Transforming Homo Economicus into Homo Ludens: a field experiment on gamification in a utilitarian peer-to-peer trading service

Juho Hamari (corresponding author)
juho.hamari@uta.fi, Tel.: +358 40 835 9563
School of Information Sciences, 33014 University of Tampere, Finland.

Pre-print

To cite:


Abstract

During recent years, the addition of game mechanics to non-game services has gained a relatively large amount of attention. Popular discussion connects ‘gamification’ to successful marketing and increased profitability through higher customer engagement, however there is a dearth of empirical studies that confirm such expectations. This paper reports the results of a field experiment, gamifying a utilitarian peer-to-peer trading service by implementing the game mechanic of ‘badges’ that users could earn from a variety of tasks. The users (N=3234) were randomly assigned to treatment groups and subjected to different versions of the badge system (a 2x2 design). Results show that the mere implementation of gamification mechanics does not automatically lead to significant increases in use activity in the studied utilitarian service, however those users who actively monitored their own badges and those of others in the study showed an increased user activity.

Keywords: gamification; service marketing; social commerce; badges; achievements; game design; collaborative consumption
1. Introduction

During recent years, the use of game design elements for marketing purposes has rapidly gained a substantial amount of traction among service marketing practitioners, both in games (Hamari and Lehdonvirta, 2010, Hamari and Järvinen, 2011), as well as in non-game contexts (Deterding et al. 2011, McGonigal, 2011, Zichermann and Cunningham, 2011, Huotari and Hamari, 2012). This development of enhancing services with game elements has been dubbed as ‘Gamification’. Following the successes of social networking services (Facebook), games (Angry Birds) and location-based services (Foursquare), marketers have started to apply these innovations in non-game contexts. Gamification has already been applied in several areas, including the promotion of greener energy consumption (EcoIsland), 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). Gartner (2011) predicts that more than 50% of organisations will gamify innovation processes by 2015, as gamification provides accelerated feedback, clear goals and challenging tasks. Clearly, much has been invested in the idea of gamification and its primary mechanic has been the use of ‘badges’ which are used to reward users for favorable pre-defined behaviors in a service. The strong belief in the effectiveness of gamification has mainly been based on the conception that because games are "fun", any service that uses the same mechanics should also prove to be more valuable and engaging. However, there is a dearth of empirical studies investigating the effects which result from gamifying services, in regard to customer and user behavior.

In this paper, we study the effects of gamification on user retention, namely those of usage activity and quality, as well as social interaction within a service. The research problem was approached through a field experiment where a utilitarian peer-to-peer trading service was gamified by a process of implementing badges (Hamari and Eranti, 2011). Badges have been regarded as the blueprint of gamification to such a degree, that gamification has been even referred to as “badgification”. In the experiment, people could unlock badges by completing common actions and tasks within the service. The
experiment focused on investigating whether: 1) the mere implementation of the goal oriented features and the social features of badges, and 2) whether the active pursuit of badges, were positively associated with increased service usage, quality, as well as social interaction within the case service.

The paper proceeds as follows: In the background section we discuss gamification and badges. In the following hypotheses section we outline possible theoretical foundations that could explain the effects of badges and gamification and propose hypotheses for the study. In the fourth section we describe the research process, data collection, and the service in which the experiment was conducted. The fifth section outlines the results and the final section elaborates on these results and suggests direction for further research on gamification in utilitarian and hedonic information systems.

2. Background

2.1. Gamification

The concept of gamification has strongly divided opinions. Whilst some deem it as a new name for old marketing tools, others think of it as the latest way of exploiting customers and it has been received as a genuine way of enhancing the value of services. Nevertheless, gamification has quickly become a trend in service marketing. Gamification can be situated in a previously unoccupied space of marketing thinking. For example, previously, full games have been used as a value added service on product webpages and ‘serious games’ have been used in educating consumers. Additionally, loyalty programs can resemble game mechanics, which have been used to offer economic benefits to customers who in exchange demonstrate customer loyalty. However, the previous ways in which games and consumer behavior have come together in marketing are not exactly the same as gamification’s popular conception.

Gamification can be defined in two ways:

1) The use of game elements in non-game contexts (Deterding et al., 2011).
A process of providing affordances for gameful experiences which support the customers’ overall value creation (Huotari and Hamari, 2012).

The latter conceptualization is rooted in service dominant logic (Vargo and Lusch, 2004), which suggests that customers are the creators of value and the company can merely provide affordances for the customer to experience ‘gamefulness’. This conceptualization of gamification implicitly states that the customer in the end determines whether they are engaged in gameful experiences and whether consequently the perceived value of the service is increased. The other difference between the definitions is that Huotari and Hamari (2012) emphasize that for gamification to have an effect on retention and customer loyalty, the customers should first be engaged in gameful experiences and that the mere addition of game elements does not necessarily guarantee a successful gamification. However, in the general discussion of gamification, the idea prevails that gamification simply refers to adding game mechanics into a service which in turn automatically becomes more engaging and attains a better retention of customers.

