opwis
Department of Cognitive Psychology and Methodology
University of Basel
4055 Basel, Switzerland
klaus.opwis@unibas.ch

CHI 2013 Extended Abstracts, April 27 – May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

Abstract
Interest in gamification is growing steadily. But as the underlying mechanisms of gamification are not well understood yet, a closer examination of a gamified activity's meaning and individual game design elements may provide more insights. We examine the effects of points – a basic element of gamification, – and meaningful framing – acknowledging participants' contribution to a scientific cause, – on intrinsic motivation and performance in an online image annotation task. Based on these findings, we discuss implications and opportunities for future research on gamification.

Author Keywords
Gamification, gameful design, games with a purpose, meaning, motivation

ACM Classification Keywords
H.5.m [Information Interfaces and Presentation (e.g., HCI)]: Miscellaneous; K.8.0 [Personal Computing]: Games; J.4 [Social and Behavioral Sciences]: Psychology.

General Terms
Human Factors; Design; Experimentation

Introduction

More and more web-based and mobile applications look to gamification, the use of game design elements (e.g., points, leaderboards and badges) in non-game contexts [7], to promote user engagement. Interest is increasing steadily, as reflected in the rapidly growing body of literature on the subject. Searching for gamification on Google Scholar nets nearly 1700 publications, with 1180 published only in the year 2012 (as of December 19th). While many discuss the potential and pitfalls of gamifying products and services, there is to date little empirical research on how gamification works and whether it succeeds in promoting user motivation.

Gamification relies on elements characteristic of games to shape user behavior and motivation. Although the motivational appeal of full-fledged games has already been discussed [16], it is hard to pinpoint which individual game design patterns actually affect player motivation, thereby offering only limited carry-over to gamification [6]. To properly understand how gamification functions, one promising approach would be to examine the effects of individual game design elements [6]. While there are efforts to link motivation to individual game design patterns [9, 12], to date only one study attempted to experimentally assess the motivational affordance of game-like elements [10].

Meaning is another crucial aspect of engaging gamification [5, 14]. An activity may be framed as meaningful, when embedded within a narrative, supporting users' personal goals and interests, or having a purpose that is deemed valuable by users (e.g., playing "Phylo" [11] to advance scientific progress) [5]. McGonigal states that meaning forms a type of intrinsic reward in itself, because "[...] we want to belong to and contribute to something that has lasting significance beyond our own individual lives" ([14], pp. 50). Surprisingly, even though meaning is considered such an integral part of well-thought-out gamification, there is a lack of research examining the interplay of game design elements and meaning.

We aim to counter this lack of empirical evidence and plan a series of experiments that cover different forms of meaning and game design elements, and their effects on user motivation and behavior. In a first step, the present study examines the effects of meaningful re-framing of a task and points, as many other game design elements typical of gamification (e.g., levels, high-score lists) build upon points. We chose to employ an image annotation task, as 1) we assume it is sufficiently tedious a task to benefit from gamification and/or meaningful framing, as suggested by previous studies [2, 8, 17, 18], and because 2) user performance can easily be measured through the quantity and quality of generated tags.

Related work

The term *Game with a purpose* (GWAP) describes the use of game elements, such as score keeping and high-score lists, to make human computation tasks, such as image annotation more enjoyable [18]. And indeed, Goh et al. [8] found that users preferred "gamified" versions of an image tagging task over a non-game version, even though the latter yielded more quality tags. These studies however solely focused on combinations of game design elements, which makes it difficult to assess how individual game elements are linked to behavioral and motivational outcomes.

Another approach to promote user participation in human computation is to imbue tasks with meaning. Chandler

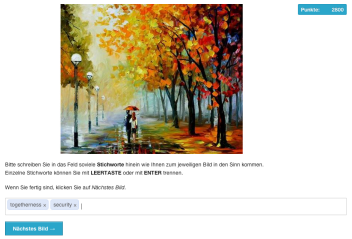

Figure 1: A screenshot of the image annotation task with the score displayed in the upper right corner.

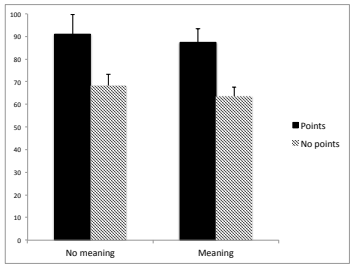

Figure 2: Average amount of tags per participant.

and Kapelner [2] provided a meaningful frame by explaining participants that they would identify tumor cells, which resulted in more images tagged and an increased likelihood to participate in the task, compared to participants who were asked to identify mere "objects of interest". In another study, participants were told that their effort would help a non-profit organization (i.e., the meaningful frame), which resulted in higher quality work, albeit not in higher quantity [17].

While the aforementioned studies show that providing meaning or adding game elements to human computation tasks may promote performance and intrinsic motivation, to our knowledge none have examined the effects of combining the two approaches.

Experiment

In order to investigate the effects of points and meaning on user performance and motivation, we implemented an online experiment with a 2 x 2 between-subject design. The independent variables were points (points vs. no points) and meaningful framing (framing vs. no framing).

