Do badges increase user activity? A field experiment on the effects of gamification

Juho Hamari[*]

Game Research Lab, School of Information Sciences, FIN-33014 University of Tampere, Finland
Department of Information and Service Economy, Aalto University School of Business, P.O. Box 21220, 00076 Aalto, Finland

ABSTRACT

During recent years, the practice of adding game design to non-game services has gained a relatively large amount of attention. Popular discussion connects gamification to increased user engagement, service profitability, goal commitment and the overall betterment of various behavioral outcomes. However, there is still an absence of a coherent and ample body of empirical evidence that would confirm such expectations. To this end, this paper reports the results of a 2 year (1 + 1 year − between-group) field experiment in gamifying a service by implementing a game mechanic called 'badges'. During the experiment a pre-implementation group ($N = 1410$) was monitored for 1 year. After the implementation, the post-implementation (the gamified condition) group ($N = 1579$) was monitored for another full year. Results show that users in the gamified condition were significantly more likely to post trade proposals, carry out transactions, comment on proposals and generally use the service in a more active way.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

During recent years, the boundary between games and other systems and services has become increasingly blurred. This development can be seen to be bi-directional: On one hand, within games, users are increasingly subjected to decision making situations pertaining to outside-game concerns (especially with the rise of Free-to-play games about how people use money − Alha, Koskinen, Paavilainen, Hamari, & Kinnunen, 2014; Hamari, 2015; Hamari & Järvinen, 2011; Hamari & Lehdonvirta, 2010; Paavilainen, Hamari, Stenros, & Kinnunen, 2013). On the other hand, in non-game contexts, game design is increasingly being used to direct people's motivations towards intrinsically motivated, gameful experiences and behavior (Deterding, Dixon, Khaled, & Nacke, 2011; Hamari, Huotari, & Tolvanen, 2015; Huotari K. & J., 2012; McGonigal, 2011). This phenomenon is commonly referred to as gamification (Deterding et al., 2011; Hamari et al., 2015; Huotari & Hamari, 2012). Gamification has already been applied in several areas, including the promotion of greener energy consumption (Nissan Leaf), building loyalty towards TV channels

(GetGlue), taking care of one's health (Fitocracy), and even for gamifying the tracking of one's aspirations in life (Mindbloom). Predictions about the diffusion of gamification have varied from extremely positive outlooks (e.g. Gartner, 2011; IEEE, 2014 − Most organizations will adopt gamification strategies in the near future), to less optimistic ones (Gartner, 2012 − most adoptions will fail).

Popular positive belief in the effectiveness of gamification has often been based on the anecdotal conception that because most games are 'fun' and intrinsically motivating, then any service that uses the same mechanics should also prove to be 'fun' and effective in invoking positive further behavioral outcomes. It is clear that gamification has attracted significant interest and opinion, although its conceptions remain scant and there is a relative dearth of a coherent body of empirical evidence on its effectiveness. Moreover, meta-studies have detected that the field is strongly dispersed and often afflicted with sub-par study designs with regards to controls, sample sizes and experiment durations (see Hamari, Koivisto, & Pakkanen, 2014a; Hamari, Koivisto, & Sarsa, 2014b). Therefore, it is not surprising that the discussion around gamification is still relatively divergent.

In this paper, we studied the effects of gamification (a badge system) on user activity in a sharing economy service (a peer-to-peer marketplace). In our experiment, people could unlock badges by completing common actions and tasks within the

* Address: Game Research Lab, School of Information Sciences, FIN-33014 University of Tampere, Finland. Tel.: +358 50 318 6861.
E-mail address: juho.hamari@uta.fi.

service. The experiment focused on investigating whether the implementation of badges positively affects usage activity. Since the experiment was carried out in a peer-to-peer marketplace, usage activity was measured via four related dependent variables: the amount of posted trade proposals, accepted transactions, posted comments, and general usage activity as measured via page views. The field experiment spanned 2 years $(1 + 1$ year $-$ between-group). During the experiment a pre-implementation group $(N = 1410)$ was firstly monitored for 1 year. After the implementation, a post-implementation group $(N = 1579)$ was monitored for another full year.

2. Background

2.1. Related literature

Industry studies have found that the addition of badges to games has led to better critical reception and increased revenue (Electronic Entertainment Design, 2007). In fact, large game console publishers such as Microsoft, demand that game developers include badges in games that are published for Xbox consoles (see Jakobsson, 2011). However, there is a dearth of literature as to how badges affect user behavior in a gamification setting where users are not predisposed to gaming.

Badges consist of optional rewards and goals, the fulfillment of which is located outside the scope of the core activities of a service. On a systemic level, a badge consists of a signifying element (the visual and textual cues of the badge), rewards (the earned badge), and the fulfillment conditions which determine how the badge can be earned (Hamari, 2013; Hamari & Eranti, 2011; Jakobsson, 2011; Montola, Nummenmaa, Lucerano, Boberg, & Korhonen, 2009). Furthermore, because of their visual element (the badge itself) and the included descriptions regarding the goal and how to unlock a badge, they may also be accompanied by narrative elements and challenges that have been found to give rise to intrinsic motivations (Malone, 1981).

Badges have been one of most common mechanics investigated in gamification studies and studied in a variety of contexts (Hamari et al., 2014b) (Table 1). In an educational context, Domínguez et al., 2013 found that while badges did have a positive effect on practical assignments, they had a possible negative effect on written assignments. Hakulinen, Auvinen, and Korhonen (2013) found that results depend upon the badge type, as well as the users. Denny (2013), on the other hand, found only positive effects regarding the level of contributions, as well as on the time a student engaged with the system.

In a commerce context, Hamari (2013) found that enabling people to compare their badges and to use them as service user goals, had little significant effect on either the amount or quality of service use. However, those people who actively followed up on the accumulation of badges showed an increased service use.