In addition to linking gamification with service dominant logic, the ‘gameful experience’ could be linked with hedonic usage patterns and consumption (Hirschman and Holbrook, 1982), as well as intrinsic (as opposed to extrinsic) motivations (Deci and Ryan, 1985) towards the use of information systems and services. Therefore, gamification can be viewed as an attempt to convert utilitarian services into more hedonically oriented (see e.g. van der Heijden, 2004 on hedonic information systems). In terms of IS theory, this sits well within the long-run context of studying technology acceptance (e.g. Davis, 1989, Venkatesh, 1999), continuous usage intentions (Bhattacherjee, 2001, Hsieh et al., 2008) and especially with the more recent understanding of the hedonic nature of novel services which has called for the measurement of more hedonic constructs such as perceived enjoyment, flow, immediate feedback, clear goals (Csíkszentmihályi, 1990) and social comparison (Festinger, 1954). As such, gamification might offer an interesting vein for this continuum of research.

In principle, this is also how gamification differs from loyalty programs, although it is often used for pursuing similar goals. Most loyalty programs aim to offer economic benefits (redeemable by points) from
the continuous use of services which most likely invokes extrinsic motivations. These have in turn been demonstrated to be detrimental to intrinsic motivations, autonomy and creativity (Deci et al., 1999). Game mechanics in themselves however, do not provide economic benefits for the users, but are believed to add value to the service via transformation of the usage motivations and intentions (Huotari and Hamari, 2012).

2.2. Badges - the blueprint of gamification

Badges have been considered as the blueprint of gamification and have been the primary game mechanic in popular gamified applications such as Foursquare. Industry studies have found that the addition of badges to games has led to better critical reception and increased revenue (EEDAR, 2007). In fact, large game console publishers such as Microsoft, demand that game developers include badges in the games that are published for Xbox consoles. 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 whose fulfillment is stored outside the scope of the core activities of a service (Montola et al., 2009, Jakobsson, 2011). 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 and Eranti, 2011). 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). Previous works hypothesize that badges can provide clear goals, and signal reputation and status as well as affirm it (Hamari and Eranti, 2011). In this paper, we report a field experiment where we study the effects gamification that aims to provide features for clear goals and a social comparison on usage activity.
3. Hypotheses

We propose two sets of hypotheses divided between social comparison (hypotheses 1-4) and clear goals (hypotheses 5-8), as well as two sets between investigating whether the mere addition of game mechanics (marked a) and active exposure (marked b) are positively associated with increased usage behavior. The latter sets of hypotheses are related to internal validity; whether the impact from being actually exposed to the gamified features is associated with usage activity. The hypotheses were divided in this manner because it was thought it may have been possible that those users who have been clearly exposed to gamification might show a greater level of activity, or that it might not be possible to determine any significant association between gamifying a service and increased usage activity. The dependent variables are the number of trade proposals a user has posted, the number of transactions a user has completed, the number of comments a user has posted, and the number of page views a user has generated.

3.1. Hypotheses 1-4: Social comparison increases usage activity

One of the rationales behind gamification has been to harness the persuasive power that emerges when people compare their points and badges amongst each other, and so benchmark themselves. In general, this phenomenon is called social comparison (Festinger, 1954), and forms an over-arching concept for other, more specific theories related to the 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 and Koivisto, 2013).

Social proof theory (Cialdini, 2001a, 2001b, Goldstein et al., 2008) predicts that individuals are more likely to engage in behaviors that 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 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 the same behaviors. Van de Ven et al. (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. If these phenomena are present to a significant degree, then an increase in use activity for those users who were enabled to compare badges as well as for those users who have actively viewed the badges of other users would be anticipated. We propose the following hypotheses related to social comparison in a gamified setting:

**Hypothesis 1a (Social comparison: productive actions).** Users who are enabled to compare their badges with the badges of other users create more trade proposals.

**Hypothesis 1b (Social comparison: productive actions).** The number of times a user has viewed the badges of other users has a positive effect on the number of trade proposal the user makes.

**Hypothesis 2a (Social comparison: quality of action).** Users who are enabled to compare their badges with the badges of other users complete more transactions.

**Hypothesis 2b (Social comparison: quality of action).** The number of times a user has viewed the badges of other users has a positive effect on the number of transactions the user completes.

**Hypothesis 3a (Social comparison: social interaction).** Users who are enabled to compare their badges with the badges of other users post more comments.

**Hypothesis 3b (Social comparison: social interaction).** The number of times a user has viewed the badges of other users has a positive effect on the number of comments the user posts.

**Hypothesis 4a (Social comparison: usage activity).** Users who are enabled to compare their badges with the badges of other users generate more page views.

