CHI 2015 Workshop "Researching Gamification: Strategies, Opportunities, Challenges, Ethics" Collected Papers

Sebastian Deterding, Seth Cooper, Alessandro Canossa, Lennart E. Nacke, Casper Harteveld, Jennifer R. Whitson

Researching Gamification

Simon Scheider, Peter Kiefer, Paul Weiser, Martin Raubal and Christian Sailer
Score design for meaningful gamification

Noopur Singh, Hylke W. van Dijk, Adelita V. Ranchor and Robbert Sanderman
Family Support to Improve Physical Activity in COPD: Game as a Research Tool

Ayoung Suh
Measuring User Engagement in an Enterprise Gamified System

Rob van Roy and Bieke Zaman
Moving Beyond the Effectiveness of Gamification

The Effects of Adding Premise and Backstory to Psychological Tasks

Max V. Birk
University of Saskatchewan,
Department of Computer Science
Interaction Lab
max.birk@usask.ca

Benjamin Buttlar
University of Trier
Department of Psychology
s1bebutt@uni-trier.de

Regan L. Mandryk
University of Saskatchewan,
Department of Computer Science
Interaction Lab
regan@cs.usask.ca

Jason Bowey
University of Saskatchewan,
Department of Computer Science
Interaction Lab
jtb134@mail.usask.ca

Abstract
Psychological tasks are used for assessment, induction, and treatment in both research and therapeutic contexts. Adding game-elements (e.g., leaderboards, premise, or points) could be beneficial for participant motivation; however, it has been argued that the value of gamification could come at a cost to experience or reliability of the task. We replicated four psychological tasks and added premise and backstory to investigate the effect of gamification on task performance and player experience. Our results show that adding game elements has the potential to negatively influence both player experience and performance.

Author Keywords
gamification; psychology; performance task; cognition

Introduction
Psychological tasks are used in both research and applied psychology (e.g., educational, clinical, or industrial) for a variety of reasons, including: to measure performance, induce a psychological state, or assess and treat psychological issues. To allow for better standardization as well as cheaper deployment and analysis, the presentation of these tasks is often computer-based. As such, designers have considered leveraging the power of gamification to motivate participation in standard psychological tasks.

Participants
218 participants (46.3% female) with an average age of 32.79 (*SD=10.29*) participated. Participants received $6 compensation paid through the platform.

Platform
Amazon Mechanical Turk (MTurk), a platform that acts as a broker between parties offering a range of *Human Intelligence Tasks* (HITs) (e.g., marketing questionnaires, or research studies) and paid workers.

Player Experience
PENS: Measures *competence*, e.g. experiencing success and failure based on one's own skills, *autonomy*, e.g. accepting challenge under one's own volition, and *relatedness*, e.g. experiencing relations to others.

IMI: Measures *enjoyment*, e.g. "I enjoyed this game very much", *tension*, e.g. "I felt tense while playing the game", and *effort*, e.g. "I put a lot of effort into this game."

PANAS: Measures positive affect and negative affect.

Adding game-elements in a non-game context has been shown to be effective in a variety of settings (e.g., the ESP game [9], Foldit [5]). In the context of gamifying standard psychological tasks, research has shown that game elements can help to increase task accessibility [3], foster engagement with a system and motivate participation in a treatment over the long-term [6]. As intriguing as these benefits are, research has yet to show whether or not they come at a cost, such as decreased reliability or changes in performance.

Adding game-elements into computer-based psychological tasks for the purposes of assessment, induction, or treatment has several ethical and practical implications. First, a gamified task can deviate in terms of the user experience; however, the *reliability* of assessment or treatment needs to remain equal to or better than the un-gamified task. Second, gamified tasks can be embedded into games and could be used to assess or treat users without their *explicit consent*. Third, the *increased motivation* of players of a gamified task has implications for the interpretation of findings.

To begin to understand how task performance and player motivation differs in gamified and standard psychological tasks, we investigated the effects of *adding premise* to four psychological tasks – Go/No-Go, N-Back, ambiguous word interpretation, and facial feedback. We focused on adding premise as our gamification element as it would not change the nature of the standardized task through changing the mechanics, feedback, or reward structure. In addition, the use of exposition is one of the tools recommended for meaningful gamification of systems [7].

Our results show that adding game elements has the potential to negatively influence player experience and performance, demanding careful evaluation before being applied in research or therapeutic environments.

Methods

Procedure
After participants gave consent they played one of the games and filled out player experience measures (see sidebar) after each task.

Tasks with and without premise
The four different psychological tasks were presented to half of the participants as a task, and to the other half as a game with premise added. We kept all other elements of the task stable, and only added premise through graphical assets and a backstory (reasoning) presented prior to the task. We used a Zombie theme, which allowed us to present a premise for different scenarios, e.g., shooting, running away, selecting.

Go/No-Go Task: To test executive functioning we used a Go/No-Go Task [7]. The task presents a sequence of stimuli for 500ms. In the task, participants respond to circles, but not to squares; in the game condition, players shoot blond zombies, but not a mole with a yellow hat. Higher precision and accuracy indicate better functioning.

N-Back: For each stimulus in the N-back task [4], participants are asked to indicate if the stimulus is the

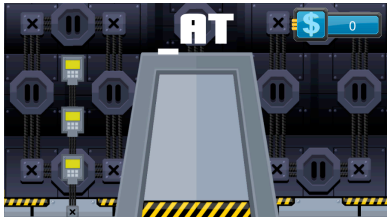

Figure 1 The gamified version of the ambiguous word task.

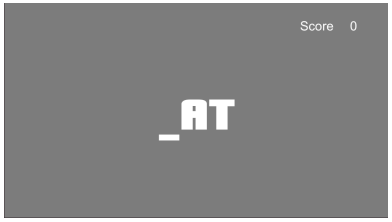

Figure 2 The standard version of the ambiguous word task

Figure 3 The gamified version of the N-back task

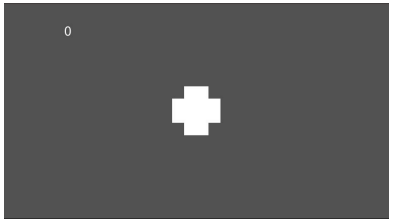

Figure 4 The task version of the N-back task

same as the one presented 2-back (press E), or not (press O). The stimuli were letters in the task version and zombies in the game. Higher precision and accuracy indicate better short-term memory.

Facial Recognition: In this task [2], a neutral face is presented and changes for 500ms to an emotion (happy, sad, angry, surprised), which participants must identify through a key press. The game was presented as a school for zombies to learn to recognize human emotion. The sum of correctly-identified expressions is a measure of emotion recognition.

Ambiguous Word Task: The ambiguous word task [1] presents words that can be interpreted as neutral or aggressive, e.g. S_AY (SLAY or STAY). In the game, participants were told to escape from a warehouse by guessing the password to open doors. Performance is the sum of hostile words chosen.

Results

Go/No-Go Task: We found no differences in player experience. However, the sensitivity was higher for the task *($F_{1,40}$=5.74, p<.021, η^2=0.126).*

N-Back: We found significant differences neither for player experience measures, nor for task performance.

Facial Recognition: Players experienced higher relatedness *($F_{1,59}$=4.74, p=.03, η^2=0.07)* in the plain task. Performance did not statistically differ.

Ambiguous Word Task: Players experience higher levels of enjoyment *($F_{1,68}$=5.62, p=.021, η^2=0.08)*, autonomy *($F_{1,68}$=17.02, p<.001 η^2=0.20)*, relatedness *($F_{1,68}$=9.80, p<.01, η^2=0.126)*, and immersion *($F_{1,68}$=8.05, p<.01 η^2=0.106)* in the plain task.

The ratio of hostile words to words overall differed *($F_{1,68}$=4.00, p<.049, η^2=0.056),* revealing that the game increases aggressive interpretation of words.

Interpretation

Our results show that the gamification of the N-Back task and the Facial Recognition task do not show differences in performance measures; however, gamification of the Go/No-Go task decreased performance in a strong effect. This difference is likely a result of the discrepancy of the stimuli – in the task, a circle is very different from a square, while the complexity of the more detailed characters in the game version need longer to be processed and result in lower sensitivity. Gamification of the ambiguous word task also reduced performance. One explanation is that this difference is due to the expectations participants bring with them into the experiment; within the magic circle, games can been experienced as a place where aggression is expected and players are given license to act in aggressive ways, potentially increasing their access to hostile words when in a game environment.