Two studies (Fitz-Walter, Tjondronegoro, & Wyeth, 2011; Montola et al., 2009) share the observation that badges can have both positive and negative consequences. Undesirable usage patterns were deemed to be a potential problems as badges might entice users to excessively carry out those activities that award badges. The impact of badges on usability and their integration into the existing system were also considered as possible problems.

2.2. Theoretical underpinnings

According to Bandura (1993), set *goals* (such as those in badges) increase performance in three ways: (1) people anchor their expectations higher, which in turn increases their performance; (2) assigned goals enhance self-efficacy; (3) the completion of goals

leads to increased satisfaction, which in turn leads to increased future performance within the same activities. These effects are further strengthened if the goals are context-related, immediate, and the users are provided with (immediate) feedback. It has also been found that when goals are clearly specified in terms of how many times they have to be completed, the rate of completion of the tasks increases (Ling et al., 2005).

Another effect noted from using badges has been connected to their ability to guide user behavior because they set *clear goals*. It has been argued that badges function as a guidance mechanic (Hamari & Eranti, 2011; Jakobsson, 2011; Montola et al., 2009) in a service, providing the user with an idea of how the service is meant to be used and what is expected of the user, thus increasing the *amount* and *quality* of those actions within a service. In a larger context, goals are regarded as a central game mechanic (Salen & Zimmerman, 2004), and have been demonstrated to exert persuasive power even when the progression towards them was illusionary (Kivetz, Urminsky, & Zheng, 2006; Nunes & Drèze, 2006). Clear goals are also one of the main dimensions of flow theory (Csíkszentmihályi, 1990) which predicts that having clear goals and immediate feedback supports the emergence of a 'flow state', where the user's skills and the challenge of the task are optimally balanced.

Even though users may be offered clear goals as described above, they need to be committed to these goals in order for the hypothesized effects of increased motivation, engagement and performance to take place (Klein, Wesson, Hollenbeck, & Alge, 1999). According to Locke and Latham (1990), *goal commitment* can be defined as one's determination to reach a goal, implying that users are more likely to persist in pursuing goals and be less likely to neglect them.

Another rationale behind gamification has been to harness the persuasive power that emerges when people compare their points and badges amongst each other, effectively benchmarking themselves. In general, this phenomenon is called *social comparison* (Festinger, 1954), and this forms an over-arching concept for other more specific theories related to effects which result from comparisons between individuals such as *social influence* and the theory of planned behavior (Ajzen, 1991). The social influence and recognition that users receive through gamification have also been found to be strong predictors for the adoption and use of gamification applications (Hamari & Koivisto, 2013).

Social proof theory (Cialdini, 2001a, 2001b; Goldstein, Cialdini, & Griskevicius, 2008) predicts that individuals are more likely to engage in behaviors which they perceive others are also engaged in (Cialdini, 2001b). Gamification via badges facilitates social proof by providing a means for users to observe the activities of others, and indicating which behaviors they have been rewarded for − "*We view a behavior as correct in a given situation to the degree that we see others performing it*" (Cialdini, 2001b). The other side of this phenomenon is *social validation*, by which people signal their conformity, in that they have also engaged in same behaviors. Van de Ven, Zeelenberg, and Pieters (2011) found that people were willing to pay up to 64% more for a product that their peers had already acquired. Badges facilitate social validation by providing a means for users to display their conformity to the behavior and expectations of others.

3. Methods and data

According to a literature review on gamification, Hamari et al. (2014b), conclude that many empirical studies on gamification have suffered from methodological limitations. For instance, the studies have often had relatively small sample sizes, have been conducted in makeshift services, their experiments have lacked

Table 1
Studies on badges.

Reference	Outcome	Result	N	Context
Denny (2013)	Level and quality of participation	Positive effect on the quantity of students' contributions, without a corresponding reduction in their quality, as well as on the period of time over which students engaged with the tool	1031	Education
Domínguez et al. (2013)	Learning outcomes	Students who completed the gamified experience got better scores in practical assignments and in overall score, but our findings also suggest that these students performed poorly on written assignments and participated less on class activities, although their initial motivation was higher	195	Education
Fitz-Walter et al. (2011)	Exploration of the campus while interacting with the application	Suggests that added game elements can be enjoyable but can potentially encourage undesirable use by some, and aren't as enjoyable if not enforced properly by the technology. Consideration is also needed when enforcing stricter game rules as usability can be affected	26	A mobile information application for new university students
Hakulinen et al. (2013)	Impact on time management, carefulness and achieving learning goals	Achievement badges can be used to affect the behavior of students even when the badges have no impact on the grading. Statistically significant differences in students' behavior were observed with some badge types, while some badges did not seem to have such an effect. We also found that students in the two studied courses responded differently to the badges. Based on our findings, achievement badges seem like a promising method to motivate students and to encourage desired study practices	281	Education
Hamari (2013)	Amount of use, quality of use and social interaction	The results show that the mere implementation of gamification mechanisms [referring to the features that enable social comparison and goal attainment] does not automatically lead to significant increases in use activity, however, those users who actively monitored their own badges and those of others in the study showed increased user activity	3234	Commerce
Montola et al. (2009)	General impressions	The results suggest that there is some potential in achievement systems outside the game domain. The achievements triggered some friendly competition and comparison between users. However, many users were not convinced, expressing concerns about the achievements motivating undesirable usage patterns. Therefore, an achievement system poses certain design considerations when applied in non-game software	n/a — Qualitative study	Photosharing/social networking

control groups and different gamification elements have not been separately controlled for. The present experiment has a relatively large sample size with regards to both the number of subjects and the longevity of the measurement periods before and after the intervention. Moreover, the study was carried out in a pre-existing service which is still operational. The experiment also introduced a single gamification element (badges), rather than introducing a large set of different mechanisms. The data and methods of the study are now described in more detail.