**Hypothesis 4b (Social comparison: usage activity).** The number of times a user has viewed the badges of other users has a positive effect on the number of page views the user generates.
3.2. **Hypotheses 5-8: Goal setting - clear goals increase usage activity**

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; and 3) the completion of goals leads to increased satisfaction which, in turn, leads to increased future performance with 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 the goals are clearly specified in terms of how many times they have to be completed, the rate of completion of the tasks increases when compared to a condition where the number of times the task has to be completed is not specified (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 (Montola et al., 2009, Jakobsson, 2011, Hamari and Eranti, 2011) 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 the service. In a larger context, goals are regarded as a central game mechanic (Salen and Zimmermann, 2004) and have been demonstrated to exert persuasive power even when the progression towards them was illusionary (Kivetz et al., 2006, Nunes and Dreze, 2006). Clear goals are also one of the main dimensions of the flow theory (Csikszentmihályi, 1990) which predicts that having clear goals and immediate feedback supports the emergence of the ‘flow state’ – where the user’s skills and the challenge of the task are optimally balanced.

Even though users would be offered clear goals as described above, the users need to be committed to the goals in order for the hypothesized effects of increased motivation, engagement and performance to arise (Klein et al., 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 the goals and be less likely to neglect them.
The badges in the experiment were designed with the above goal-setting related theories in mind. They provided *clear goals* (including *the specified numeration of goals*) and *immediate feedback*; however their effect on *performance* (the dependent variables) may well be dependent on *goal commitment*. We propose the following hypotheses related to goal setting:

**Hypothesis 5a (Goal setting: productive actions):** Users who are enabled to have clear goals through badges create more trade proposals.

**Hypothesis 5b (Goal setting: productive actions):** The number of times a user has viewed their own badges has a positive effect on the number of trade proposals the user makes.

**Hypothesis 6a (Goal setting: Quality of actions):** Users who are enabled to have clear goals through badges complete more transactions.

**Hypothesis 6b (Goal setting: Quality of actions):** The number of times a user has viewed their own badges has a positive effect on the number of transactions the user completes.

**Hypothesis 7a (Goal setting: Social interaction):** Users who are enabled to have clear goals through badges post more comments.

**Hypothesis 7b (Goal setting: Social interaction):** The number of times a user has viewed their own badges has a positive effect on the number of comments the user posts.

**Hypothesis 8a (Goal setting: Usage activity):** Users who are enabled to have clear goals through badges generate more page views.

**Hypothesis 8b (Goal setting: Usage activity):** The number of times a user has viewed their own badges has a positive effect on the number of page views the user generates.
4. Methods and data

4.1. Data collection and the case service

Sharetribe (https://www.sharetribe.com/) is an international peer-to-peer trading service which offers its service package to a variety of organizations. The 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 and 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.

![Figure 1: Front page of Sharetribe](image-url)
“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 being targeted to narrow local communities such as an organization or town districts and by also 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 offered for anyone to download. The reason for having many "tribes" is to emphasize local communities, trust and information access, and also to diminish transaction costs and costs related to shipping.

Figure 2: User profile in Sharetribe
4.2. Field experiment

The field experiment was setup in the Sharetribe service and data was gathered from the implementation of badges at the beginning of December 2010 until the end of July 2012. During this time the service remained the same without any major upgrades.

The existing users were evenly and randomly assigned to four test groups (Tables 1 and 2). Users who registered after the implementation were further randomly assigned to one of the groups.

**Table 1: Experimental groups – Independent variables**

<table>
<thead>
<tr>
<th>Ability to see from which actions one can unlock badges (clear goals)</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to view other users’ badges (social comparison)</td>
<td>No</td>
<td>Group 3</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Group 1</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Group 2</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Group 4</td>
</tr>
</tbody>
</table>

The data consists of a database of users of the Sharetribe Aalto University site who registered during the experiment timeframe (n=3234), including the number of trade proposals, accepted transactions, comments posted and how many individual page views a user undertook. We selected only users who had registered during the experiment timeframe because older users have existing trade proposals in the service and would therefore have accumulated actions during the experiment timeframe which would not have been affected by the experiment. We selected the Aalto University Sharetribe site for the experiment because it is the largest implementation of Sharetribe of the several hundred installations world-wide.

The experiment was purposefully conducted as a field experiment in a real existing service, rather than in a laboratory setting in which respondents would have been asked to assume a hypothetical scenario of a badge system. In this way we could avoid using self-reported data which might potentially reflect the novel and glorified attitudes towards the idea of using game mechanics. With this approach we expected
to achieve a higher level of validity. The generalizability of our findings is further explored in the discussion section of the paper.