For player experience, we see no differences in the cognitive tasks; however, we find differences in the social and aggression tasks that point to more motivating and positive experiences in the task compared to the game. One explanation is that prior expectations of game scenarios influence players of the game – players expect a game to fulfill certain standards based on prior experience. For gamified tasks in general, the game mechanics and aesthetics may not match expectations, diminishing the experience. For experimental tasks, this is not true,

Figure 5 The game version of the executive functioning task

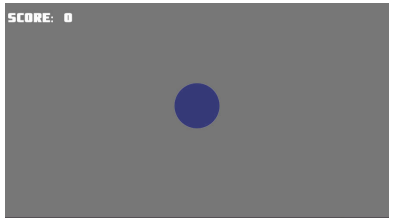

Figure 6 The task version of the executive functioning task

Figure 7 The game version of the facial recognition task

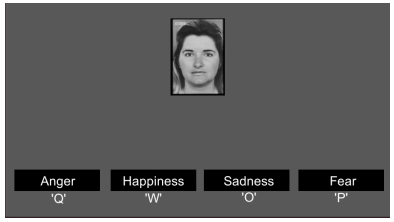

Figure 8 The task version of the facial recognition task

because we don't expect a task to stimulate fantasy or give interesting problems. In our case, the two affected conditions still felt like experimental tasks and less like a game (compared to the cognitive games), perhaps raising expectations for gameplay that were not met.

Conclusion

Opportunities: The promising take-away message from our results is that changes in experience don't necessarily affect performance. This opens up several design opportunities to create psychological tasks that can benefit from game-elements without threatening the reliability of the task. Psychological tasks can be used to assess different populations, e.g., children or older adults, by providing appeal. Because of the higher entertainment value, players may be more motivated to start and to complete multiple repetitions of a gamified task. This is often necessary, when for example, cognitive functions are being monitored.

Challenges: It might be more challenging than previously assumed to integrate game design into an already well-designed task. Considering that the mechanic is at the core of each game, gamifying the task through the addition of premise might not add as much value to simple tasks as we assumed. Additional research is necessary to understand how we can best integrate game-elements into psychological tasks to maximize benefits and minimize reliability issues.

Strategies: To create strategies that don't violate the reliability of the task, it is important to understand that adding game-elements might not foster a more positive experience, depending on the context and purpose of a task. Designers should ensure that the task is evaluated and compared to the gold standard.

Ethics: Not ensuring that a task is comparable to the gold standard has ethical implications in both research and therapeutic contexts because people are often classified based on their task performance. Additionally, embedding performance measures into a game-context has ethical implications, because games can be used for assessment without explicit participant consent.

References

[1] Anderson, C.A., Carnagey, N.L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. JPSP, 84, 960-971.

[2] Cohn, J. F., Schmidt, K., Gross, R., & Ekman, P. (2002). Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. In *Proc. of MI* (p. 491).

[3] Geurts, L., Vanden Abeele, V., Van Keer, K., & Isenborghs, R. (201). Playfully learning visual perspective taking skills with sifteo cubes. In *Proc. CHI play'14* (pp. 107-113).

[4] Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic bulletin & review*, *9*(4), 637-671.

[5] Khatib, F., DiMaio, F., Cooper, S.... & Foldit Void Crushers Group. (2011). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature structural & molecular biology*, *18*(10), 1175-1177.

[6] Mandryk, R. L., Dielschneider, S., Kalyn, M. R.... & Keiver, K. (2013). Games as neurofeedback training for children with FASD. In *Proc. IDC'13*, 165-172.

[7] Nicholson, S. (Forthcoming). A RECIPE for Meaningful Gamification. To be published in Wood, L & Reiners, T., eds. Gamification in Education and Business, New York: Springer

[8] Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia*, *46*(1), 224-232.

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

Mobile Gamification for Experiment Data Collection: Leveraging the Freemium Model

Kristen Dergousoff
Computer Science
University of Saskatchewan
Saskatoon, SK, S7N0Y7
kristin.dergousoff@usask.ca

Regan L. Mandryk
Computer Science
University of Saskatchewan
Saskatoon, SK, S7N0Y7
regan.mandryk@usask.ca

Abstract
Classic ways of gathering data on human behaviour are time-consuming, costly, and are subject to limited participant pools. Gamification provides a motivation to participate, but also requires the development of specialized, research-question specific games that can be costly to produce. Our solution leverages the popular Freemium model of play to motivate voluntary participation by rewarding players for participation in microexperiments with in-game powerups, using a robust framework to study multiple unrelated research questions within the same system. We deployed our game on the Android store and compared it to a gamified laboratory version and a non-gamified laboratory version, and found that players who used powerups were motivated to do the microexperiments.

Author Keywords
Gamification, freemium, psychophysics, crowdsourcing

Introduction
A fundamental, and limiting, step in Human Computer Interaction research is gathering data in order to understand human behaviour. Researchers often perform costly and time-consuming user studies in laboratory environments. In academic research this

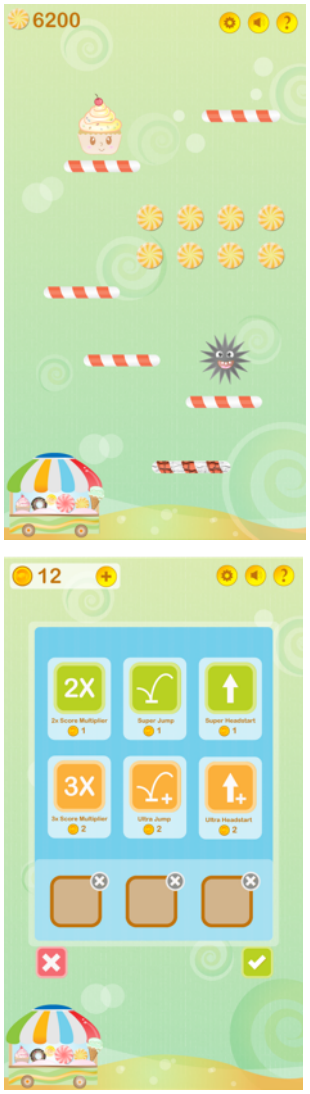

Figure 1. Top- Prototype of gameplay; Bottom- Prototype of powerup menu

usually consists of recruiting participants from a participant pool at the university, which confines participation in studies to the "boom-and-bust semester cycle" and limits the generalizability of the study to the demographic of young college students [9] [10].

One solution that has been proposed is to crowdsource experiments. Crowdsourcing platforms such as Amazon's Mechanical Turk (AMT) have been shown to successfully recreate experimental results [6], [7] and provide additional benefits such as constant any-time access to a large and diverse participant pool [9]; however, issues have been raised regarding data quality and the ethics of low pay [9].

The gamification [4] of experimental studies has been shown to motivate large numbers of voluntary participants [11] and to increase the enjoyability of tasks [5]. However, it is still unclear whether the results obtained through a gamified approach are as accurate as data gathered in traditional laboratory settings [1]. Also gamified experiments are usually highly customized to the specific research question, requiring costly development of new games for each research problem that is addressed.

Our solution is to combine the motivation of gamification with the broad participant base available through crowdsourcing by leveraging the benefits of the "Freemium" model, in which players can play a game for free, but are given access to special content, features, or advertisement-free play for a fee. We created a framework in which players of a game gain in-game advantages (i.e., powerups) for completing experimental tasks (i.e., microexperiments) [3]. We target the mobile game market to target players looking to kill time with short gameplay sessions. The main advantage of our framework is that it separates the system for completing microexperiments from the game so that multiple experimental tasks can be deployed in a single game, or multiple games can be deployed to increase the appeal for players. To evaluate the efficacy of our framework, we developed one game and two experimental tasks.

System

Our game, Sugar Rush, is an action-based vertical platformer that utilizes the accelerometer on an Android phone. In the game, players are in control of a continuously bouncing cupcake, which they guide through platforms and enemies to collect candy and coins and obtain a high score (see **Figure 1**). Players are provided with the option to purchase useful powerups with in-game currency earned through the completion of quick microexperiment tasks.

Powerup items can be accessed from the main menu (see **Figure 1**) and "purchased" by players using in-game credits. These powerups provide players with a benefit to gameplay, such as increasing the value of each candy collected, or blasting the player through the first sections of the game, collecting a large amount of coins along the way. Once powerups have been purchased, they are automatically applied to the next game session that the player starts.

