Electronic Commerce Research and Applications 12 (2013) 236–245

Contents lists available at SciVerse ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/ecra

Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service

Juho Hamari *

Information Systems Science, Department of Information and Service Economy, Aalto School of Business, P.O. Box 21220, 00076 Aalto, Finland

ARTICLE INFO

Article history:
Available online 16 February 2013

Keywords:
Badges
Collaborative consumption
Game design
Gamification
Service marketing
Social commerce

ABSTRACT

During recent years, the addition of game mechanisms 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, which gamifies a utilitarian peer-to-peer trading service by implementing the game mechanism of *badges* that users can earn from a variety of tasks. There were 3234 users who were randomly assigned to treatment groups and subjected to different versions of the badge system in a 2 × 2 design. The results show that the mere implementation of gamification mechanisms 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 increased user activity.

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 *mechanism* (also called *mechanics* in the literature) has involved 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 mechanism should also prove to be more valuable and engaging. However, there is a dearth of empirical studies investigating the effects that result from gamifying services, in regard to customer and user behavior.

This research studies 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.

This article proceeds as follows. In Section 2, we discuss gamification and badges. Section 3 outline possible theoretical foundations that may explain the effects of badges and gamification, and proposes hypotheses for study. In Section 4, describes the research process, data collection, and the service in which the experiment was conducted. Section 5 outlines the results and Section 6 elaborates on these results, limitations and discusses possible reasons for the supported and unsupported hypotheses. Section 7 concludes the article by suggesting further research directions for studying gamification.

* Address: Information System Science, Department of Information and Service Economy, Aalto School of Business, P.O. Box 21220, 00076 Aalto, Finland. Tel.: +358 40 835 9563.
E-mail address: juho.hamari@aalto.fi

1567-4223/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.elerap.2013.01.004

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 or as a new way of exploiting customers, some regard it as a genuine way of enhancing the value of service. 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 mechanisms, 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 in 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) or as (2) 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. The mere addition of game elements does not necessarily guarantee successful gamification. However, in popular discussion the idea prevails that gamification simply refers to adding game mechanisms 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) toward the use of information systems and services. Therefore, gamification can be viewed as an attempt to convert utilitarian services into more hedonically oriented. (See, for example, van der Heijden (2004) on hedonic IS.) In terms of IS theory, this sits well within the long-run context of studying technology acceptance (Davis 1989, Venkatesh 1999), continuous usage intentions (Bhattacherjee 2011, 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 that most likely invokes extrinsic motivations. These, in turn, have been demonstrated to be detrimental to intrinsic motivations, autonomy and creativity (Deci et al. 1999). Game mechanisms 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 mechanism in popular gamified applications such as Foursquare. Game industry studies have found that the addition of badges to games has led to better critical reception and increased revenue (Electronic Entertainment Design and Research 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, Hamari and Eranti 2011, 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 that determine how the badge can be earned (Hamari and Eranti 2011). Furthermore, because of their visual element – especially 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 motivation (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). This article reports results of a field experiment which studies the effects of 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 of hypotheses between investigating whether the mere addition of game mechanisms (marked with "*a*") and active exposure (marked with "*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 forthcoming).

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 views 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 views 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 views 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 views 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), setting *goals* such as those in badges increases 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 mechanism* (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 mechanism (Salen and Zimmermann, Salen and Zimmerman 2004) and have been demonstrated to exert persuasive power even when the progression towards them was illusionary (Kivetz et al. 2006, Nunes and Drèze 2006). Clear goals are also one of the main dimensions of the flow theory (Csíkszentmihályi 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 may be offered clear goals as described above, they 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 views 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 views 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 views 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 views their own badges has a positive effect on the number of page views the user generates.