3.1. The gamified service

Sharetribe (https://www.sharetribe.com/) is an international peer-to-peer trading service which offers its service package to a variety of organizations. Available localizations at the time of writing were English, Spanish, Finnish, Greek, French, Russian and Catalan. Sharetribe is used in communities all over the world. At the time of writing, there were 479 local Sharetribes world-wide. The company, Sharetribe Ltd., is a social for-profit enterprise registered in Finland. Their mission is to help people connect with their community and to help eliminate excess waste by making it easier for everyone to use assets more effectively by sharing them.

"Sharetribe is a network of "tribes", online communities where you can share goods, services, rides and spaces in a local, trusted environment. You can create a tribe for your university campus, your company, your neighborhood, your association, your sports club, your congregation, you name it!" — Sharetribe FAQ (2013)

Sharetribe's marketing strategy focuses on differentiating itself from other trading services such as eBay or Craigslist by targeting narrow local communities such as either an organization or town district, and by offering tools for non-monetary transactions including borrowing and carpooling. Users can however buy and sell goods and services. Sharetribe uses open source principles in the design of their service and the entire code is available for anyone to download. The reason for having many "tribes" is to emphasize local communities, trust and informational access, and also to diminish transaction costs and costs related to shipping (see Fig. 1).

3.2. Field experiment

The experiment setup followed users registered during one calendar year before the implementation of gamification, and users registered one calendar year after the implementation (Fig. 3 and Table 2).

The data consists of a database of user actions from a Sharetribe site of a major Finnish University. The data and measurements consist of the actions of users who were registered during the experiment timeframe ($n = 2989$), and includes the number of trade proposals, accepted transactions, comments posted, and the number of individual page views a user undertook. Selecting these for analysis allowed the experiment to be less affected by temporal usage patterns (see Tables 2 and 3 and Fig. 3). For instance, users commonly use the service less over time, and there is commonly a spike in usage right after registration. In this study, we were able to compare the behavior of two homogenous user groups pre- and post-implementation of gamification. We restricted the selection of users and dependent variable counts with mutually exclusive timeframes. This way we could prevent effects from older users, having for example an impact on existing trade proposals in the service that could have affected the dependent variable counts. We selected this specific Sharetribe site for the experiment as it is the largest Sharetribe of the several hundred installations in-place world-wide. The selection of only one 'tribe' also helped to

Fig. 1. Front page of Sharetribe.

guarantee the homogeneity between the populations of the control and post-implementation groups.

The experiment was purposefully conducted as a field experiment in a real existing service, rather than in a laboratory setting or a makeshift service in which respondents would have been asked to assume a hypothetical scenario of a badge system, and would be knowledgeable of the temporary nature of the service. In this way we could also avoid using self-reported data which might potentially reflect novel and glorified attitudes towards the idea of using game mechanics (on possible novelty effects in gamification, see e.g. Farzan et al., 2008; Koivisto & Hamari, 2014). With this approach we expected to achieve a higher level of validity.

The study aimed to provide a relatively straightforward and hence generalizable test, with which to investigate the main effects of a representative badge implementation, purposefully designed to mimic the most common implementation of gamification (see Table 4 and Fig. 4). Furthermore, to the best of our knowledge, this study is the longest duration *a/b*-test-style experiment to be conducted on badges with a reasonably high subject count.

The badges were mainly designed by the developers of the service. Researchers commented on the design and steered it towards being as generalizable as possible. This followed the lines of previous work on conceptualizing game badge design pattern (Hamari & Eranti, 2011; Jakobsson, 2011), as well as resembling popular implementation approaches such as those found in Foursquare, the Steam gaming platform and Xbox Live. According to previous work, a badge consists of three main elements: (1) signifier, (2) completion logic and (3) rewards (Hamari & Eranti, 2011). The elements of the badges are described in Table 4.

Moreover, the badges used in the experiment were designed with the above mentioned goal-setting related theories in mind. They provided *clear goals* (including *the specified numeration of goals*) and *feedback*.

The goal was to award badges for all of the core activities of the service: general use activity (frequently browsing/logging in), posting trade proposals, carrying out transactions, and asking/commenting about listed trade proposals. All of these activities were assigned a badge. Moreover, there were additional badges for types of trade proposals (carpooling, giving for free, selling, borrowing, and offers for help). Furthermore, a specialty badge was awarded for those who had given one item for free during December (a Christmas badge).

The rational for how many times a user would have to carry out a certain action to unlock a badge was based on the estimated average use case scenario, in such a way that the first level of the badge (bronze) would be quite easy to unlock so that users would get acquainted with unlocking badges. The second level (silver) was significantly more difficult, and the third (gold) would require a very active use of the service to be unlocked. Thus, users were provided with long-term goals to reach for. For example, the badges related to types of trade proposal the required 2 actions to achieve bronze, 6 for silver and 15 for gold. The unlocked badges were displayed on the users' individual profiles (Fig. 2). Users could also view badges on a separate page linked to every users' profile (Fig. 4). Here they could see which badges they had unlocked (colored) and which badges they had yet to unlock (grey). Users were notified via email for every badge they unlocked.