Players can earn in-game credits to purchase powerups by completing the quick experimental tasks that are included in the task manager. When the player chooses to participate in a microexperiment, the system randomly chooses one of the available experimental tasks. Upon the completion of the task, the player is

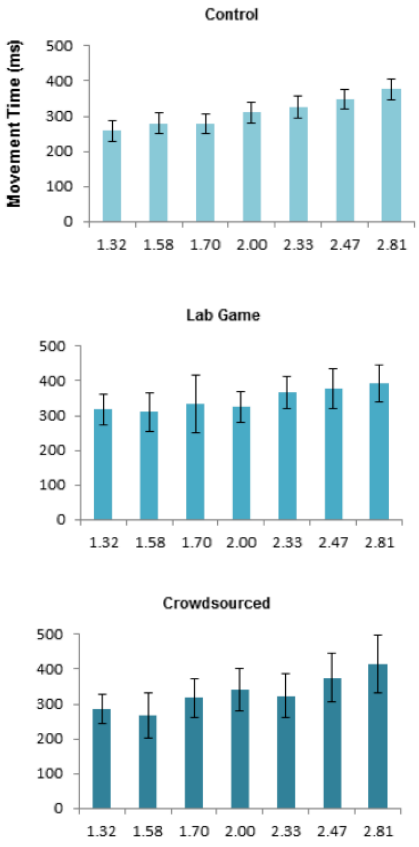

Figure 2. Mean movement time for the motor (Fitts) task for each index of difficulty across all three conditions.

awarded a single credit. The number of credits required to purchase a powerup varies between one and two credits and a player can purchase a total of three powerups at a time, providing incentives for the player to participate in multiple experimental tasks.

Evaluation

We evaluated our framework under three conditions to isolate the factors of gamification and crowdsourcing and examine their effects separately on the quality of data gathered (Crowdsourced Game, Laboratory Game, and Control, i.e., laboratory no game). We also analyzed usage data received from the Android market to inform the degree to which our prototype game and chosen Freemium model motivated voluntary participation in the tasks. Finally, we deployed a survey with the laboratory participants who played the game to gather their opinions on the topic.

We were interested in answering two main questions. First, whether the quality of the work that they are doing meets the standards set by similar experiments conducted using more traditional approaches, such as in the lab. Second, whether participants are motivated to play our game and whether they find enough value in the powerups to "work" for them by doing tasks.

Results and Discussion

Data Quality

Results of the evaluation of data quality indicated that there was no difference in performance between the three conditions for the performance-based motor task (i.e., a Fitts reciprocal tapping task [8]) – see **Figure 2**; however, the crowdsourced game resulted in worse performance on the attention-based cognitive task (i.e., a Cleveland and McGill information visualization judgment task [2]) – see **Figure 3**.

One possible explanation for this difference could be that participants are concerned with completing the task quickly in order to earn their credit. Because this concern of earning the credit quickly is in line with the performance-based nature of the motor task, it does not affect the quality of the results. However, because the cognitive task requires that participants take the time to make a judgment regarding the values they input, this goal of finishing as quickly as possible is detrimental to quality.

One way to motivate greater effort and accuracy by the participants in their responses is to make the rewards contingent on the apparent effort put into the task. If the data that is entered is easily identifiable as being false data, the credit reward can be withheld. Although this would only deter players from providing obviously false data, our results showed that when participants took time to enter a value, the value entered was not unreasonable. In addition, the value of the reward could be tied to the accuracy of the data, motivating participants to input quality data. Withholding rewards for obvious lack of effort or tying the value of the reward to the quality of the answer would encourage participants to spend more time in giving a response and likely result in better quality data.

Motivation to Participate

The survey results of player motivation suggested that players were willing to do the tasks in return for earned in-game bonuses, would prefer the tasks over in-game advertising, and that the tasks did not detract from the play experience. The usage data shows that participants who were exposed to the powerups used

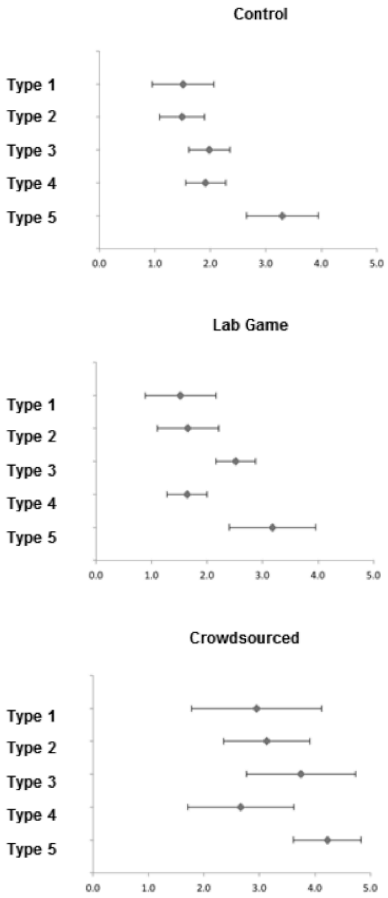

Figure 3. Mean log absolute error for the cognitive (McGill) task for each of five types of charts across all three conditions

them in about 21% of games, but that the majority of participants did not actually try the powerups.

We feel that it is possible to increase the motivation of players to complete experimental tasks in a framework such as ours through game design. Specifically, exposing the players to powerups in a limited manner would help players learn to appreciate the value of powerups (e.g., as is done in Candy Crush or Frozen's Free Fall) and likely increase the appeal. Hindering progress by unlocking content through microexperiment completion could also motivate increased participation.

Conclusions

Gathering data on human behaviour is a limiting factor in behavioural research. Crowdsourcing and gamification provide two possible solutions to this data collection problem but they each pose issues of their own. In this paper we presented our solution, which combines both gamification and crowdsourcing techniques into a smartphone-based platform to motivate voluntary participation and provide researchers with a framework that can be used to investigate multiple research questions without the need to develop costly specialized games.

Results from our initial evaluation showed that the quality of the motor task data did not suffer; however, the data from the cognitive task was of lower quality. We feel that tying the reward to the quality of the data could improve data quality for attention-based tasks. Despite a low adoption of powerup usage, participants that were exposed to the experimental tasks were supportive of participating in return for in-game benefits.

References

[1] Cechanowicz, J., Gutwin, C., Brownell, B., and Goodfellow, L. Effects of Gamification on Participation and Data Quality in a Real-World Market Research Domain. *Gamification '13*, (2013) 58-65.

[2] Cleveland, W. S., and McGill, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Am. Statistical Association, 79*, (1984), 531-554.

[3] Dergousoff, K., and Mandryk, R. Mobile Gamification for Crowdsourcing Data Collection: Leveraging the Freemium Model. CHI 2015 (to appear).

[4] Deterding, S., Bjork, S., Nacke, L. E., Dixon, D., and Lawley, E. Designing Gamification: Creating Gameful and Playful Experiences. *CHI 2013*, ACM (2013), 3263-3266.

[5] Fatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., and Mandryk, R. L. Calibration Games: Making Calibration Tasks Enjoyable by Adding Motivating Game Elements. *UIST `11*, ACM (2011). 403-411.

[6] Heer, J., and Bostock, M. Crowdsourcing Graphical Perception: Using Mechanical Turk to Asses Visualization Design. *CHI 2010*, ACM (2010). 203-212.

[7] Komarov, S., Reinecke, K., and Gajos, K. Z. Crowdsourcing Performance Evaluations of User Interfaces. *CHI 2013,* ACM (2013).

[8] MacKenzie, S. I. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction, 7*, (1992). 91-139.

[9] Mason, W., and Suri, S. Conducting Behavioural Research on Amazon's Mechanical Turk. *Behavior Research Methods*, (2012). 1-23.

[10] Reinecke, K., and Gajos, K. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. *To appear in CSCW'15.*

[11] von Ahn, L., and Dabbish, L. Labeling Images with a Computer Game. *CHI '04*, ACM (2004). 319-326.

Design Considerations for Creating Game-Based Social Experiments

Casper Harteveld
Northeastern University
140 Meserve Hall
360 Huntington Ave
Boston, MA 02115
c.harteveld@neu.edu

Steven C. Sutherland
Northeastern University
567 Holmes Hall
360 Huntington Ave
Boston, MA 02115
st.sutherland@neu.edu

Gillian Smith
Northeastern University
147 Meserve Hall
360 Huntington Ave
Boston, MA 02115
gillian@ccs.neu.edu

CHI 2015, April 18-23, 2015, Seoul, Republic of Korea.
ACM 978-1-4503-3145-6/15/04.

Abstract
In this paper, we explore the topic of games as research methods, which has received little attention and has much promise. More specifically, our aim is to provide insight into what design considerations are made for creating game-based social experiments. In order to achieve this, we discuss our experiences of translating a traditional experiment focused on the decoy effect into a game-based experiment. Our preliminary results suggest that monetary incentives with no real-world consequences can affect behavior even when time delays with real-world consequences were a competing factor. This has implications for how research can be gamified.

Author Keywords
games as research method; social experiments; game design; citizen science; crowdsourcing.