Table 2: Users in treatment groups

<table>
<thead>
<tr>
<th>Group 1: Both features disabled (control)</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2: Social comparison condition</td>
<td>802</td>
<td>24,8</td>
</tr>
<tr>
<td>Group 3: Clear goals condition</td>
<td>790</td>
<td>24,4</td>
</tr>
<tr>
<td>Group 4: Both conditions enabled</td>
<td>837</td>
<td>25,9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>3234</td>
<td>100 / 100</td>
</tr>
</tbody>
</table>

For the experiment, the badges were designed in adherence to previous work on conceptualizing the badge game design pattern (See Hamari and Eranti, 2011), as well as to resemble popular implementation approaches such as that found in Foursquare, the Steam gaming platform and Xbox Live. Table 3 describes the elements of the badges. According to previous works, a badge consists of three main elements: 1) signifier, 2) completion logic and 3) rewards (Hamari and Eranti, 2011).

The users could unlock badges for typical actions within the service, such as commenting on other peoples' trade proposals, submitting proposals of their own, completing trades and even for using the service for a prescribed amount of consecutive days. The unlocked badges were displayed on the users' individual profiles which were viewable both by the owner of the badges and other users in the respective treatment groups. Users were notified via email for every badge they unlocked.
Users could also view badges on a separate page linked to every users' profile (Figure 3), where they could see which badges they had unlocked (colored) and which badges they were yet to unlock (grey). Furthermore, users in the respective treatment groups could also see which activities were unlocked or could unlock the badges.

Table 3: The badge design

<table>
<thead>
<tr>
<th>Element / component</th>
<th>Implemented in Sharetribe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signifier (name, visual, description)</td>
<td>The badges have a humoristic name and a badge itself 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).</td>
</tr>
<tr>
<td></td>
<td>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). This text is only visible for people in the experimental conditions related to clear goals.</td>
</tr>
</tbody>
</table>
The completion logic does not include any hidden rules. All that has to be done in order to unlock a badge is mentioned in the description component unless the person is in one of the experiment conditions which are not able to see clear goals. The badges have no pre-requirements for unlocking them.

### Reward

As in other popular services, the only reward from unlocking the badge is that it will be unlocked in the user’s profile.

## 5. Results

A simple t-test (Table 4) on the dependent variables did not show any significant differences between the experiment conditions.

### Table 4: t-tests on dependent variables between the experimental groups

<table>
<thead>
<tr>
<th>Test group id</th>
<th>Social comparison / Clear goals</th>
<th>Trade proposals Mean</th>
<th>Trade proposals SD</th>
<th>Accepted transactions Mean</th>
<th>Accepted transactions SD</th>
<th>Comments Mean</th>
<th>Comments SD</th>
<th>Page views Mean</th>
<th>Page views SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No / No</td>
<td>1.00</td>
<td>3.08</td>
<td>0.52</td>
<td>1.61</td>
<td>0.62</td>
<td>2.20</td>
<td>115.18</td>
<td>263.77</td>
</tr>
<tr>
<td>2</td>
<td>Yes / No</td>
<td>1.15</td>
<td>3.55</td>
<td>0.49</td>
<td>1.64</td>
<td>0.70</td>
<td>2.61</td>
<td>106.47</td>
<td>281.01</td>
</tr>
<tr>
<td>3</td>
<td>No / Yes</td>
<td>0.92</td>
<td>3.25</td>
<td>0.48</td>
<td>1.48</td>
<td>0.63</td>
<td>2.14</td>
<td>97.32</td>
<td>243.55</td>
</tr>
<tr>
<td>4</td>
<td>Yes / Yes</td>
<td>1.04</td>
<td>3.13</td>
<td>0.40</td>
<td>1.35</td>
<td>0.63</td>
<td>2.72</td>
<td>97.84</td>
<td>252.58</td>
</tr>
</tbody>
</table>

A multivariate test (MANOVA) was performed on the effects of the possibilities to compare badges with other users, $F(4, 3227) = 1.679, p = 0.152, \text{Wilk's} = 0.998, \eta^2 = 0.002$; the ability to see from what actions
one can unlock badges, $F(4, 3227) = 0.709, p = 0.568$, Wilk’s = 0.999, $\eta^2 = .001$; and the interaction of the features, $F(4, 3227) = 0.716, p = 0.581$, Wilk’s = 0.999, $\eta^2 = .001$, on the dependent variables: the amount of trade proposals, accepted transactions, comments posted or page views. These tests did not yield any significant results.

However, this sample included all the users in the data who had registered into the service during the experiment timeframe. Therefore, we moved to a more confined sample population in order to increase the level of internal validity. From the data, we selected only users who had actively used the service after the implementation of badges by selecting only those users from the data that had at least 100 page views. This way we could be confident that all the users in the sample had the possibility of being exposed to the experimental conditions. Even with this sub-sample however, the results did not change remarkably: social comparison feature: $F(4, 716) = 1.549, p = 0.186$, Wilk’s = 0.991, $\eta^2 = 0.009$, or the clear goals feature, $F(4, 716) = 0.320, p = 0.865$, Wilk’s = 0.998, $\eta^2 = 0.002$, or their interaction , $F(4, 716) = 0.507, p = 0.731$, Wilk’s = .997, $\eta^2 = 0.003$.