The badge system underwent general technical and usability
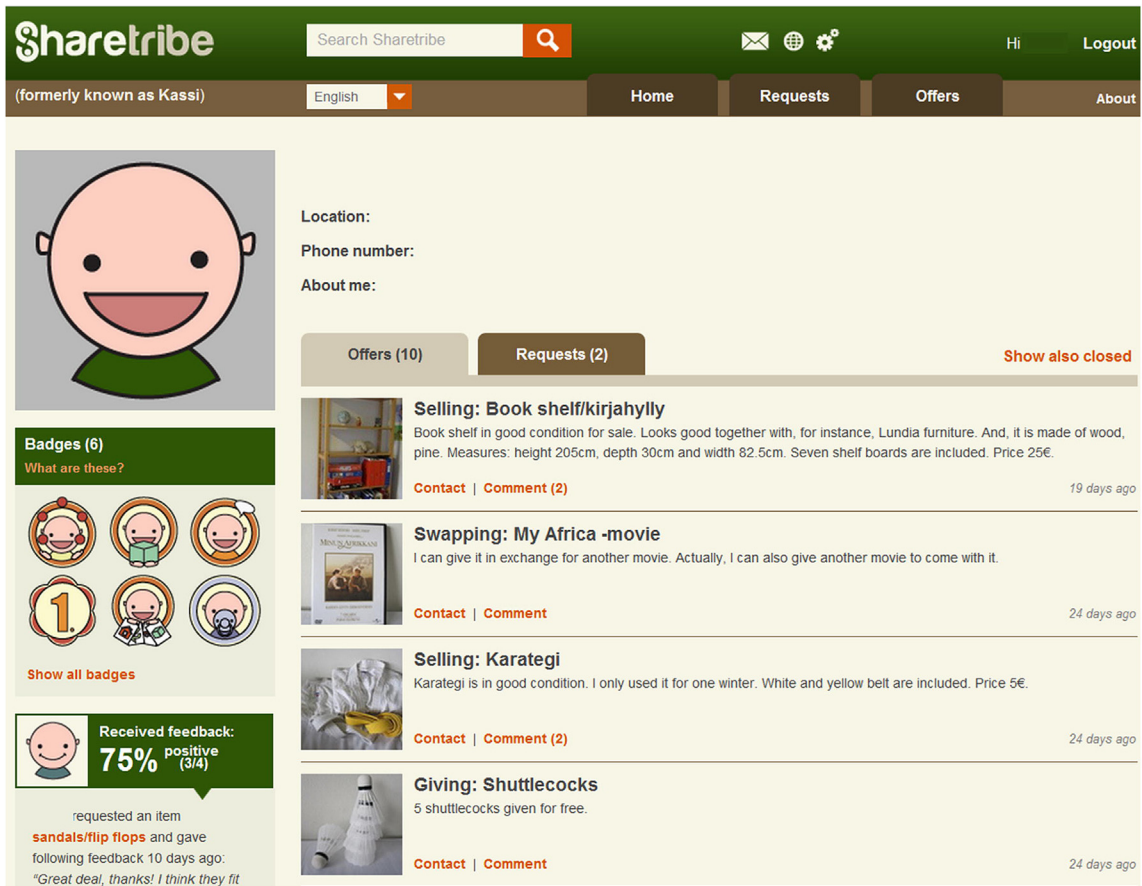
Fig. 2. User profile in Sharetribe.
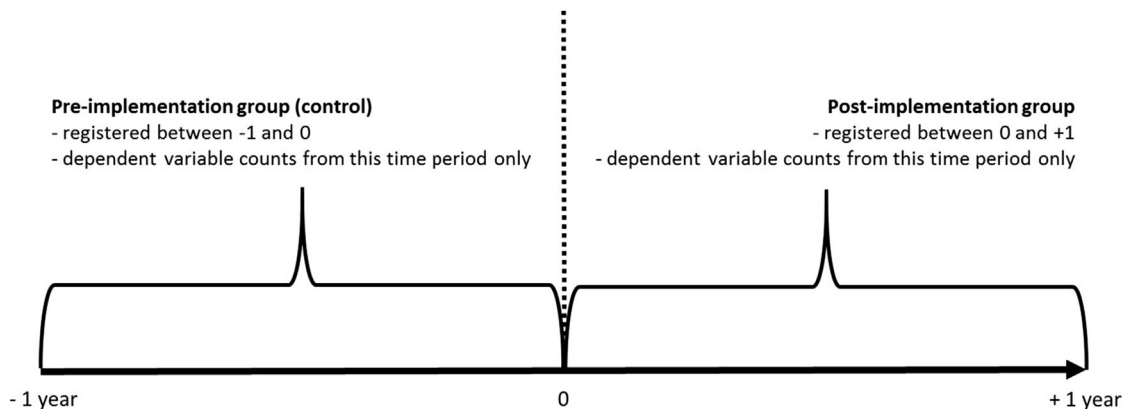
Fig. 3. Pre-implementation and post-implementation groups.

Table 2
Users in treatment groups.

	Count
Group 1: Registered within 1 year before implementation (control)	1410
Group 2: Registered within 1 year after the implementation	1579
Total	2989

testing to guarantee that the system worked as intended and that it was intuitive to use. After implementation, no technical or usability issues were raised.

4. Results

A *t*-test (Table 5) showed a significant difference in the means of all of the dependent variables between the users in the pre-implementation and the post-implementation systems.

A MANOVA test on the dependent variables showed a significant overall effect: $F(4, 2984) = 29.937^{***}$, $p = 0.000$, Wilk's $= 0.961$. We then tested the effects of the treatment condition on different dependent variables separately using ANOVA analyses (Table 5):

Table 3
List of variables.

Independent variable	Dependent variables
Binary variable: • 0 = registered pre-implementation • 1 = registered post-implementation	Number of individual posted trade proposals during the user's registered timeframe[a] Number of individual transactions during the user's registered timeframe[a] Number of individual posted comments during the user's registered timeframe[a] Number of individual page views during user's registered timeframe[a]

[a] For users registered in the pre-implementation phase, the number includes only those activities which occurred between −1 years from implementation to implementation. For users registered in the post-implementation phase, the number includes only those activities which occurred between the implementation and +1 years.

Table 4
The badge design.