A total of 172 participants (46 male, 123 female, 3 not specified; mean age 32.95 years, range 15 - 74) were randomly assigned to one of the four experimental conditions.

After filling in demographic information, participants were introduced to the image annotation task, namely providing tags that describe the mood of 15 abstract paintings. We used the images from Machajdik and Hanbury's study on affective image classification [13] and adapted the task so that participants could freely associate tags to the images, instead of having to choose from a set number of tags. A test trial, which was the same for every condition with no points displayed and no additional explanation provided, preceded the actual experiment. The images were presented in random order. After completing the image annotation task, participants filled out the Intrinsic Motivation Inventory (IMI) [3] (7-point Likert scale, 1 = not at all true, 7 = very true) and had the option to comment on the study.

To establish a meaningful frame, *recognition* and *purpose* were provided, as suggested by Ariely et al. [1]. After the test trial, participants in the meaningful framing conditions were informed that their tags would help improve computerized affective image categorization and that their contribution would thereby advance science. In the points conditions, participants earned 100 points for each tag they entered. The current score was displayed in the upper right corner of the screen (see Figure 1). Points had no further meaning, other than depicting how many tags a participant had entered. After finishing all 15 images, participants were presented with their final score.

Results

In order to examine the effects of points and meaningful framing on performance, tag quality and intrinsic motivation ratings, analyses of variance (ANOVA) were calculated, unless otherwise stated.

Data preparation

To determine tag quality, tags were matched with a German dictionary with over 1.3 million entries (http://germandict.sourceforge.net/) and subsequently all articles (e.g., *the*) removed. Then, we checked whether the framing manipulation was successful and found that participants perceived the task as significantly more valuable ($p = .024$) and tended to perceive it as personally more important ($p = .088$), when a meaningful frame was given, regardless of points.

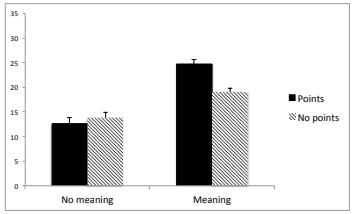

Figure 3: Average time spent per tag in seconds.

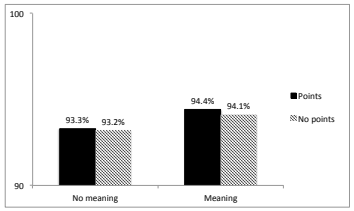

Figure 4: Quality tags in %.

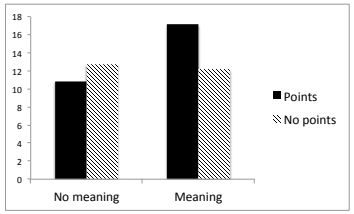

Figure 5: Average amount of affective tags per participant.

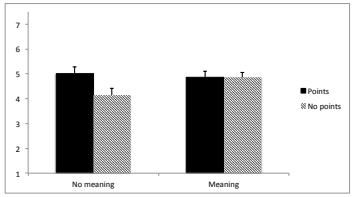

Figure 6: Mean intrinsic motivation.

Tagging performance
Participants in the points conditions generated significantly more tags (F (1, 168) = 13.613, $p < .001$), while framing did not affect the total amount of tags ($p = .995$; Figure 2). No points x framing interaction occured.

For time spent per tag (a proxy for participants' effort to generate task-relevant labels), there were no significant main effects. However, a significant points x framing interaction (F (1, 168) = 5.262, $p = .023$) was found, indicating that participants spent more time per tag, when meaningful framing was combined with points (Figure 3).

Lastly, we checked whether points and framing had an effect on tag quality (i.e., whether the generated tags were sensible or nonsensical). A chi-square test showed that points did not affect the quality of the tags (i.e., number of sensible tags), whereas a significant effect was found for framing ($\chi^2 = 6.46$, $p = .011$), indicating that meaningful framing elicited around 100 more sensible tags than the no meaning condition (Figure 4). Because the annotation task asked for tags describing the mood/emotional content of the images, we ran a Linguistic Inquiry and Word Count (LIWC) Analysis [15] as an additional measure of quality, to check for the amount of affective tags per condition. Again, a significant effect for meaningful framing was found ($\chi^2 = 93.39$, $p < .001$). Participants in the points/meaningful frame condition generated on average more affective tags (Figure 5).

Intrinsic motivation ratings
As illustrated in Figure 6, there was a significant points x framing interaction (F (1, 172) = 4.405, $p = .037$), as well as a main effect for points (F (1, 172) = 5.755, $p = .018$). This means that both framing and points on their own, as well as the combination of both factors, increased participants intrinsic motivation to a similar degree compared to the control condition (i.e., no meaningful frame and no points). In other words, it did not matter whether participants got points, a meaningful frame or both. Only the absence of both factors led to a lower level of intrinsic motivation.

Conclusions and Future Work

Our motivation for the present study was to experimentally assess how providing points and a meaningful frame for an image annotation task affects participants' performance and intrinsic motivation. While points did motivate participants to generate more tags, a meaningful frame inspired them to do better at the task and create more quality tags. Overall, the combination of points and meaningful framing yielded the best results. Interestingly, both points and meaning on their own and the combination thereof increased intrinsic motivation in equal measure.