We then tested the individual hypotheses by exploring the effects on individual dependent variables using separate ANOVA analyses. However, even here, we were unable to determine any significant effects from the two features on any of the dependent variables (Table 5) and therefore we were unable to find evidence to support hypothesis 1-8a.

**Table 5: Test of hypotheses 1-8a**

<table>
<thead>
<tr>
<th>H#</th>
<th>Independent variable</th>
<th>Dependent variable</th>
<th>Result (users registered during the experiment n=3234)</th>
<th>Results (only active users - at least 100 page views n=723)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Comparison Number of trade proposals</td>
<td>$1.265$</td>
<td>$0.261$</td>
<td>$0.000$</td>
</tr>
<tr>
<td>5a</td>
<td>Goal Number of accepted transactions</td>
<td>$0.695$</td>
<td>$0.405$</td>
<td>$0.000$</td>
</tr>
<tr>
<td></td>
<td>Comparison x Goal</td>
<td>$0.022$</td>
<td>$0.882$</td>
<td>$0.000$</td>
</tr>
<tr>
<td>2a</td>
<td>Comparison Number of accepted transactions</td>
<td>$1.131$</td>
<td>$0.288$</td>
<td>$0.000$</td>
</tr>
<tr>
<td></td>
<td>Goal Number of accepted transactions</td>
<td>$1.405$</td>
<td>$0.236$</td>
<td>$0.000$</td>
</tr>
<tr>
<td></td>
<td>Comparison x Goal</td>
<td>$0.143$</td>
<td>$0.705$</td>
<td>$0.000$</td>
</tr>
<tr>
<td>3a</td>
<td>Comparison Number of</td>
<td>$0.174$</td>
<td>$0.677$</td>
<td>$0.000$</td>
</tr>
</tbody>
</table>
We then tested whether we could find support for hypothesis 1-8b pertaining to whether the active exposure to gamified elements has a positive effect on the dependent variables. The exposure was measured via the number of views of the badge pages of other users (social comparison condition 1-4b) and the number views of the users own badge page (clear goals condition – 5-8b). Multivariate testing (MANCOVA) on the effects derived from viewing other users’ badges \((F(4, 3228) = 5.814, p = 0.000***, \text{Wilk's} = 0.993, \eta^2 = 0.007)\), viewing the users own badges \((F(4, 3228) = 565.361, p = 0.000***, \text{Wilk's} = 0.588, \eta^2 = 0.412)\) and their interaction \((F(4, 3228) = 58.324, p = 0.000***, \text{Wilk's} = 0.933, \eta^2 = 0.067)\) all showed significant results. However, the effect of viewing other users’ badges was relatively small.

We then moved on to testing hypotheses individually by using ANCOVA tests. The results showed that the amount of views of the users own badges was positively associated with all the dependent variables, whereas the amount of views of other people’s badge pages was only positively associated with the number of submitted trade proposals (Table 6). Based on these tests we can conclude that comparing badges does seem to have a positive effect on use, however, it is so small that the effects were difficult to establish for the different dependent variables independently and the only significant effect from comparison could be established in the amount of trade proposals a user makes.

**Table 6: Test of Hypotheses 1-8b.**

<table>
<thead>
<tr>
<th>H#</th>
<th>Independent variable - Views to</th>
<th>Dependent variable</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>badge pages of others</td>
<td>Number of trade proposals</td>
<td>(F = 5.450) (p = 0.020**) (\eta^2 = 0.002)</td>
</tr>
<tr>
<td>5b</td>
<td>own badge page</td>
<td>Number of trade proposals</td>
<td>(F = 810.885) (p = 0.000***) (\eta^2 = 0.201)</td>
</tr>
<tr>
<td>2b</td>
<td>badge pages of others</td>
<td>Number of accepted transactions</td>
<td>(F = 2.247) (p = 0.134) (\eta^2 = 0.001)</td>
</tr>
<tr>
<td>6b</td>
<td>own badge page</td>
<td>Number of accepted transactions</td>
<td>(F = 1034.045) (p = 0.000***) (\eta^2 = 0.242)</td>
</tr>
</tbody>
</table>
Additionally, we found that only 38 users had visited other users’ badge pages, whereas 664 users had visited their own badge page. Thus, it might not be surprising that we could not find any strong effects derived from comparing badges with other users. Instead, the fact that so few users had demonstrated any interest in the badges of other users implies even more essential challenges to gamification are to be faced. This result implies that gamification clearly does not seem to be effective in all contexts, not necessarily because it would fail to arouse the anticipated psychological effects previously proposed, but instead because it can fail with regards to users becoming interested in the gamified features of the service in the first place. If we consider hedonic services such as games, where people by default are oriented towards gameful interaction, then the situation can be dramatically different.