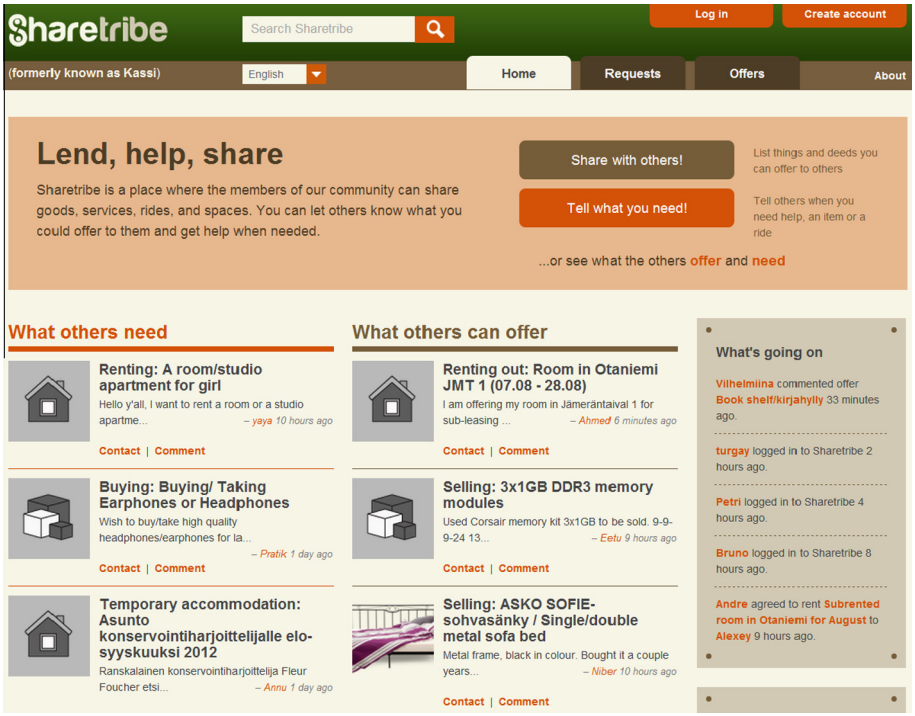

Fig. 1. Sharetribe homepage "Sharetribe is a network of "tribes", online communities where you can share goods, services, rides and spaces in a local, trusted environment. You can create a tribe for your university campus, your company, your neighborhood, your association, your sports club, your congregation, you name it!" – Sharetribe FAQ (2013).

4. Methods and data

4.1. Data collection and the case service

Sharetribe (www.sharetribe.com) is an international peer-to-peer trading service that offers its service package to a variety of organizations. See Figs. 1 and 2.

The available localizations as of early 2013 were English, Spanish, Finnish, Greek, French, Russian and Catalan. Sharetribe is used in communities all over the world, and there were 479 local Sharetribes worldwide. The company, Sharetribe Ltd., is a social for-profit enterprise registered in Finland. Its 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'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 buy and sell goods and services though. 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.

4.2. Field experiment

The field experiment was set up in the Sharetribe[1] service and data were 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. See Tables 1 and 2. Users who registered after the implementation were further randomly assigned to one of the groups.

The data consists of a database of 3234 users of the Sharetribe Aalto University site, who registered during the experiment timeframe. They include the number of trade proposals, accepted transactions, comments posted and how many individual page views a user undertook. Only users who had registered during the experiment timeframe were selected because older users had existing trade proposals in the service and would therefore have had accumulated actions during the experiment timeframe that would not have been affected by the experiment. Aalto University Sharetribe site was selected for the experiment because it is the largest implementation of Sharetribe of the several hundred installations worldwide.

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, using self-reported data could be avoided that might have potentially reflected novel and glorified attitudes toward the idea of using game mechanisms. With this approach, it is expected to achieve a higher level of validity. The generalizability of the findings is explored later in this article.

For the experiment, the badges were designed in adherence to previous work on conceptualizing the badge game design pattern (Hamari and Eranti 2011), as well as to resemble popular implementation approaches such as those 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 were able to 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

[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.