ACM Classification Keywords
H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous; I.6.8. [Simulation and Modeling]: Types of Simulation—*Gaming*; K.4.m. [Computers and Society]: Miscellaneous

Introduction
Games have increasingly become a tool for researchers [1]; however, using *games as research method* has received far

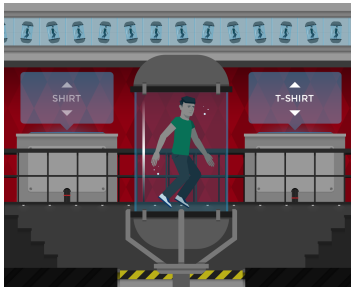

Figure 1: Choosing existing characters or creating new ones with the character creator tool.

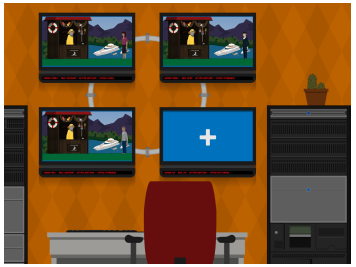

Figure 2: Manipulating the conditions for creating an experiment with the manipulator.

less attention in the past decade than their use for impact. The fundamental difference between both possible uses is that the former is used by the researcher to observe behavior, whereas with the latter the intent is to change behavior. With "games as research method," a game is used as the particular form or procedure for obtaining observations instead of surveys or laboratory experiments; on the other hand, with "games for impact," a game is used as an intervention to solve a problem or achieve objectives (e.g., education, persuasion) instead of lectures or workshops [4]. A useful way to distinguish between the two is that the focus for games for impact is on the person or system "to-be": the goal is to *change* behavior. For games as research method the focus is on the person or system "as-is": the goal is to study *existing* behavior.

Some scholars purposefully incorporate both uses [7]. However, the purpose of each use is significantly different and designers will need to make different considerations that will either emphasize the use as research method or for impact [5]. For example, for research, it may not be desirable to provide immediate feedback as this could influence subsequent behavior. In contrast, from designing a game for impact such behavior change may be necessary.

In this paper, we aim to dedicate more attention to the topic of games as research methods. A number of scholars have used gameful environments for research successfully [1] and a special issue has been dedicated to the subject too [2]; however, to our knowledge few have elaborated on the *design considerations* in gamifying research. It is our intention to provide more insight into such design considerations through our experiences of creating game-based social experiments in *Mad Science*, which is still in its development phase. Once the game is completed, players will be able to create their own experiments and to participate in other researchers' experiments. However, to begin testing impacts of our design decisions, several playable scenarios have been created to attempt to replicate previous research findings.

Although we developed a number of experiments using *Mad Science*, we will specifically focus on the (re-)development of one experiment. This experiment is a replication of an existing experiment based on the well-known phenomenon of the *decoy effect* [6]. Before we discuss this experiment in detail, we will describe the concept of *Mad Science* and how it distinguishes from related work on using games as research method. Our work contributes to HCI in thinking on how gamification can help support research activities, and specifically in thinking what it takes to translate typical laboratory experiments to a game-based variant.

Concept of Mad Science

With regards to the use as research method, games have predominantly been used to study physiological traits, such as eye-hand coordination and visual attention [3]. With the emergence of gamification, scholars are more recently considering how to apply game techniques to traditional research methods, such as surveys. Then, game-like environments have been used for years to study behavior by a few scholars but mostly also in physical laboratory settings [1]. The game *Mad Science* differs from this related work in that it is accessible on any mobile device or browser. It is further based on experimenting with social interactions, hence why we refer to the experiments as *social experiments*.

In terms of its concept, *Mad Science* is a digital game where players join the corporation Mad Science, Inc. as one of their new "mad" scientists—people who are

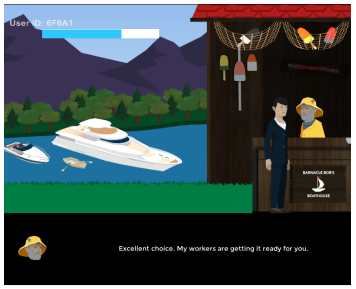

Figure 3: Decoy experiment with three boats for purchase. Below are the options offered to players.

1. Will cost $25,000.00 and will take 30 seconds to build.
2. Will cost $50,000.00 and will take 20 seconds to build.
3. (a) Will cost $30,000.00 and will take 31 seconds to build.
   (b) Will cost $55,000.00 and will take 21 seconds to build.

intrinsically curious and show that curiosity in every aspect of their behavior. Mad Science Inc.'s mission is to "understand why people do what they do." Players have to learn to use the corporation's proprietary machinery to study human behavior, such as a character creator, object creator, and a manipulator (i.e., for creating research conditions). Once players are familiar with the tools, they can go to Mad World, a world where players can design and perform research, participate in the research of other players, and share effective strategies.

Our long-term goal is to accomplish through Mad Science what we have coined "participatory crowdsourced research." This term refers to having large numbers of players authentically and collaboratively participate in creating and experiencing scientific research. For the game to be successful, a crucial requirement is that it teaches players how to perform research. Therefore, *Mad Science* is an example of a game that is both used for research as well as for impact.

Design Considerations for the Decoy Effect

One of the experiments we developed was with the aim to replicate the decoy effect, which is a decision bias that has been supported by prior studies [6]. The decoy effect describes the tendency of preferences between two options (e.g., two different cars) to be affected by a third, asymmetrically dominated option. Asymmetrical dominance occurs when the third option is better than one alternative but is clearly worse than the other alternative. For example, buyers will not buy a large popcorn of $7 if a small popcorn of $3 is offered; however, if a medium is offered for $5.50, then suddenly more people will buy the large. The medium serves in this example as the decoy.

To replicate this experiment seemed non-trivial. One design problem we considered is how to make players feel the consequences of their decisions. With the popcorn example participants were actually spending money. Therefore, in making decisions, they will consider the consequences of spending their money. We were concerned that this may not be replicated with fictitious currency because spending such money has no consequences on the players themselves. We decided to use real time delays in addition to fictitious currency as time delays have real-world consequences.

In the first iteration of the experiment, participants were advised that they would receive $100,000 when they arrived at Mad World (the island where they would be able to conduct their own research in the future). Currently, players are unable to access Mad World thus the $100,000 is completely fictitious currency. Players entered a boathouse and were required to interact with a Non-Player Character (NPC), a character controlled by no human player. Players were given three options for boats that could be used to get them to Mad World. Each boat had an associated cost (in-game currency) and an amount of time to build (real time delays). Players were required to wait for the entire time, associated with their choice, to elapse before progressing to the final scenario. The first two options were the same for every participant. The third option randomly varied between participants. The third option varied in order to change which of the other two options asymmetrically dominated it.

As part of a classroom exercise, 99 students participated in this experiment. It turned out that in both conditions the least expensive Option 1 was preferred [8]. Although we were not able to replicate the decoy effect, the results suggest that participant decisions were under the control of the fictitious money stimulus. This is an interesting

finding because the money has no real-world consequence but the time delay does have a real world consequence.

We considered that replication did not happen due to small differences in the amount of time required to build. Therefore, in our next iteration we multiplied the time delay by 10, increasing the delays to minimally 200 seconds. Interestingly enough, we observed in another classroom exercise with 59 students that players still preferred Option 1 in both conditions. However, players were more frustrated in playing this exercise. Other than that players are biased by the money stimulus, the increased time delay increased player frustration, which should be of concern to designers as player retention and engagement are important.

Conclusion

Blascovich and Bailenson [1] argue that, for the most part, individuals do behave similarly in both virtual and real environments, but that researchers must be cautious because situations may differ in ways not previously considered. Our experience and results can attest that assertion. Based on this work, further caution is necessitated for game-based environments because engagement is a factor that is added on top of other design considerations, especially if participation is voluntary. Although our preliminary work does not provide conclusive answers, it does suggest that monetary incentives with no real-world consequences can affect behavior even when time delays with real-world consequences are a competing factor. This may actually be an example of real world human behavior because it might just be that people have a harder time with evaluating time than money. Further research will consider other designs and provide additional insights into how we can create game-based social experiments.

Acknowledgements

We thank our students Nolan Manning, Yuyang Zhao, and Huichen Guan for realizing our vision. We further thank the College of Arts, Media & Design and Northeastern University for funding this project.

References

[1] Blascovich, J., and Bailenson, J. *Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution:*. HarperCollins, New York, NY, 2011.

[2] Calvillo-Gámez, E., Gow, J., and Cairns, P. Introduction to special issue: Video games as research instruments. *Entertainment Computing 2*, 1 (2011), 1–2.

[3] Green, C. S., and Bavelier, D. Action video game modifies visual selective attention. *Nature 423*, 6939 (2003), 534–537.

[4] Harteveld, C. *Triadic game design: Balancing reality, meaning and play*. Springer, London, UK, 2011.