It is commonplace to use ANOVA or similar types of analysis even though the data was non-normal. Also in this study, the dependent variables are not normally distributed as there were more users with 0 actions than users with 1 action, more users with 1 action than 2 actions and so forth. Therefore we ran the test again using the Mann-Whitney U test which is nonparametric and especially suitable for handling non-normal data. Even here however, the results remained insignificant (p-values: H1a 0.972, H2a 0.256, H3a 0.795, H4a 0.193, H5a 0.965, H6a 0.745, H7a 0.430, and H8a 0.169). The same was also the case with the sub-sample consisting only of active users (≥ 100 page views): (p-values: H1a 0.084, H2a 0.136, H3a 0.568, H4a 0.509, H5a 0.916, H6a 0.934, H7a 0.882, and H8a 0.399).

**Table 7: Confirmation of hypotheses**

<table>
<thead>
<tr>
<th>H#</th>
<th>Hypothesis</th>
<th>Supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td><em>Social comparison: productive actions</em> Users who are enabled to compare their badges with the badges of other users create more trade proposals.</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Social comparison: productive actions</td>
<td>The number of times a user has viewed the badges of other users has a positive effect on the number of trade proposals the user makes.</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2a</td>
<td>Social comparison: quality of actions</td>
<td>Users who are enabled to compare their badges with the badges of other users complete more transactions.</td>
</tr>
<tr>
<td>2b</td>
<td>Social comparison: quality of actions</td>
<td>The number of times a user has viewed the badges of other users has a positive effect on the number of transactions the user completes.</td>
</tr>
<tr>
<td>3a</td>
<td>Social comparison: social interaction</td>
<td>Users who are enabled to compare their badges with the badges of other users post more comments.</td>
</tr>
<tr>
<td>3b</td>
<td>Social comparison: social interaction</td>
<td>The number of times a user has viewed the badges of other users has a positive effect on the number of comments the user posts.</td>
</tr>
<tr>
<td>4a</td>
<td>Social comparison: usage activity</td>
<td>Users who are enabled to compare their badges with the badges of other users generate more page views.</td>
</tr>
<tr>
<td>4b</td>
<td>Social comparison: usage activity</td>
<td>The number of times a user has viewed the badges of other users has a positive effect on the number of page views the user generates.</td>
</tr>
<tr>
<td>5a</td>
<td>Clear goals: productive actions</td>
<td>Users who are enabled to have clear goals through badges create more trade proposals.</td>
</tr>
<tr>
<td>5b</td>
<td>Clear goals: productive actions</td>
<td>The number of times a user has viewed their own badges has a positive effect on the number of trade proposals the user makes.</td>
</tr>
<tr>
<td>6a</td>
<td>Clear goals: quality of actions</td>
<td>Users who are enabled to have clear goals through badges complete more transactions.</td>
</tr>
<tr>
<td>6b</td>
<td>Clear goals: quality of actions</td>
<td>The number of times a user has viewed their own badges has a positive effect on the number of transactions the user completes.</td>
</tr>
<tr>
<td>7a</td>
<td>Clear goals: social interaction</td>
<td>Users who are enabled to have clear goals through badges post more comments.</td>
</tr>
<tr>
<td>7b</td>
<td>Clear goals: social interaction</td>
<td>The number of times a user has viewed their own badges has a positive effect on the number of comments the user posts.</td>
</tr>
<tr>
<td>8a</td>
<td>Clear goals: usage activity</td>
<td>Users who are enabled to have clear goals through badges generate more page views.</td>
</tr>
<tr>
<td>8b</td>
<td>Clear goals: usage activity</td>
<td>The number of times a user has viewed their own badges has a positive effect on the number of page views the user generates.</td>
</tr>
</tbody>
</table>
5.1. Limitations & discussion on unsupported hypotheses

Given that the phenomenon under examination is relatively novel, it is still difficult to say which exact psychological theories can explain the effects stemming from badges or other game mechanics. We did however discuss several theories from motivational and social psychology as probable candidates to explain the possible effects of gamification. Therefore, for further studies we suggest measuring latent psychological variables through surveys in order to attain more accurate linkages between game mechanics, psychological effects and resultant behavioral manifestations.

With regards to the present experiment, there is no way to infer directly whether game mechanics were able to arouse the hypothesized psychological effects, such as social influence or goal commitment, but merely to measure whether the gamified features had an effect on behavioral outcomes. Furthermore, the unsupported hypotheses do not imply that the hypothesized psychological effects do not exist, but rather that gamification failed to arouse such psychological effects in the sporadic utilitarian context of the experiment. We hypothesized that gamification would positively affect the number of productive actions users carry out within a utilitarian service. These hypotheses were mainly based upon the considerably large hype in the service marketing sector (e.g. Gartner, 2011). There was no strong previous scientific evidence on the effectiveness of gamification and instead the hypotheses here are more based on anecdotal evidence and therefore the study, although confirmatory in nature, has to be regarded as exploratory at this stage of the research on gamification. From this perspective and taking into account more recent discussion on gamification (see e.g. Gartner, 2012 which hints that 80% of gamified applications will fail), then the results might not be so surprising after all and therefore it could have been equally hypothesized that there would be no effect. Positive results from mere implementation of gamification alone might have required more elaborations as to why such an effect exists.