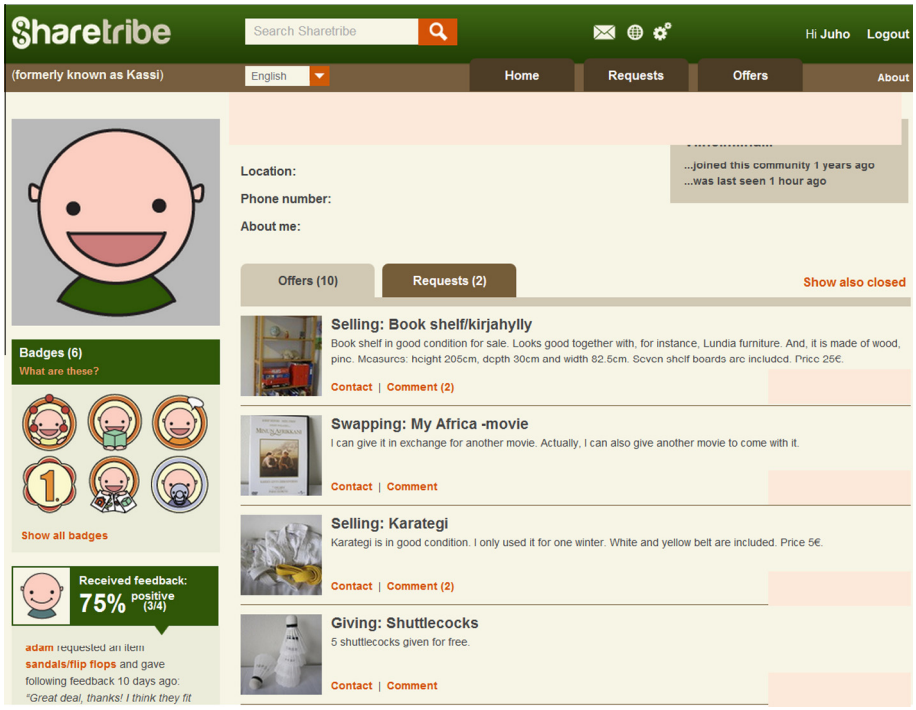

Fig. 2. User profile in Sharetribe.

Table 1
Experimental groups – independent variables.

Ability to view other users' badges (social comparison)	Ability to see from which actions one can unlock badges (clear goals)	
	No	Yes
No	Group 1	Group 3
Yes	Group 2	Group 4

Table 2
Users in treatment groups.

	Count	%
Group 1: Both features disabled (control)	805	24.9
Group 2: Social comparison condition	802	24.8
Group 3: Clear goals condition	790	24.4
Group 4: Both conditions enabled	837	25.9
Total	3234	100/100

for using the service for a prescribed number of consecutive days. The unlocked badges were displayed on the users' individual profiles, and 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 also were able to view badges on a separate page linked to every users' profile. See Fig. 3. This enabled them to 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.

5. Results

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

Table 3
The badge design.

Element/component	Implemented in Sharetribe
Signifier (name, visual, description)	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). 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* Example image of a badge (Commentator badge) in Sharetribe: 
Completion logic (trigger, pre-requirement, condition, multiplier)	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 in which she is 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.

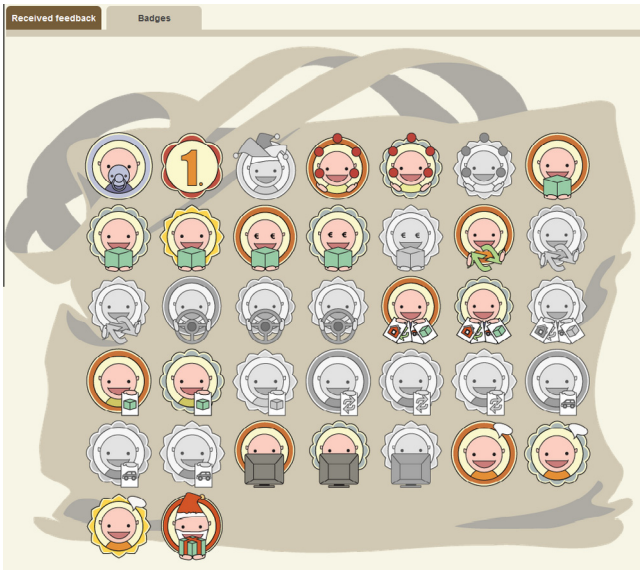

Fig. 3. View of the user's badges in Sharetribe.