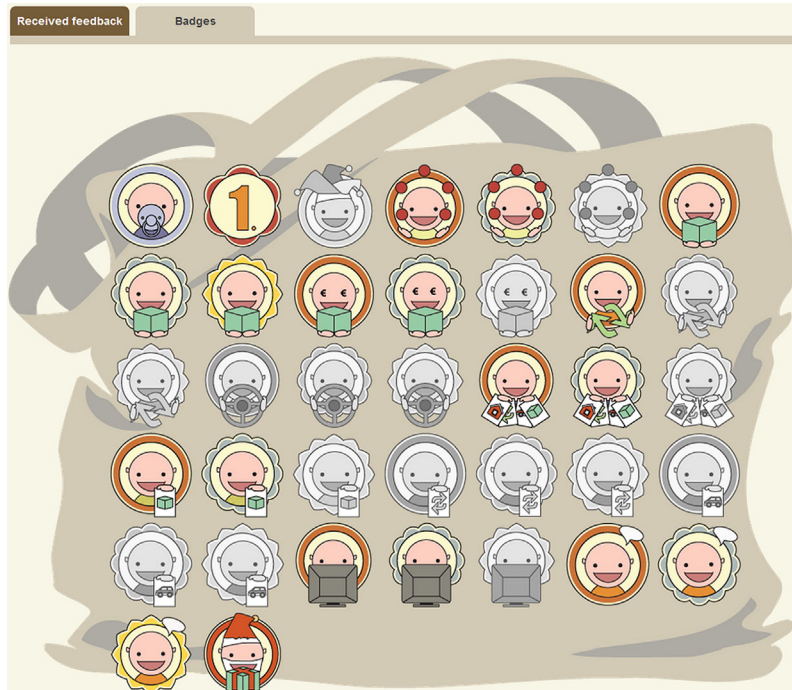| Element/component	Implemented in Sharetribe
Signifier (name, visual, description)	Example badge names were: "Jack of all trades", "Rookie", "Generous", and "Chauffeur". The badge itself and the name represents the type of activity that was carried out in order to unlock the badge. Both are also associated with the level of that badge with color coding and text (bronze/silver/gold) The particular visual style for badges was based on the default generalized avatar picture already present in the service (Fig. 2). This image was already used as the default avatar image for users who had not uploaded one. The same avatar was used in the feedback system for communicating good user feedback (sad <−> happy face). Therefore it was natural to employ a graphical style that connected to the existing theme in the service. In this way the, the badges depicted the same avatar as carrying out different activities within the service (see Figs. 2 and 4) The description describes what the user has to do/has done in order to unlock the badge. For example: "You've been in Sharetribe on five different days. It seems you are on your way to become a regular." (Regular badge) Example image of a badge (Commentator badge) in Sharetribe:
Completion logic	The badge was unlocked when the user had carried out the pre-defined amount of actions. There were no pre-requisites, thus implying that all users were automatically eligible to unlock all badges
Reward	As in other popular services, the only reward from unlocking the badge is that it will be unlocked in the user's profile

Fig. 4. View of the user's badges in Sharetribe.

Table 5
Tests for the differences between the non-gamified and the gamified condition.

		t-Test					ANOVA			Mann −Whitney U	
		Mean	Std. D	*t*	*df*	*p*	*F*	*p*	r^2	*p*	*r*
Trade proposals	Pre	.4504	1.854	5.435	2986.977	.000	29.161	.000	.010	.000	.167
	Post	.8417	2.082								
Accepted transactions	Pre	.0872	.449	9.347	1996.841	.000	79.965	.000	.026	.000	.193
	Post	.4098	1.286								
Comments	Pre	.0943	.432	9.844	1871.545	.000	88.066	.000	.029	.000	.194
	Post	.4794	1.486								
Page views	Pre	45.0695	115.110	7.424	2821.909	.000	52.988	.000	.017	.000	.094
	Post	83.4870	165.652								

$r^2 > 0.010 =$ small, $>0.090 =$ medium, $>0.250 =$ large effect (Cohen, 1988).
$r > 0.100 =$ small, $>0.300 =$ medium, $>0.500 =$ large effect (Cohen, 1988).

trade proposals ($F = 29.161^{***}$, $p = 0.000$, $r^2 = 0.010$), accepted transactions ($F = 79.965^{***}$, $p = 0.000$, $r^2 = 0.026$), comments ($F = 88.066^{***}$, $p = 0.000$, $r^2 = 0.029$), and page views ($F = 52.988^{***}$, $p = 0.000$, $r^2 = 0.017$). The sizes of effect can be classified as falling between small (>0.01) and medium (>0.09) (Cohen, 1988). See Table 5.

The dependent variables are not normally distributed as there are more users with 0 actions than users with 1 action, more users with 1 action than 2 actions and so forth (closer to a Poisson distribution). Therefore we ran the Mann–Whitney U test which is nonparametric. Similarly, we could establish significant (all *p*-values <0.000) differences between the non-gamified and the gamified conditions for all of the dependent variables. See Table 5.

Although these simple tests for differences in means and variances show significant differences between the pre and post implementation groups, we wanted to make sure no confounding factors would affect the results. Firstly the service had grown during the two year duration of the experiment, and therefore we were worried that this might have caused the observed increase in use of the service. The more users there are, the more trades can potentially occur, so in order to control for these network effects (see e.g. Katz & Shapiro, 1985) of the growing service, we calculated a count variable that represented how many users existed in the service at the time a given user was registered. Secondly, the users registering to the service during the pre and post-intervention phases had varying times to accumulate dependent variable counts: users who

registered during the beginning of the studied period had more time before the intervention was introduced (users who registered in the pre-intervention phase), or the end of the experiment (users who registered in the post-adoption phase) (see Fig. 5). Thirdly, we further wanted to employ the GLM Poisson regression model, since the distribution of the dependent variables was naturally closer to a Poisson distribution than to a normal distribution.

As Table 6 indicates, when the potential confounding factors were also included in the regression model, we could still establish a positive effect between the intervention and the dependent variables for accepted transactions, comments and page views. However, the effect on posting trade proposals was no longer significant. According to our expectations, the network effects and the length of time users could potentially use the service also positively affected the dependent variables (excluding network effects on accepted transactions). Being able establish the relationships between the control variables and dependent variables further strengthened the reliability and validity of the study, since both the main effects of the intervention and the effects from the control variables could be established concurrently and also independently.