[5] Harteveld, C. *Making sense of virtual risks: A quasi-experimental investigation into game-based training*. IOS Press, Amsterdam, the Netherlands, 2012.

[6] Huber, J., Payne, J. W., and Puto, C. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research 9*, 1 (June 1982), 90–98.

[7] Meijer, S. A. *The organisation of transactions: Studying supply networks using gaming simulation*. Wageningen Academic Publishers, Wageningen, the Netherlands, 2009.

[8] Sutherland, S. C., Harteveld, C., Smith, G., Schwartz, J., and Talgar, C. Exploring digital games as a research and educational platform for replicating experiments. In *NEDSI Conference* (2015).

BudgetWiser: Gamification Design Opportunities in the Government Budget Domain

Nam Wook Kim
Computer Science, Harvard University
Cambridge, MA 02139 USA
namwkim@seas.harvard.edu

Chang Won Lee
Computer Science, KAIST
Daejeon, Republic of Korea
chiyah@kaist.ac.kr

Jihee Kim
Business and Technology Management, KAIST
Daejeon, Republic of Korea
jiheekim@kaist.ac.kr

Eun-Young Ko
Mathematical Sciences, KAIST
Daejeon, Republic of Korea
key44@kaist.ac.kr

Juho Kim
MIT CSAIL
Cambridge, MA 02139 USA
juhokim@mit.edu

Jonghyuk Jung
Industrial Design, KAIST
Daejeon, Republic of Korea
heretor92@kaist.ac.kr

CHI'15, April 18–23, 2015, Seoul, Republic of Korea.

Abstract
In our research project, dubbed BudgetWiser, we focus on the government budget, one of the most important policy documents of a government. Our goal is to increase public interests in the budget and encourage public participation in the budgeting process by leveraging open government data. We envision interactive tools in which taxpayers' budget navigation and sensemaking activities lead to meaningful resources for future taxpayers. A primary design goal is incentivizing taxpayers – often with low interests and awareness of the budget – to engage in such activities. We believe gamification can provide a viable solution to our challenge. We share some design lessons from our exploration in the budget domain with two prototype systems: Factful and BudgetMap. In the workshop, we hope to discuss various ways to incorporate gamification into our systems to encourage public participation.

Author Keywords
Open government data; Crowdsourcing; Gamification; Government budget.

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction

With the recent E-Government movement in many countries and government organizations, there are opportunities to improve the public awareness of important government operations and to elicit public participation in discussions and decision making. For example, open access to a government budget helps the public understand and evaluate how a government spends taxpayers' money, which is fundamental to a democracy [2]. In our research project, dubbed BudgetWiser, we focus on the government budget, which is considered as the single most important policy document of a government [1]. Our goal is to increase public interests in the budget and encourage public participation in the budgeting process.

Despite recent efforts in opening government data, designing interactive channels for taxpayers to make sense of extensive and multi-faceted budget data remains an open challenge. Even though the budget proposals and plans are available online in many countries, the existing resources suffer from two main drawbacks: 1) they fail to reduce the complexity of the budget in their way of delivery, and 2) their static format cannot accurately reflect public interests that dynamically evolve over time. In addition, while taxpayers are capable of understanding complex issues and making informed decisions, government organizations lack suitable tools for leveraging the wisdom of crowds [6, 7]. We envision interactive tools in which taxpayers' budget navigation and sensemaking activities lead to meaningful resources for future taxpayers. A primary design goal is incentivizing taxpayers – often with low interests and awareness of the budget – to engage in such activities. We believe gamification can provide a viable solution to our challenge.

Gamification has been proven effective in motivating desired behaviors in non-game contexts, including education (Khan Academy), science (Foldit), and user-generated content (Quora) [3]. A few serious games have been introduced in the government budget domain, such as Next 10's Budget Challenge and The New York Times' Budget Puzzle, by asking users to solve budget deficit problems by layout out budget priorities. While these systems are more 'serious' games with less focus on entertainment, Budget Hero, created by American Public Media and the Woodrow Wilson Center, introduced more game-like components as its primary goal is to educate the public about the federal budget and the budget tradeoffs. While our research shares design goals with Budget Hero, we envision applying gamification for crowdsourced budget labeling and navigation enhancement. We are currently exploring various incentives for participating taxpayers.

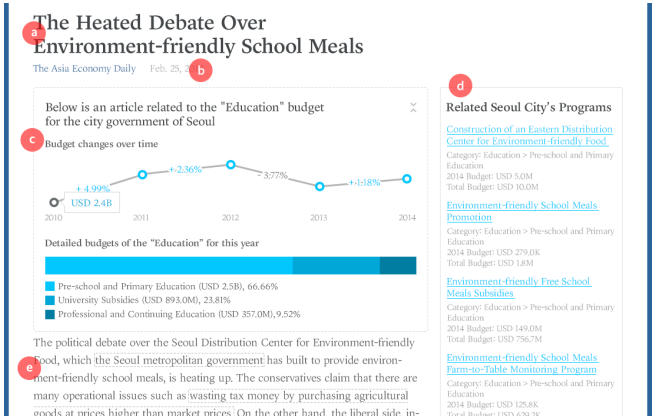

Figure 1: Overview of Factful powered by open budgetary data from the Seoul metropolitan government: (a) title of the article, (b) posted date and news outlet, (c) contextual budget category panel (education), (d) individual budget programs relevant to the article, (e) article content.

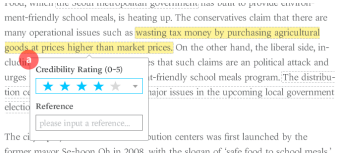

Figure 2: Annotative fact-checking interactions: clicking on the fact-check button opens a popup (a), so that a user can evaluate the credibility of the highlighted phrase in a 5-point scale, and add a URL reference that verifies the phrase.

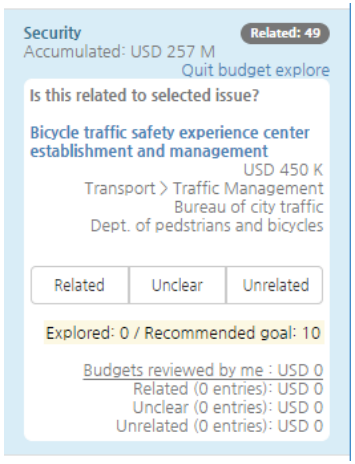

Figure 4: Passive tagging: a user is presented with a randomly chosen program and asked to determine the relationship with the selected issue.

Lessons from Factful and BudgetMap

We began our exploration in this space with two prototype systems: Factful [4](Figure 1) is an annotative news reading application powered by taxpayers' fact-checking and discussion activities; BudgetMap [5](Figure 3) is an interactive budget navigation tool powered by taxpayers' labeling data. We designed human computation tasks in both systems, in which natural fact-checking or labeling activities can yield useful information for future taxpayers.

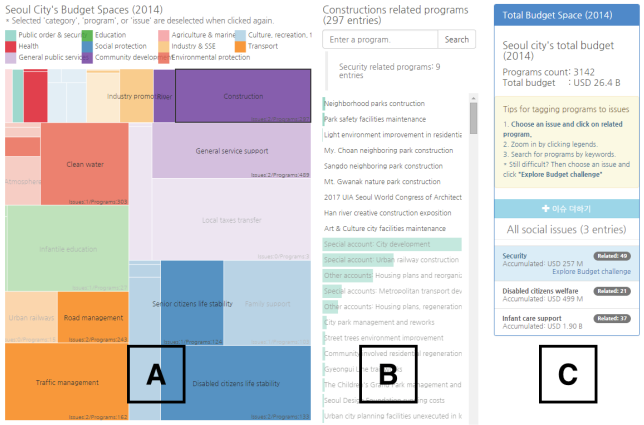

Figure 3: Overview of the BudgetMap Interface: (A) budget category information of Seoul city, (B) a list of programs sorted by the budget size, (C) a list of social issues

In our lab study, we found that the fact-checking support in Factful helped people hold more critical views. However, participants were sometimes reluctant to perform fact-checking (Figure 2), as it requires significant expertise and effort in searching for references, comparing multiple resources, and making the final judgment. We believe the barrier to contribution can be lowered, as there are bite-sized tasks and multiple ways even novice taxpayers can easily contribute.

BudgetMap (Figure 3) is an interactive tool for navigating budgets of government programs through a lens of social issues of public interests. It elicits the public to tag government programs with public-generated social issues via active and passive tagging methods. In our preliminary live deployment of BudgetMap in September 2014, many users commented that they appreciated the ability to view and navigate the policy programs and their budgets. This shows that presenting the budget data in a publicly accessible way provides value to the public. However, we discovered challenges in guiding the public to actively engage in tagging tasks (Figure 4) while making meaningful contributions. While taxpayers' reactions to tagging were overall positive, casual users on the web have not participated in the tagging as actively as we initially expected. We speculate that it may be due to the inherent complexity and difficulty in budgetary information. Overcoming these challenges will be crucial for BudgetMap to have broader social impact.