A multivariate test (MANOVA) was performed on the effects of the possibilities to compare badges with other users, $F(4, 3227) = 1.679$, $p = 0.152$, 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, only users who had actively used the service after the implementation of badges were selected by selecting only those users who had at least 100 page views. This way, it could be ensured that all the users in the sample had 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, and therefore, unable to find evidence to support Hypotheses 1–8a. See Table 5.

We then tested whether support for Hypotheses 1–8b could be found 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 – H1–H4b) and the number views of the users own badge page (*clear goals* condition – H5–H8b). Multivariate testing (MANCOVA) on the effects derived from viewing other users' badges ($F(4, 3228) = 5.814$, $p = 0.000^{***}$, Wilk's = 0.993, $\eta^2 = 0.007$), viewing the users own badges ($F(4, 3228) = 565.361$, $p = 0.000^{***}$, Wilk's = 0.588, $\eta^2 = 0.412$) and their interaction ($F(4, 3228) = 58.324$, $p = 0.000^{***}$, 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 onto 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. See Table 6.

Based on these tests, it can be concluded 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.

Additionally, it was 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 any strong effects derived from comparing badges with other users could not be established. Instead, the fact that so few users had demonstrated any interest in the badges of other users implies even more essential challenges to gamification. 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 zero actions than users with one action, more users with one action than two actions and so forth. Therefore, the test were ran 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, with the following *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 ($\geqslant$100 page views). The *p*-values are – H1a: 0.084; H2a: 0.136; H3a: 0.568; H4a: 0.509; H5a 5a: 0.916; H6a 6a: 0.934; H7a: 0.882; and H8a 8a: 0.399.

Table 7 presents an overview of the results.

6. Limitations and 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 mecha-

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

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

Table 5
Test of Hypotheses 1–8a.

H#	Independent variable	Dependent variable	Result (3234 users registered during the experiment)			Results (723 active users with at least 100 page views)		
			F	p	η^2	F	p	η^2
1a	Comparison	Number of trade proposals	1.265	0.261	0.000	2.186	0.140	0.003
5a	Goal		0.695	0.405	0.000	0.166	0.683	0.000
	Comparison × Goal		0.022	0.882	0.000	0.333	0.564	0.000
2a	Comparison	Number of accepted transactions	1.131	0.288	0.000	0.790	0.374	0.001
6a	Goal		1.405	0.236	0.000	0.715	0.398	0.001
	Comparison × Goal		0.143	0.705	0.000	0.716	0.398	0.001
3a	Comparison	Number of comments	0.174	0.677	0.000	0.244	0.621	0.000
7a	Goal		0.110	0.741	0.000	0.015	0.901	0.000
	Comparison × Goal		0.248	0.619	0.000	0.769	0.381	0.001
4a	Comparison	Number of page views	0.200	0.655	0.000	0.052	0.820	0.000
8a	Goals		2.087	0.149	0.001	0.598	0.440	0.001
	Comparison × Goal		0.254	0.614	0.000	0.017	0.897	0.000

Table 6
Test of Hypotheses 1–8b.

H#	Independent variable - Views to	Dependent variable	Results		
			F	p	η^2
1b	Badge pages of others	Number of trade proposals	5.450	0.020**	0.002
5b	Own badge page		810.885	0.000***	0.201
2b	Badge pages of others	Number of accepted transactions	2.247	0.134	0.001
6b	Own badge page		1034.045	0.000***	0.242
3b	Badge pages of others	Number of comments	1.957	0.162	0.001
7b	Own badge page		720.280	0.000***	0.182
4b	Badge pages of others	Number of page views	1.398	0.239	0.000
8b	Own badge page		2253.084	0.000***	0.411

Table 7
Confirmation of hypotheses.