5. Discussion

This paper reports results of a 2 (1 + 1) year-long field experiment on gamifying a utilitarian trading service by the

- IV A = intervention dummy variable
- IV B = number of months the user had the possibility to accumulate the dv counts
- IV C = network effects: the number of users in the service at time of registration
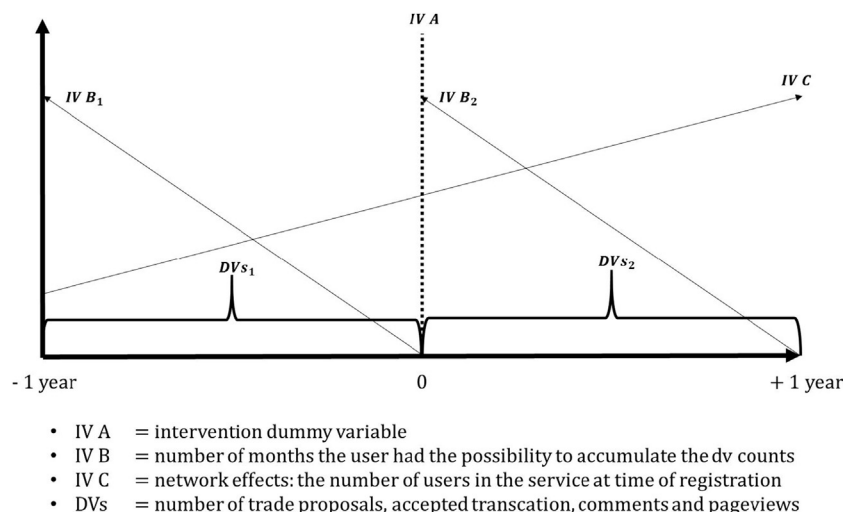- DVs = number of trade proposals, accepted transcation, comments and pageviews

Fig. 5. Experiment variables including controls.

Table 6
Poisson regression with controls.

IVs/DVs	Trade proposals			Accepted transactions			Comments			Page views		
	Wald	Sig.	Exp(B)	Wald	Sig.	Exp(B)	Wald	Sig.	Exp(B)	Wald	Sig.	Exp(B)
Control: Tenure (IV B)	100.172	0.000	1.150	4.735	0.030	1.050	12.130	0.000	1.081	1099.682	0.000	1.048
Control: Network effects (IV C)	10.277	0.001	1.000	0.053	0.819	1.000	24.495	0.000	1.001	5.492	0.019	1.000
Intervention (IV A)	1.785	0.182	1.223	41.371	0.000	4.536	8.678	0.003	1.907	1686.979	0.000	1.819

implementation of badges, which are considered as one of the primary mechanics by which services may be gamified. The study indicates that all of the use-related dependent variables were significantly higher for the post-implementation group which was exposed to gamification. Users were seen to be more likely to actively use the service, list their goods for trade, and comment on listings and to complete transactions. However, when controlling for network effects and tenure, the relationship between the intervention and the amount of trade proposals was no longer significant (see Table 7).

The present study investigated the direct relationship between the gamification implementation and behavioral outcomes. Therefore, there is no reliable way to infer which psychological aspect mediated the effects. However, as Hamari et al. (2014a, 2014b) note, a large portion of the studies on gamification and similar studies in general seem to directly refer to the relationship between the affordances of the system and behavioral change. This was also seen to be the case in this study. A related strong point of the study was that it directly measured actual use, instead of self-reported use. For future studies, we suggest combining experimental setups with surveys that measure latent psychological variables in order to attain more accurate linkages between game mechanics, psychological effects and resultant behavioral manifestations.

Possible such mediators between gamification mechanics and behavior may include, for example, the attainment of clear goals that badges provide. Previous research has demonstrated that clear goals increase behavior, based on the expectations they set and that the completion of goals increases positive emotions such as the experience of self-efficacy and satisfaction (Bandura, 1993). In addition to the goals linked to specific service activities, another goal emerges from the collection behavior associated with badges − unlocking all the badges can also be considered as another goal that badges provide. A more cognitively oriented mechanism by which badges have been postulated to increase goal-related behavior is the way that clear goals make it easier for users to understand how to use the service, and therefore become more efficient (Hamari & Eranti, 2011; Jakobsson, 2011; Montola et al., 2009). Badges also provide feedback which is regarded as an important antecedent to flow and engagement (Csíkszentmihályi, 1990), and this has also been reported to be strongly linked to gamification (Hamari & Koivisto, 2014). However, since activity in

the case service examined in the present study is rather sporadic and the feedback from unlocking badges is not as instantaneous as other implementations. Thus the effect of 'instant feedback' is not likely to be a key mediator in this study, although it might be so in other gamification settings.

Badges also function as social markers, since the earned badges are publicly visible to other users. Therefore, another possible explanation of why badges (and gamification in general) can affect behavior is the social comparison theory (Festinger, 1954). Individuals are more likely to engage in behaviors that they perceive others are also engaged in. Moreover, seeing that other users have earned certain badges and have thus carried out specific activities, provides a social validation that these activities are worthwhile (Cialdini, 2001b).

In the more general discussion around gamification, gamefulness and playfulness (on playfulness see e.g. Martocchio & Webster, 1992; Webster & Martoccio, 1992) are discussed as overarching psychological mediators between gamifying mechanics and behavior (Deterding et al., 2011; Hamari et al., 2015). In traditional theory, these can be linked to the continuum between ludus (structured experience) and paidia (free-form explorative experience) (Caillois, 1961). While badges and gamification intuitively add structure and goals to the experience, they may also increase the experimental nature of service use, since badges might provoke users to try out different aspects of the service with an explorative mindset. Therefore, another interesting vein of further inquiry would be to investigate the relationship of these factors and their role in mediating the effects of gamification efforts.