In the case of Sharetribe, it can be fathomed that badges do not offer considerable value to the users. In retrospect, it would have been more surprising to find that the mere addition of badges and enabling users to compare them and attain clear goals would have significantly increased usage activities
in a service where people use the service only as much as they need to in order to carry out their sporadic trading. This utilitarian use though is not unique to the case service and we believe that the results are generalizable to other utilitarian services. The results do however bring forth an interesting further question: How does gamification work in more hedonic services where people use the service not because of extrinsic reasons, but rather for its enjoyment value and in services where users return to use the service because they either enjoy the activity or want to keep in touch with other people in the service. Further such research questions are discussed in the following section.

We also considered whether the measured dependent variables are truly representative of the possible user activities within the service and discussed the issue with the developers of the service. The dependent variables were deemed to well represent the entire 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 means of how many individual page loads users had made. We intentionally did not report whether the independent variables affected how many private messages users had sent to each other as there was no badge to be earned from 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.

Although the data for the study is sufficiently large with regards to registered users, the number of times users have carried out different activities (dependent variables) on average is quite low (Table 8). This also further justified us to run extra analyses with a more active sub-sample in which users had an acceptable number of actions carried out. As reported above, this sub-sample further strengthened the results.

**Table 8: Means of the dependent variables in the data**
Sharetribe represent a typical start-up looking to grow customer engagement via gamification (Zichermann and Cunningham, 2011). The experiment conducted in this study well emulates a typical scenario where gamification is commonly implemented into a relatively new service with a relatively small initial user base. As discussed previously, we found that only a relatively small portion of users became interested in badges and therefore we were unable to find support for the first hypotheses (marked a) which pertained to the question as to whether the mere implementation of gamification is effective in encouraging overall user behavior. A probable explanation for the failed gamification implementations (see e.g. Gartner, 2012) in general can stem from the lack of interest towards such mechanics when the user motivations are otherwise extrinsic to the service itself, such as selling ones belongings. However, we did find that for those users who actively monitored their own badges, the usage activity was also higher. This suggests that in a large service with a larger user base, gamification can be highly effective since it will affect at least some proportion of the users.

6. Discussion and directions for further research

This paper reported results of a 1.5 year-long field experiment on gamifying a utilitarian trading service by the implementation of badges, which have been considered the primary mechanic through which services have been gamified. In the study we could confirm that users who had actively exposed themselves to
badges in Sharetribe were also significantly more likely to actively use the service, list their goods for trade, comment on listings and complete transactions. Furthermore, our results indicate that actively browsing other users’ badges was positively associated with posting trade proposals in the service. However, we could not find support for the claims that implementing gamified features would alone lead to significant overall increases in usage frequency, quality or social interaction in a utilitarian trading service.

The unexciting result related to the lack of overall effects achieved by the introduction of gamified elements could be explained by several factors, such as a low goal commitment (Locke and Latham 1990, Klein et al., 1999) towards the badges, which we hypothesized to be pre-requisite for the badges to arouse the wanted effects. A low goal commitment could be explained by a number of different conditions within the gamified setting, such as the nature of the underlying service. We hypothesize for future studies that users in such a focused utilitarian service concentrate more on pre-meditated utilitarian activities and exercise a considerably more cognitive involvement rather than affective involvement (see Zaichkowsy, 1994). Therefore, the hedonic service elements could be chosen to be ignored by the majority of the user population. For future studies, we suggest measuring the involvement of the users (Zaichkowsy, 1994) and using it as a moderator for predicting behavioral intentions towards continuous use intentions (Bhattacherjee, 2001) and other measurements related to usage activities. We would also suggest in this context, that the direct measurement of goal commitment may provide useful information for future service development.

Another possible explanation for low goal commitment and affective involvement could be that badges were introduced long after the launch of the service. As such, the user population had not expected ‘gameful’ interactions. If we consider popular gamified services, such as Foursquare, they have been advertised as gameful services from the outset, and consequently these services attract users who have preferences that match towards gameful interaction. Therefore, it might be easier to demonstrate the effectiveness of gamification in environments which have attracted a user populace that would be
receptive to gameful interaction. In the present experiment, gamification was implemented in a strictly utilitarian service where the user population had registered in order to trade goods and services, without any knowledge of the future implementation of gamified features. However, we suggest that further studies also be undertaken to investigate how temporal differences in implementation, affect the technology acceptance (Davis, 1989) of gamified features.