H#		Hypothesis	Supported
1a	*Social comparison: productive actions*	Users who are enabled to compare their badges with the badges of other users create more trade proposals.	No
1b	*Social comparison: productive actions*	The number of times a user views the badges of other users has a positive effect on the number of trade proposals the user makes.	Yes
2a	*Social comparison: quality of actions*	Users who are enabled to compare their badges with the badges of other users complete more transactions.	No
2b	*Social comparison: quality of actions*	The number of times a user views the badges of other users has a positive effect on the number of transactions the user completes.	No
3a	*Social comparison: social interaction*	Users who are enabled to compare their badges with the badges of other users post more comments.	No
3b	*Social comparison: social interaction*	The number of times a user views the badges of other users has a positive effect on the number of comments the user posts.	No
4a	*Social comparison: usage activity*	Users who are enabled to compare their badges with the badges of other users generate more page views.	No
4b	*Social comparison: usage activity*	The number of times a user views the badges of other users has a positive effect on the number of page views the user generates.	No
5a	*Clear goals: productive actions*	Users who are enabled to have clear goals through badges create more trade proposals.	No
5b	*Clear goals: productive actions*	The number of times a user views their own badges has a positive effect on the number of trade proposals the user makes.	Yes
6a	*Clear goals: quality of actions*	Users who are enabled to have clear goals through badges complete more transactions.	No
6b	*Clear goals: quality of actions*	The number of times a user views their own badges has a positive effect on the number of transactions the user completes.	Yes
7a	*Clear goals: social interaction*	Users who are enabled to have clear goals through badges post more comments.	No
7b	*Clear goals: social interaction*	The number of times a user views their own badges has a positive effect on the number of comments the user posts.	Yes
8a	*Clear goals: usage activity*	Users who are enabled to have clear goals through badges generate more page views.	No
8b	*Clear goals: usage activity*	The number of times a user views their own badges has a positive effect on the number of page views the user generates.	Yes

nisms. Several theories from motivational and social psychology were discussed as probable candidates to explaining the possible effects of gamification. Therefore, for further study, wesuggest measuring latent psychological variables through surveys in order to attain more accurate linkages between game mechanisms, psychological effects and resultant behavioral manifestations.

With regard to the present experiment, there is no way to infer directly whether the game mechanisms were able to arouse the

hypothesized psychological effects, such as social influence or goal commitment as the study investigated 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 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. Thus, this 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 (e.g., see Gartner (2012) which hints that 80% of gamified applications will fail), then the results might not be so surprising after all. Thus, 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 is possible that badges do not offer that much value to 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.

It was 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 were sufficiently large with regards to registered users, the number of times users have carried out different activities (dependent variables) on average is quite low. See 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-sampling analysis work further strengthened the results.

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, it was 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 with "*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 (e.g., Gartner 2012) in general can stem from the lack of interest toward such mechanisms when 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.

7. Conclusion

This article reported results of a one and a half year-long field experiment on gamifying a utilitarian trading service by the implementation of badges, which have been considered the primary mechanism through which services have been gamified. This study was able to 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, the results indicate that actively browsing other users' badges was positively associated with posting trade proposals in the service. However, 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 could not be found.

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) toward the badges, which was hypothesized to be a prerequisite 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. It can be hypothesized 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* (Zaichkowsy 1994). Therefore, 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*

Table 8
Means of the dependent variables in the data.

	3234 Users registered for experiment		723 Active users with > 100 page views	
	Mean	SD	Mean	SD
Trade proposals per user	1.03	3.24	3.77	6.05
Accepted transaction per user	0.47	1.51	1.87	2.72
Comments per user	0.65	2.45	2.59	4.60
Page views per user	103.47	259.35	399.80	434.99

intentions (Bhattacherjee 2011) 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 that have attracted a user population 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 into carry out pre-meditated 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 means 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 (Electronic Entertainment Design and Research 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 mechanisms 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 mechanisms, 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* (Deterding et al. 2011) and (2) *experiential* (Huotari and Hamari 2012). The first approach defines games based on what elements or mechanisms their system has. Therefore, the addition of game mechanisms would (according to such an approach) transform services into games. However, the systemic perspective to gamification is in conflict with how we understand gameful experiences.