Another possible explanation for behavioral change could be that the feeling of novelty and curiosity towards the badges. This has been suggested as a possible factor for increased user behavior (Hamari et al., 2014b) and is supported by the findings of other studies (Farzan et al., 2008; Koivisto & Hamari, 2014). The novelty theory of gamification purports that gamification has the ability to change behavior because people are curious towards gamification and look to try it out, thus changing their behavior. However, when the novelty wears off, the changed behavior levels also decrease (see e.g. Farzan et al., 2008; Koivisto & Hamari, 2014). So far, however, novelty has not been directly measured in related studies and therefore further research on the effects of novelty in gamification are needed.

In relation to the psychological aspects that might moderate the

Table 7
Confirmation of hypotheses.

H#	Category	Hypothesis	Supported
1	Productive use	Badges have a positive effect on the number of trade proposals a user makes	Yes, however, the effect remained insignificant in the Poisson regression model when possible confounding factors were controlled for
2	Quality of Use	Badges have a positive effect on the number of transactions a user makes	Yes
3	Social use	Badges have a positive effect on the number of comments a user makes	Yes
4	General use activity	Badges have a positive effect on the number of page views a user has	Yes

effects of gamification: even though badges may provide clear goals, users would need to be committed to pursuing them or else the gamification might end up to be perceived as something trivial (on goal commitments, see Klein et al., 1999; Locke & Latham, 1990). These issues may further depend on several factors such as the nature of the underlying system (for example the utilitarian (Davis, 1989) versus hedonic (Hirschman & Holbrook, 1982; van der Heijden, 2004) aspects of the system), and the type and degree of involvement (cognitive versus affective – Zaichkowsky, 1994) of the user.

Prior studies have also demonstrated individual differences in how the benefits of gamification are perceived (Koivisto & Hamari, 2014). Therefore, future research could also consider the moderating role of, for example, personality differences (McCrae & John, 1992) and player types (Hamari & Tuunanen, 2014; Yee, 2006) on the use and experiences of gamification initiatives. Furthering this line of research could refine the understanding of moderating demographical and user related factors.

5.1. Limitations

Although the present experiment was designed with a high degree of internal validity, it may be possible that some uncontrollable factors remain, related to the sampling. As both groups consisted mainly of university students, this is unlikely to be an issue. It was impossible to gather demographic information on the users. However, as the majority of the users were university students (around 90%) and the rest university staff, we can assume a certain age distribution (approx. 18–30), and a gender split customary to a basic sciences education. This issue was raised with the developers of the service, however they also believed that this would not be an issue. Another possible limitation that we controlled for was the fact that during the treatment phase, there were more service users, simply because the service had grown. There was therefore a greater potential for positive network effects to be seen, i.e. the more users there are, the more potential trade listings there are for people to browse, which might in turn bloat their individual page view counts. We operationalized these network effects by forming a count variable indicating how many users existed at the time a given user registered with the service. By using this variable as part of the analysis we could control these effects (see Table 6).

We also considered whether the measured dependent variables were truly representative of possible user activities within the service, and discussed the issue with the service developers. The dependent variables were deemed to well-represent the full variety of relevant actions available for users of the core activity of the service, including making trade proposals, carrying out trades and commenting on trade proposals. Furthermore, browsing trade proposals was measured by how many individual page loads users had made. We intentionally have not reported whether the independent variables affected how many private messages users had sent to each other, as there was no badge to be earned by sending messages and because the number of messages may have depended upon the other trade activity of the user. Similarly, we did not report how the number of badges was affected by the independent variables for the same reason and there was no significant relationship between the independent variables and the number of messages or the number of earned badges.

Another way of conducting the experiment could have been to compare how the behavior of the same group of users differed between pre and post-implementation timeframes, however it was deemed unfeasible given that the usage rate tends to decline from early use. Should this assumption hold and be strong in effect, there would have been no feasible way to discern the effects of badges from any general usage rate decline. Therefore, in this experiment we followed users registered for one calendar year before the implementation of gamification and users registered one calendar year after the implementation, and believe that any possibility of a slight heterogeneity between the user groups poses far less of an issue than the impact of a declining tendency of use.

Acknowledgements

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.

Alha, K., Koskinen, E., Paavilainen, J., Hamari, J., & Kinnunen, J. (2014). Free-to-play games: Professionals' perspectives. In *Proceedings of Nordic Digra 2014, Gotland, Sweden*.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117–148.

Caillois, R. (1961). *Man, play, and games*. New York, NY: Free Press.

Cialdini, R. (2001). Harnessing the science of persuasion. *Harvard Business Review, 79*(9), 72–79.

Cialdini, R. (2001). *Influence: Science and practice* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Csíkszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper and Row.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Denny, P. (2013). The effect of virtual achievements on student engagement. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 763–772). Paris, France, April 27–May 2, 2013, ACM Press, New York, NY.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic mindtrek conference: envisioning future media environments* (pp. 9–15). Tampere, Finland, September 2011, ACM Press, New York, NY.

Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers and Education, 63*, 380–392.

Electronic Entertainment Design and Research (2007). EEDAR study shows more achievements in games leads to higher review scores. <http://www.eedar.com/News/Article.aspx?id=9> (Accessed 2011).

Farzan, R., DiMicco, J. M., Millen, D. R., Dugan, C., Geyer, W. & Brownholtz, E. A. (2008). Results from deploying a participation incentive mechanism within the enterprise. In *Proceedings of CHI, Florence, Italy*, April 5–10, 2008, pp. 563–572.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140.

Fitz-Walter, Z., Tjondronegoro, D., & Wyeth, P. (2011). Orientation passport: Using gamification to engage university students. In *Proceedings of the 23rd Australian computer-human interaction conference* (pp. 122–125). ACM.