Gamification Opportunities

With the working prototypes and promising initial results from lab studies, we plan to deploy these systems to real taxpayers. We now describe some gamification ideas for encouraging continuous participation from taxpayers.

Decompose fact-checking activities into quiz-like microtasks: Inspired by multi-stage crowd workflows (e.g., three-stage in-video prompts for learners [8]), a fact-checking workflow can be presented as multiple microtasks: finding references, comparing multiple resources, scoring the found references, and reaching the final consensus. Each type of microtasks can be presented as quiz questions with some scores, and users can become

Nam Wook Kim
http://namwkim.org/
PhD student in the Harvard School of Engineering and Applied Sciences, studying information visualization and human-computer interaction. Prior to Harvard, he worked at Samsung and LG for four years. He received his M.S. from Stanford and B.S. (dual degree) from Ajou and Stony Brook.

Jihee Kim
http://jiheekim.net/
Assistant professor of Economics at KAIST. She holds a Ph.D. in Management Science and Engineering, a M.A. in Economics, and a B.S. in Computer Science. While her main research interest lies in economics, she's expanding her research agenda to designing civic engagement systems by leveraging open data.

Juho Kim
http://juhokim.com/
Ph.D. candidate at MIT CSAIL. He builds interactive systems powered by large-scale data from users. His research introduces learnersourcing, in which learners collectively generate novel content and interfaces for future learners while engaging in a meaningful learning experience themselves. He earned his M.S. in Computer Science from Stanford University, and B.S. in Computer Science and Engineering from Seoul National University.

an ''expert" in each fact-checking activity or subject area (e.g., childcare, transportation, environment) of their contribution scores.

Map exploration to exploit the map features of BudgetMap: Inspired by map-based arcade games whose goal is to claim all areas on the map, we plan to visualize parts of BudgetMap that a user has explored and encourage navigating to nearby (e.g., semantically related) budget items. For the user to be able to claim an area, certain amount of budgetary exploration and tagging can be required. Since many budget items span multiple areas, which was commonly found by participants in our lab study, we also plan to explore collaborative play: a team of taxpayers covering different areas on the map can join forces to discover overlapping items for bonus points.

Question & Answering game for budget tagging: In information literacy, question asking & answering plays a central role in demonstrating understanding of a text, which is essential for information retrieval to achieve higher-level comprehension. An idea is to incorporate the Q&A mechanism into a role-playing game where a user can choose to be a teacher or a student. Alternatively, the teacher can be an intelligent agent that automatically generates tagging questions.

In all these ideas, we plan to explore common game design elements such as badges, levels, points, and leaderboards. Moreover, we would like to address political bias and quality issues in designing interactive systems for civic participation. In the workshop, we hope to share design lessons from building Factful and BudgetMap, and discuss ways to incorporate gamification into our systems to encourage public participation.

References

[1] Oecd best practices for budget transparency. *OECD Journal on Budgeting* (2002).

[2] Baxandall, P., and Magnuson, B. Transparency. gov. 2.0: Using the internet for budget transparency to increase accountability, efficiency and taxpayer confidence. *Boston: MASSPIRG Educational Fund* (2008).

[3] Deterding, S. Gamification: designing for motivation. *interactions 19*, 4 (2012), 14–17.

[4] Kim, J., Ko, E.-Y., Jung, J., Lee, C. W., Kim, N. W., and Kim, J. Factful: Engaging taxpayers in the public discussion of a government budget. In *Proc. CHI 2015 (to appear)*, ACM (2015).

[5] Kim, N. W., Lee, C. W., Jung, J., Ko, E.-Y., Kim, J., and Kim, J. Budgetmap: Issue-driven navigation for a government budget. In *Proc. CHI 2015 EA (to appear)*, ACM (2015).

[6] Piotrowski, S. J., and Van Ryzin, G. G. Citizen attitudes toward transparency in local government. *The American Review of Public Administration 37*, 3 (2007), 306–323.

[7] Tanaka, S. Engaging the public in national budgeting: A non-governmental perspective. *OECD Journal on Budgeting 7*, 2 (2007), 139.

[8] Weir, S., Kim, J., Gajos, K. Z., and Miller, R. C. Learnersourcing subgoal labels for how-to videos. In *Proc. CSCW 2015*, ACM (2015).

Gamification Ethics: Exploitation and Manipulation

Author
Tae Wan Kim
Assistant Professor of Business Ethics
Tepper School of Business
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA
twkim@andrew.cmu.edu

Abstract

Two main accusations made against gamification are exploitation and manipulation [3, 4, 5]. I explain under what circumstances gamification can be exploitative or manipulative. I maintain that gamification is characteristically vulnerable to an expressive form of exploitation and a non-reason-tracking form of manipulation. I suggest solutions.[1]

Author Keywords

Gamification ethics, exploitation, manipulation.

Author Keywords

H.5.m [Information Interfaces and Presentation (e.g., HCI)]: Miscellaneous; K.4.1 [Computers and Society]: Ethics; K.8.0 [Personal Computing]: Games.

[1] This article is an abridged version of two unpublished working papers [13, 14]. I would like to thank the workshop organizers and two anonymous referees for their comments.

Problem: Exploitation

Notably, Bogost [4] dubbed gamification "exploitaionware." Bogost claims that players gain only a small fraction of the entirely created value, while the company gains relatively too large a share of it, and the imbalance is exploitative. Yet, just because an outcome is asymmetrically distributed between employees and the employer, does not itself constitute a wrong. Suppose that a transplant surgeon gains $35,000 by saving a patient. The patient benefits significantly more than the surgeon, but it would be odd to say that the patient exploits the surgeon. This does not mean, however, that gamification can never be exploitative. Below, I test gamification with two influential accounts of exploitation: the fairness account and the mere-means account.

According to the fairness account [21], an exploitative transaction is one in which *A* takes *unfair* advantage of *B* and a fair price is determined in a hypothetical environment in which some number of well-informed and unpressured buyers and sellers transact. Consider a famous exploitive case, *The Port of Caledonia and the Anna*, in which the master of a vessel in danger asked for assistance from a nearby tug and the master of the tug offered £1,000 or no help. The master of the vessel voluntarily agreed to pay £1,000. If there were at least one competitor, the offering price would be significantly lower than £1,000. This means that the tug master took an unfair advantage, so that the agreement was, although voluntary, exploitative.

Imagine a labor market in which labor for gamification is clearly specified within employment advertisements. In this market, some number of competing companies advertises, for instance, cashier jobs. The job descriptions of some companies include reference to the gamified working condition; other companies' descriptions do not. Would the addition of the gamified work change workers' salaries in the hypothetical market? Probably not, or, at least, it is unclear. In theory, offering a gamified working environment could allow a company to even slightly decrease wages for those who preferred gamified to non-gamified employment. From the fairness account, hence, it's difficult to say that gamification is exploitative.

The other influential view is the Kantian account of exploitation as the use of others as a mere means [1]. For Kant, what makes a person an end is the rational nature of humanity that enables her to deliberatively set moral and practical ends. First of all, respecting the rational nature demands that we refrain from coercing or deceiving people. As a conceptual matter, however, gamification is voluntary [19, 20], so players are not coerced.[2] And it makes sense that the intent of using gamification (or any other incentives, especially at the workplace) is straightforward: gamification providers want to solve some problem *x* by making a working environment game-like.

The Kantian account requires more. Treating others as ends in themselves also demands us not to be "indifferent" to minimally maintaining their rational nature of humanity [1]. Typically, this requires, in the context of organizational life, meeting minimum or reasonable safety standards and providing a minimum or living wage for employees [1]. Most companies that use gamification operate in the U.S. and other developed countries in which companies typically provide more than the legally defined minimum wage as well as other benefits including discounts on products, services, and merchandise.

The Kantian account can also be interpreted as an expressive theory, according to which an act that does not *substantively* (through e.g., deception, coercion, or minimum wage violation) use people as a mere means can *expressively* treat them as a mere means by showing disrespectful attitudes such as humiliation, contempt, offense, insult, etc. toward the rational nature of humanity [13]. For instance, treating adults as if they were children can often express the insulting attitude that the adults actually lack the rational capacity to act as adults [7]. Some cashiers who play Target's Checkout game, for instance, might reasonably believe that they are being treated as if they were in a preschool or kindergarten, in which once a four-year-old child behaves well and solves given problems, she earns a sticker or sometimes a candy or a marshmallow; the child is thrilled and wants to earn more stamps.