Gamification often attempts to direct user or consumer decision-making toward 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 1971, Ryan et al. 2006), rather than from the outcomes of that decision-making. In the same vein, Huotari and Hamari (2012) proposed 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 this experiment. According to game theorists, this is a conflict that might negatively affect the general attitudes of users toward such an implementation. Therefore, we suggest that further studies be undertaken which measure the attitudes of users towards artificially assigned badges 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 (see Huotari and Hamari 2012). Although, we may 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* (Csíkszentmihályi 1990) and *playfulness* (Webster and Martocchio 1992, Martocchio and Webster 1992).

To conclude, this study proposes the following questions and lines of inquiry for 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 *goal commitment* (Locke and Latham 1990, Klein et al. 1999) toward badges, and *attitude* towards gamification? (2) How does the *involvement* – *cognitive* versus *affective* (Zaichkowsy 1994) – of the user or consumer affect *goal commitment* and *attitude* towards gamification? (3) How do the temporal differences in the implementation or removal of the gamification affect *continuous use intention* (Bhattacherjee 2011)? (4) How does the level of *goal commitment* (Locke and Latham 1990) toward badges affect *continuous use intention* (Bhattacherjee 2011)? And (5), does the typical implementation of gamification mechanisms, such as badges, arouse experiences related to gamefulness and playfulness (Caillois 1961) 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.

References

Ajzen, I. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 2, 1991, 179–211.
Avedon, E., and Sutton-Smith, B. *The Study of Games*. John Wiley & Sons, New York, NY, 1971.
Bandura, A. Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28, 2, 1993, 117–148.
Bhattacherjee, A. Understanding information systems continuance: an expectation–confirmation model. *MIS Quarterly*, 25, 3, 2011, 351–370.
Caillois, R. *Man, Play, and Games*. Free Press, New York, NY, 1961.
Cialdini, R. Harnessing the science of persuasion. *Harvard Business Review*, 79, 9, 2001, 72–79.