Gartner (2011). Gartner says by 2015, more than 50 percent of organizations that manage innovation processes will gamify those processes. Engham, UK, April 12, 2011. <http://www.gartner.com/it/page.jsp?id=1629214> (Accessed 07.05.14).

Gartner (2012). Gartner says by 2014, 80 percent of current gamified applications will fail to meet business objectives primarily due to poor design. Stamford, CT, November 27, 2012. <http://www.gartner.com/it/page.jsp?id=2251015> (Accessed 07.05.14).

Goldstein, N., Cialdini, R., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*(3), 472–482.

Hakulinen, L., Auvinen, T., & Korhonen, A. (2013). Empirical study on the effect of achievement badges in TRAKLA2 online learning environment. In *Learning and teaching in computing and engineering (LaTiCE)* (pp. 47–54). Macau, March 21–14, 2013. IEEE.

Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications, 12*(4), 236–245.

Hamari, J. (2015). Why do people buy virtual goods? Attitude towards virtual good purchases versus game enjoyment. *International Journal of Information Management, 35*(3), 299–308.

Hamari, J., Eranti, V., 2011. Framework for designing and evaluating game achievements. In *Proceedings of digra 2011 conference: Think design play, Hilversum, Netherlands*, September 14–17.

Hamari, J., & Koivisto, J., (2013). Social motivations to use gamification: An empirical study of gamifying exercise. In *Proceedings of the 21st European conference on information systems, Utrecht, Netherlands,* June 5–8, 2013.

Hamari, J., Koivisto, J., & Pakkanen, T. (2014a). Do persuasive technologies persuade? – A review of empirical studies. In Spagnolli, A. et al. (Eds.), *Persuasive technology, LNCS 8462* (pp. 118–136). Springer International Publishing Switzerland.

Hamari, J., Koivisto, J., & Sarsa, H. (2014b). Does gamification work? – A literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii international conference on system sciences, Hawaii, USA,* January 6–9, 2014.

Hamari, J., Huotari, K., & Tolvanen, J. (2015). Gamification and economics. In S. P. Walz, & S. Deterding (Eds.), *The gameful world: Approaches, issues, applications* (pp. 139–161). Cambridge, MA: MIT Press.

Hamari, J., & Järvinen, A. (2011). Building customer relationship through game mechanics in social games. In M. Cruz-Cunha, V. Carvalho, & P. Tavares (Eds.), *Business, technological and social dimensions of computer games: Multidisciplinary developments.* Hershey, PA: IGI Global.

Hamari, J., & Koivisto, J. (2014). Measuring flow in gamification: Dispositional flow scale-2. *Computers in Human Behavior, 40,* 133–143.

Hamari, J., & Lehdonvirta, V. (2010). Game design as marketing: How game mechanics create demand for virtual goods. *International Journal of Business Science and Applied Management, 5*(1), 14–29.

Hamari, J., & Tuunanen, J. (2014). Player types: A meta-synthesis. *Transactions of the Digital Games Research Association, 1*(2), 29–53.

Hirschman, E., & Holbrook, M. B. (1982). Hedonic consumption: Emerging concepts, methods and propositions. *Journal of Marketing, 46*(3), 92–101.

Huotari K., & Hamari, J. (2012). Defining gamification – A service marketing perspective. In *Proceedings of the 16th international Academic Mindtrek conference* (pp. 17–22). Tampere, Finland, October 3–5, 2012. ACM, New York, NY, USA.

IEEE (2014). Everyone's a gamer – IEEE experts predict gaming will be integrated into more than 85 percent of daily tasks by 2020. <http://www.ieee.org/about/news/2014/25_feb_2014.html> (Accessed 07.05.14).

Jakobsson, M. (2011). The achievement machine: Understanding Xbox 360 achievements in gaming practices. *Game Studies, 11*(1), 1–22.

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review, 75*(3), 424–440.

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research, 43*(1), 35–58.

Klein, H., Wesson, M., Hollenbeck, J., & Alge, B. (1999). Goal Commitment and the goal setting process: Conceptual clarification and empirical synthesis. *Journal of Applied Psychology, 84*(6), 885–896.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior, 35,* 179–188.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., et al. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication, 10*(4).

Locke, E., & Latham, G. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice-Hall.

Malone, T. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*(4), 333–369.

Martocchio, J., & Webster, J. (1992). Effects of feedback and cognitive playfulness on performance in microcomputer software training. *Personnel Psychology, 45*(3), 553–578.

McCrae, R., & John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world.* London, UK: Jonathan Cape.

Montola, M., Nummenmaa, T., Lucerano, A., Boberg, M., & Korhonen, H. (2009). Applying game achievement systems to enhance user experience in a photo sharing service. In *Proceedings of the 13th international academic mindtrek conference: Everyday life in the Ubiquitous Era* (pp. 94–97). Tampere, Finland, September 30–October 2, 2009.

Nunes, J., & Drèze, X. (2006). The endowed progress effect: How artificial advancement increases effort. *Journal of Consumer Research, 32*(4), 504–512.

Paavilainen, J., Hamari, J., Stenros, J., & Kinnunen, J. (2013). Social network games: Players' perspectives. *Simulation and Gaming, 44*(6), 794–820.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals.* Cambridge, MA: MIT Press.

Van de Ven, N., Zeelenberg, M., & Pieters, R. (2011). The envy premium in product evaluation. *Journal of Consumer Research, 37,* 984–998.

van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly, 28*(4), 695–704.

Webster, J., & Martoccio, J. (1992). Microcomputer playfulness: Development of a measure with workplace implications. *MIS Quarterly, 16*(2), 201–226.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology and Behavior, 9*(6), 772–775.

Zaichkowsy, J. (1994). The personal involvement inventory: Reduction, revision, and application to advertising. *Journal of Advertising, 23*(4), 59–70.