Trading services can be seen to have patterns of sporadic use where users log in to carry out premeditated searches for offers and to list their own goods or services. Gamification and badges, on the other hand, rely on persistence. Badges are reputation indicators and rewards that persist in the users’ profile as a social indicator. However, in the larger context of the use of such services, their role might not be significant enough to fundamentally change the way these services are being used. It is conceivable that if the use of a service or a system is sporadic, then gamification might not be seen to hold enough value by the majority of users. The sporadic nature of such services also means that there are no peers who actively use the service for hedonic or social purposes and therefore the role of aspects related to social comparison (Festinger, 1954) are diminished.

In the game context however, badges seem to be a notable vessel for players’ goal-oriented and social behavior. For instance, along with the publication of the FPS-game, Battlefield 3, EA Games also published a web service solely for monitoring and comparing player activities and badges. On the Xbox game console, every game publisher is required to implement badges in their games. In addition, it has been found that games with badges receive better ratings (EEDAR, 2007). Therefore, it seems that the effectiveness of game elements depends upon the nature of the service in which they are used, as well as the intentions and use scenarios of the users. The reason why people use different services can greatly differ between services that are of a different nature (van der Heijden, 2004). Therefore, game mechanics that are mostly hedonic are likely to provide little to the usage considerations of utilitarian services. This suggests that the gamification of utilitarian services might not be efficient unless the service also adds some hedonic emphasis, for example by being gamified more consistently through perhaps narrative and
other game mechanics, or if the core activity within the service already resembles a game, which is the setting in which the use of badges seems to yield positive results.

In the field of game studies, there are two main perspectives by which games (and therefore gamification) may be defined: 1) systemic (see Deterding et al., 2011) and 2) experiential (see Huotari and Hamari, 2012). The first approach defines games based on what elements or mechanics their system has. Therefore, the addition of game mechanics would (according to such an approach), transform services into games. However, the experiential perspective gamification is in conflict with how we understand gameful experiences. Gamification often attempts to direct user or consumer decision making towards choices that are desirable to a third party. Games themselves however, attempt to do the opposite. Games create choice spaces that are separate from deeply consequential outcomes (Caillois, 1961). The enjoyment of games emerges from mastering autonomous decision making activity, regulated by free will (Avedon and Sutton-Smith, 1977, Ryan et al., 2006), rather than from the outcomes of that decision making. In the same vein, Huotari and Hamari, (2012) propose that gamification then refers to design that aims to bring about these ‘gameful experiences’. In a common gamification implementation however, the goals are strictly tied into the consequential utilitarian activities of the service, and this was also the case in our experiment. According to game theorists, this is a conflict which might negatively affect the general attitudes of users towards such an implementation. We suggest that further studies be undertaken which measure the attitudes of users towards artificially assigned badges which are awarded for demonstrating certain behaviors within the service.

This conflict is also connected to a further issue: If we accept that gamification is about creating “gameful experiences” as the name suggests and not just about directly increasing customer relationship metrics, then the successfulness of gamification should also reflect metrics that measure the user experience. Although, we could hypothetically find that gamification increased the retention of users and other usage activities, it would still be unknown whether users experienced gameful or playful experiences. Therefore, for further studies we suggest focusing on the experiential aspects of such
engagement, such as *perceived enjoyment* (van der Heijden, 2004), *flow* (Csikszentmihályi, 1990) and *playfulness* (Webster and Martocchio, 1992, Martocchio and Webster, 1992).

In summary, we propose the following lines of enquiry for the future research on badges and gamification:

1) How does the nature of the underlying service (*utilitarian* versus *hedonic* - Hirschman and Holbrook, 1982, van der Heijden, 2004) affect
   a. the *goal commitment* (Locke and Latham, 1990, Klein et al., 1999) towards badges
   b. the *attitude* towards gamification

2) How does the *involvement* (*cognitive* versus *affective* – Zaichkowsy, 1994) of the user or consumer affect
   a. the *goal commitment*
   b. the *attitude* towards gamification

3) How do the temporal differences in the implementation/removal of gamification affect the *continuous use intention* (Bhattacherjee, 2001).

4) How does the level of *goal commitment* (Locke and Latham, 1990) towards badges affect the *continuous use intention* (Bhattacherjee, 2001).

5) Does the typical implementation of gamification mechanics such as badges, arouse experiences related to gamefulness and playfulness (Caillois, 1961, Avedon and Sutton-Smith, 1977, Csikszentmihályi, 1990, Ryan et al., 2006, Webster and Martocchio, 1992, Martocchio and Webster, 1992) and further promote hedonic use (van der Heijden, 2004).

**Acknowledgements**

I would like to thank the editors and reviewers for their useful comments. I would also like to thank Sharetribe.com (Antti Virolainen and Juho Makkonen), Malte Elson, Veikko Eranti, Matti Nelimarkka, Juha Tolvanen and Antti Ukkonen.

This study was supported by Finnish Cultural Foundation.
7. References


---

1 The previous name of Sharetribe was Kassi. Sharetribe was given its current name in May 2012. We use the current name in the paper to make it easier for readers to find the service.