Cialdini, R. *Influence: Science and Practice*, fourth ed.. Allyn & Bacon, Needham Heights, MA, 2001.
Csíkszentmihályi, M. *Flow: The Psychology of Optimal Experience*. Harper and Row, New York, NY, 1990.
Davis, F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 3, 1989, 319–340.
Deci, E., and Ryan, R. *Intrinsic Motivation and Self-Determination in Human Behaviour*. Plenum, New York, NY, 1985.
Deci, E., Koestner, R., and Ryan, R. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 6, 1999, 627–668.
Deterding, S., Dixon, D., Khaled, R., and Nacke, L. From game design elements to gamefulness: defining gamification. *In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, Tampere, Finland, September 2011, ACM Press, New York, NY, 2011.
Electronic Entertainment Design and Research. EEDAR study shows more achievements in games leads to higher review scores. 2007. Available at www.eedar.com/News/Article.aspx?id=9.
Festinger, L. A theory of social comparison processes. *Human Relations*, 7, 2, 1954, 117–140.
Gartner. Gartner says by 2015, more than 50 percent of organizations that manage innovation processes will gamify those processes. Engham, UK, April 12, 2011. Available at www.gartner.com/it/page.jsp?id=1629214.
Gartner. 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. Available at www.gartner.com/it/page.jsp?id=2251015.
Goldstein, N., Cialdini, R., and Griskevicius, V. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35, 3, 2008, 472–482.
Hamari, J., and Koivisto, J. 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, forthcoming.
Hamari, J., and Lehdonvirta, V. Game design as marketing: how game mechanics create demand for virtual goods. *International Journal of Business Science and Applied Management*, 5, 1, 2010, 14–29.
Hamari, J., and Eranti, V. Framework for designing and evaluating game achievements. In *Proceedings of Digra 2011 Conference: Think Design Play*, Hilversum, Netherlands, September 14–17, 2011.
Hamari, J., and Järvinen, A. Building customer relationship through game mechanics in social games. In M. Cruz-Cunha, V. Carvalho, and P. Tavares (eds.), *Business, Technological and Social Dimensions of Computer Games: Multidisciplinary Developments*, IGI Global, Hershey, PA, 2011.
Hirschman, E., and Holbrook, M. B. Hedonic consumption: emerging concepts, methods and propositions. *Journal of Marketing*, 46, 1982, 92–101.
Hsieh, J., Rai, A., and Keil, M. Understanding digital inequality: comparing continued use behavioral models of the socio-economically advantaged and disadvantaged. *MIS Quarterly*, 32, 1, 2008, 97–126.
Huotari K., and Hamari, J. Defining gamification: a service marketing perspective. In *Proceedings of the 16th International Academic MindTrek Conference*, Tampere, Finland, October 3–5, 2012, New York: ACM Press, New York, NY, 2012.
Jakobsson, M. The achievement machine: understanding Xbox 360 achievements in gaming practices. *Game Studies*, 11, 2011, 1.
Kivetz, R., Urminsky, O., and Zheng, Y. The goal-gradient hypothesis resurrected: purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, 43, 2006, 35–58.
Klein, H., Wesson, M., Hollenbeck, J., and Alge, B. Goal Commitment and the goal setting process: conceptual clarification and empirical synthesis. *Journal of Applied Psychology*, 84, 1999, 885–896.
Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., Rashid, A., Resnick, P., and Kraut, R. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10, 2005, 4.
Locke, E., and Latham, G. *A Theory of Goal Setting and Task Performance*. Prentice-Hall, Englewood Cliffs, NJ, 1990.
Malone, T. Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 4, 1981, 333–369.
Martocchio, J., and Webster, J. Effects of feedback and cognitive playfulness on performance in microcomputer software training. *Personnel Psychology*, 45, 1992, 553–578.
McGonigal, J. *Reality is Broken: Why Games Make Us Better and How They Can Change the World*. Jonathan Cape, London, UK, 2011.
Montola, M., Nummenmaa, T., Lucerano, A., Boberg, M., and Korhonen, H. 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*, Tampere, Finland, 2009, ACM Press, New York, NY, 2009.
Nunes, J., and Drèze, X. The endowed progress effect: how artificial advancement increases effort. *Journal of Consumer Research*, 32, 4, 2006, 504–512.
Ryan, R., Rigby, C., and Przybylski, A. Motivational pull of video games: a self-determination theory approach. *Motivation and Emotion*, 30, 4, 2006, 344–360.
Salen, K., and Zimmerman, E. *Rules of Play. Game Design Fundamentals*. MIT Press, Cambridge, MA, 2004.
Sharetribe FAQ. 2013. Available at www.sharetribe.com/en/faq.
van der Heijden, H. User acceptance of hedonic information systems. *MIS Quarterly*, 28, 4, 2004, 695–704.
Van de Ven, N., Zeelenberg, M., and Pieters, R. The envy premium in product evaluation. *Journal of Consumer Research*, 37, 2011, 984–998.
Vargo, S., and Lusch, R. Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68, 1, 2004, 1–17.
Venkatesh, V. Creation of favorable user perceptions: exploring the role of intrinsic motivation. *MIS Quarterly*, 23, 2, 1999, 239–260.
Webster, J., and Martoccio, J. Microcomputer playfulness: development of a measure with workplace implications. *MIS Quarterly*, 16, 2, 1992, 201–226.
Zaichkowsy, J. The personal involvement inventory: reduction, revision, and application to advertising. *Journal of Advertising*, 23, 4, 1994, 59–70.
Zichermann, G., and Cunningham, C. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*, first ed.. O'Reilly Media, Sebastopol, CA, 2011.