The most likely reason why points elicited more tags from participants may be due to them functioning as feedback. Points establish a clear connection between user effort (typing tags) and performance (amount of tags generated) [10, 14, 18], which then might have motivated participants to generate more tags. Moreover, people have an inherent need to accomplish tasks and providing them with individual performance information (i.e., points as feedback) may facilitate the fulfillment of this need [10]. This may hold especially true for people with a high need for achievement [8].

Providing the task with a meaningful frame seemed to inspire participants to work harder and take their time when labeling the images. This in turn, may have proven beneficial for the quality of the tags, which conforms to previous findings [17]. Moreover, the combination of

points and a meaningful frame led to the highest increase in quality, even though points were only awarded for the number of tags generated. Future studies should score participants on tag quantity and quality – for instance, by having participants simultaneously agree on tags [18], – to allow for a more detailed examination of this interaction. Caution is advised though, as research on motivation suggests that rewarding points may undermine participants' interest in the task (i.e., its meaningfulness) and hence impair intrinsic motivation and performance [4].

Surprisingly, participants' reported intrinsic motivation did not mirror their performance in the image annotation task. All participants were motivated to similar degrees, apart from those in the control condition, which featured neither a meaningful frame nor points. For a better understanding of this phenomenon, our next study will take into account aspects that may determine intrinsic motivation, such as participants' motivational need for achievement and competence. Participants in the points condition may have felt more competent, as suggested by Deterding [6] and Jung et al. [10], which then furthered their intrinsic motivation. In the meaningful framing condition, participants were perhaps content to contribute to the scientific cause [5, 14].

In order to investigate how enduring these findings are, the present study has to be expanded to include more images and/or sets of images and allow participants to choose themselves how many images they wish to tag. This way, we may observe whether and for how long points and meaningful framing increase participants' willingness to continue tagging images, along the lines of the study designs employed by [1, 2, 17]. Moreover, we plan to introduce a new experimental condition, where points and meaning are more tightly interwoven. The GWAP "Phylo" [11] provides a good example of this, as there is a clear connection between the scoring system and the purpose of the activity (i.e., aligning multiple DNA sequences).

Eventually, we will gradually introduce other game design elements typical of gamification, such as high-score lists, as suggested by previous studies [10, 18], thereby adding a social dimension. This will in turn raise new issues, as competition may either promote intrinsic motivation by offering participants an opportunity to experience competence [10], or reduce it, if perceived as involuntary and controlling [6].

Finally, it has yet to be seen whether the findings of the present study can be applied to the gameful redesigning of non-game contexts, other than human computation. In some contexts, the combination of game design elements and meaning (e.g., earning points for making someone a compliment) may pose problems due to social circumstances [5]. Sustainable living, on the other hand, seems to be a likely candidate, as many people perceive it as a meaningful and worthwhile cause, even if the personal gains of such activities may sometimes seem unclear. In sum, by disassembling gamification into its components and implementing it in different non-game contexts, we may gain a deeper understanding of how gamification works and how it may be applied to solve real-world problems.

Acknowledgements

We thank Roland Hübscher and the reviewers for their very helpful comments in preparing this paper.

References

[1] Ariely, D., Kamenica, E., and Prelec, D. Man's search for meaning: The case of legos. *Journal of*

Economic Behavior & Organization 67, 3 (2008), 671–677.

[2] Chandler, D., and Kapelner, A. Breaking monotony with meaning: Motivation in crowdsourcing markets. *University of Chicago mimeo* (2010).

[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] Deterding, S. Meaningful play: getting "gamification" right. *Google Tech Talk* (2011).

[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., 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.

[8] Goh, D., and Lee, C. Perceptions, quality and motivational needs in image tagging human computation games. *Journal of Information Science 37*, 5 (2011), 515–531.

[9] Holopainen, J., and Björk, S. Gameplay design patterns for motivation. *ISAGA 2008* (2008).

[10] 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.

[11] Kawrykow, A., Roumanis, G., Kam, A., Kwak, D., Leung, C., Wu, C., Zarour, E., Sarmenta, L., Blanchette, M., and Waldispühl, J. Phylo: a citizen science approach for improving multiple sequence alignment. *PloS one 7*, 3 (2012), e31362.

[12] Lewis, C., Wardrip-Fruin, N., and Whitehead, J. Motivational game design patterns of'ville games. In *Proceedings of the International Conference on the Foundations of Digital Games*, ACM (2012), 172–179.

[13] 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.

[14] McGonigal, J. *Reality is broken: Why games make us better and how they can change the world*. Penguin Press HC, 2011.

[15] Pennebaker, J., Chung, C., Ireland, M., Gonzales, A., and Booth, R. The development and psychometric properties of liwc2007. *Austin, TX, LIWC. Net* (2007).

[16] Przybylski, A., Rigby, C., and Ryan, R. A motivational model of video game engagement. *Review of General Psychology 14*, 2 (2010), 154.

[17] Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., and Vukovic, M. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *Proc. of AAAI Conference on Weblogs and Social Media*, vol. 11 (2011).

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