Solution: A Duty of Hermeneutics

In order not to express disrespect to others, we should consider how others would publicly *interpret* our acts. Hence, gamification designers have a duty to pay attention to the hermeneutical aspect of game elements. How can designers pay enough attention to the hermeneutical dimension? They can invite the input

[2] It is controversial whether or not gamification by definition is voluntary. I assume that it is. For a non-voluntary perspective, see [15].

of future interpreters, e.g., the Target cashiers, ask them how they interpret the game, and carefully consider their feedback. Perhaps a more structured way of fulfilling the duty of hermeneutics is possible. For instance, designers can develop an interpretive process based on the philosophy of Habermas's communicative approach [18], also suggested to courtroom judges who face interpretive issues [12]. No doubt, such a deliberative process cannot aim to lead involved parties to an ideally convergent interpretation. But it can at least promise a realistic and reasonably acceptable framework by which all parties can understand each other's perspectives and determine a realistic exit rather than stagnating in a dead-end interpretive debate. In addition, recognizing others' deliberative abilities is itself often valuable as an apt expression of respecting the rational nature of their humanity [17].

Problem: Manipulation

Relying on a philosophical analysis of bullshit [8], Bogost [3] claims that providers of gamification are bullshitters because they are indifferent to the justificatory quality—i.e., the truth value—of gamification. Here, Bogost, probably unknowingly, uses a reason-tracking account of manipulation [9], according to which gamification providers are manipulative to the extent that they do not track the justificatory quality of their means of influence. In typical cases, however, gamification providers seem to be committed to solving their problems using gamification. In other words, if it turns out that the gamified incentive does not work well—i.e., not true—, they find a new solution. That is, they track justificatory quality. So, Bogost's own charge is not typically true.

I develop a new perspective from Bogost's insight about bullshit. Providers who are not themselves bullshitters can attempt to influence players to be bullshitters—those who do not track the justificatory quality of their plays. Consider a well-known gamified activity, the ALS ice bucket challenge. Suppose that Alan genuinely cares about ALS patients and hopes more people become aware of the disease. He knows that his friends Ken and Taylor like to be recognized by others on Facebook. So Alan takes the ice bucket challenge himself on his Facebook and then nominates Ken and Taylor. During preparation for the challenge, Ken and Taylor realize that some filmed challenges are "liked" and "shared" many times—that is, earned points and badges, while others are simply ignored. So they decide to film their pouring ice bucket with a certain funny idea. They mention ALS in passing, but do not mean it. Many "like" and "share" their video. Ken and Taylor are excited about the points and badges. Many who would not otherwise know about ALS are now aware of ALS. Alan is happy about the outcome. So, everyone becomes happier.

Participating in the ice bucket challenge is itself a desirable act, and the worthiness of an action in part depends on the desirability of the act itself. Nonetheless, two actions that are equal in moral desirability may be of different moral worth, because the worthiness of an action also significantly depends upon the extent to which one is motivated to perform it with reasons that make it desirable [2, 14, 15]. Consider an often-invoked example: that a man wants to save a drowning child because he believes it is the right thing to do. The same man now wants to save the drowning child because he wants to seduce her mom. In both cases, the acts themselves are exactly the

same. But not many of us would deny that the moral worth of the act in the first case significantly differs from that of the inadequately motivated act in the second case.

I submit that Alan attempts to get Ken and Taylor's decision-making to fall short of a certain important moral ideal through a mental process I call "bullshitification,"[3] which I define as a decision-making process in which, due to the influence of game design elements, a decision maker becomes detached from the reason that makes her action desirable, which can put the action at risk of significantly losing its moral worth. For this reason, Alan manipulates Ken and Taylor. To put it more generically, Person *A*'s act *x* manipulates person *B* when A, through *x*, attempts to get *B*'s decision-making to fall short of the moral ideal about moral worth and a relevant norm, "Do the desirable act with the motivation that makes the act desirable."

I maintain that a similar moral phenomenon can be attempted in many other cases of gamification. That is, providers of gamification often attempt to get workers' or customers' decision-making to fall short of the ideal about moral worth and to violate a relevant norm, "Do the right/good/desirable act with the motivation that makes it right/good/desirable." In other words, the aim of gamification is to lead players to get motivated with points, badges, and leaderboards, and so to become indifferent to the reason that makes their action desirable, which can put their action at ethical risk of significantly losing the points of its moral worth.

[3] My notion of bullshitification differs from the so-called "motivation crowding out" [9], which says that extrinsic (especially monetary) incentives decrease intrinsic motivation. "One is said to be intrinsically motivated to perform an activity when one receives no apparent reward except the activity itself"[6: 105]. Hence, a gamification player is intrinsically motivated when he or she receives no apparent reward except the game experience itself. Nevertheless, the intrinsically motivated player can be induced to be indifferent to the reason that makes the act morally desirable.

Solution: Solemn Time

The problem with bullshitification is that players are induced to be motivated with a reason that does not make their action desirable, which in turn leads them to be indifferent to the right motivation that could make their job morally (more) worthy. Thus, any adequate solution must be something that can help players not to be indifferent to the right motivating reason. My suggestion is that players—that is, workers or customers—in gamified environments have what I want to call "Solemn Time," in which they learn about what their works or jobs are really doing, such as helping others, contributing to society, or enhancing important moral goods such as friendship or sustainability. For instance, the IT technicians who are excited about the points and badges in *OmniQuest* could have a chance during Solemn Time to get in touch with the reality of what they are doing through playing the game: helping senior patients in nursing homes.

One might worry that Solemn Time could distract workers from being immersed in the game and decrease the practical effectiveness of gamification. This idea about the incompatibility of fun with solemnity is not necessarily true, however. Endorsing Solemn Time can even boost the effectiveness of gamification as well as making it morally enlightened. In a field experiment about stressed out fundraising callers' performances [11], fundraising callers who were read stories about how their job could make a positive

difference in others' lives (e.g., a story about a person who would not have finished his college without the scholarship that fundraising callers created) showed significantly higher job performance than those who did not have such an intervention or Solemn Time. Robust empirical research is required to support this idea.

References

[1] Arnold, D. G., and Bowie, N. Sweatshops and respect for persons. *Business Ethics Quarterly* 13, 2 (2003), 221-242.

[2] Arpaly, N. Moral worth. *Journal of Philosophy* 99 (2002), 223-245.

[3] Bogost, I. *Gamification is bullshit.* August 9, (2011b.) http://www.theatlantic.com/technology/archive/2011/08/gamification-is-bullshit/243338

[4] Bogost. I. *Persuasive games: Exploitaionware.* May 3, (2011a). http://www.gamasutra.com/view/feature/6366/persuasive_games_exploitationware.php.

[5] Bréville, B., and Rimbert, P. *Losing on points: Do you play games, or are they playing you?* January (2014). http://mondediplo.com/2014/01/09videogames

[6] Deci, E. L. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 1 (1971), 105-115.

[7] Fahmy, M. S. Love, respect, and interfering with others. *Pacific Philosophical Quarterly* 92 (2011), 174-192.

[8] Frankfurt, H. G. *On bullshit.* Princeton, NJ: Princeton University Press, (2005).

[9] Frey, B., and Jegen, R. Motivation crowding theory. *Journal of Economic Surveys* 15, 5 (2001), 589-611.

[10] Gorin, M. Towards a theory of interpersonal manipulation. In *Manipulation: Theory and Practice*, by C. Coons and M. Weber, 73-97. New York: Oxford University Press, (2014).

[11] Grant, A. M. The significance of task significance: Job performance effects, relational mechanisms, and boundary conditions. *Journal of Applied Psychology* 93, 1 (2008), 108-124.

[12] Hellman, D. The expressive dimension of equal protection." *Minnesota Law Review* 85, 1 (2000), 1-70.

[13] Kim, T.W. Gamification and exploitation. Unpublished working paper.

[14] Kim, T.W. Gamification and manipulation. Unpublished working paper.

[15] Markovits. Acting for the right reasons. *Philosophical Review* 119 (2010), 201-242.

[16] Mollick, E., and Rothbard, N. Mandatory fun: Consent, gamification and the impact of games at work. Unpublished working paper http://www.ivey.uwo.ca/cmsmedia/1332866/rothbard-03-13-2015.pdf

[17] Scanlon, T. M. *What we owe to each other.* Cambridge, MA: Harvard University Press, (1998).

[18] Scherer, A. G., and Palazzo, G. Toward a political conception of corporate responsibility: Business and society seen from a Habermasian perspective. *Academy of Management Review* 32, 4 (2007), 1096-1120.

[19] Werbach, K., and Hunter, D. *For the Win: How Game Thinking Can Revolutionize Your Business*. Philadelphia: Wharton Press (2012).

[20] Werbach, K. (Re)defining gamification: A process approach. *Persuasive Technology: Lecture Notes in Computer Science* 8462 (2014), 266-272.

[21] Wertheimer, A. *Exploitation.* Princeton, NJ: Princeton University Press, (1996).

BlobSnake: Gamification of Feature Selection for Human Activity Recognition

Reuben Kirkham
Digital Interaction Group.
Newcastle University.
r.kirkham@newcastle.ac.uk

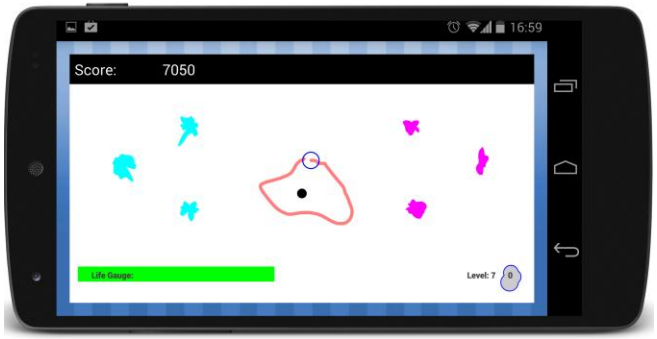

Figure 1: Illustration of BlobSnake. The goal of the game is to draw shapes that more effectively fit one side (representing one class of data) better than the other (whch represents different class or activity). The game incorporates a metric which is very similar to a snake fitting process from computer vision. The design is somewhat inspired by previous work that involved effectively visualizing sensor data [3]

Abstract
This paper discusses nascent work at Newcastle University's Digital Interaction Group, focused upon gameifying feature selection for Human Activity Recognition (HAR). The goals are two fold; the first is to mitigate the current need for a HAR expert to develop a feature selection for novel activity recognition problems, whilst the second is to address the need for science communication of this domain, especially in the legal setting. The initial game that has been developed – BlobSnake – is also briefly presented.

Author Keywords
Evidence, Gamification, Human Activity Recognition, Machine Learning, Science Communication.

Introduction
Human Activity Recognition (HAR) is an emerging field, which often involves the use of wearable sensors in order to detect specific activity patterns in the wild. Each system is hand developed to take a sensor stream – most likely from one or more inertial sensors (accelerometers or gyroscopes) and then points (or windows) in time are classified by a decision rule that partitions them into one of the particular activities under study, or a null (none of the above) category.

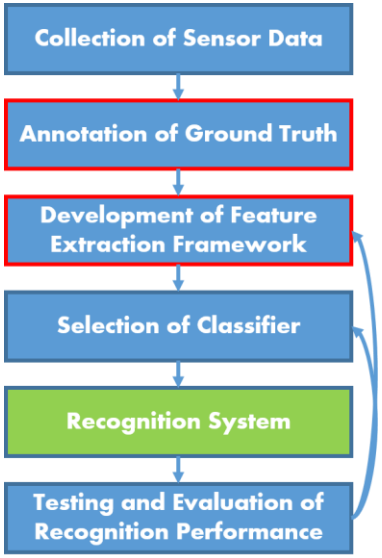

Figure 2: Illustration of the approach that is normally taken towards developing HAR systems. It can be seen that this involves a large iterative attempts at designing a system. In red outline, the two most challenging and burdensome aspects of this process are highlighted; these represent barriers to the wider deployment of HAR approaches.

Developing such a system is usually a non-trivial task. It involves the collection and precise temporal annotation of sensor data, followed by the use of an expert to develop a carefully selected set of features to use in a machine learning system (See **Figure 2**). Whilst it is true that standard features, such as the ECDF [2] can be used, it is usual that additional features should be added in order to improve the performance to a more reasonable level. The existing limitations are serious barriers to the wider adoption of HAR systems, and in particular have the effect of helping to restrict them to certain pluralities of typical (and usually, for instance, non-disabled) people that a system was originally trained upon. Reducing the amount of expertise required to develop a system based upon annotated training data would be of great assistance in broadening the reach of HAR systems. This is a primary goal of this work, by reducing the task of bespoke additional feature selection into being a matter of being a game that anyone can play, HAR can begin to become more inclusive as an endeavor. At the same time, there is a serious deficit in science communication, in that there is no existing work that attempts to explore HAR from a public engagement perspective. We attempt to fill this gap, noting that this application is most important in a legal context. This submission also explains our approach towards gamification of feature selection.

Science Communication, HAR, and the Law

If a HAR system could be used effectively in a legal context, then there would be profound benefits. Indeed, with the recent case involving FitBit, it appears that the legal community are on the verge of attempting to realize this potential opportunity. This author considers that such advances – if realized -are most likely to have an impact upon civil cases, particularly in respect of 'mental capacity', where someone's liberty is at stake, but to a civil standard (50.1%) rather than the criminal beyond reasonable doubt test. For example, in the case of *Neary vs Hillingdon Borough Council*, a dispute about the behavior of the plantiff's behavior, led to an extended deprivation of their own liberty; an effective recognition system on the line of Ploetz et al [5] might well have provided the court with sufficient evidence to address this, as well as empowering parties, such as a carer or supporter to more effectively oppose existing authorities and decision makers.

The current problem is that HAR systems have no real framework for which they could be used in a court context. The root of this problem lies with the fact that most judges are mathematically illiterate [4], with past judicial practice including accepting a p value of >0.5 as sufficient proof, before eventually switching to values more widely used in the scientific community; other cases have focused upon using actuarial calculations in order to reduce compensation payouts based upon gender and ethnicity. HAR represents a novel means of potentially providing evidence, and is therefore highly likely to encounter similar issues; this is especially of concern given the lack of agreed metrics for assessing a systems performance [6].

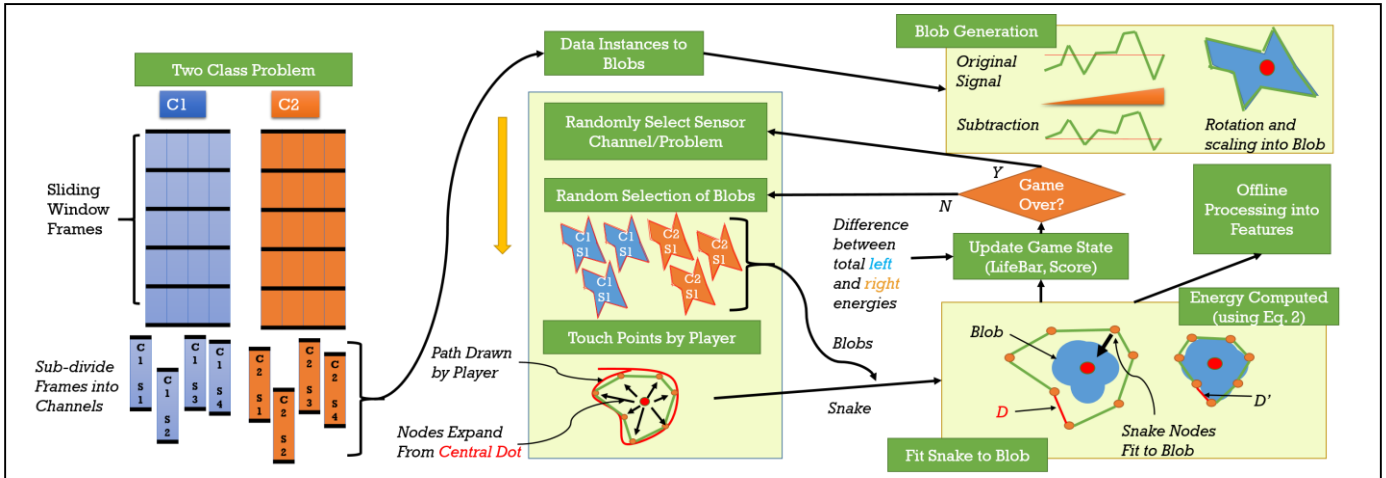

Figure 3: The algorithms involves in BlobSnake. On the left, the current two class machine learning problem that we will be using to explore the system in the context of an initial evaluation – although it is hoped that BlobSnake will scale to multi-class problems via a 1 vs all approach commonly used already in the machine learning

Skillful science communication – such as through BlobSnake – of the complexities and limitations of HAR to lay people, including lawyers and judges, is an essential step towards realizing the fairness of HAR usage in any evidential context. It is hoped that this work can be one initial step towards this far reaching goal, as well as assisting in furthering the debate around developing appropriate metrics for HAR systems that can be used in the legal setting.

BlobSnake

BlobSnake (**Figure 1**) is an Android game under development at Newcastle University. It comprises three components:

1. A script that deterministically maps existing sensor data into blobs that can be interacted within the game.
2. The game itself, where the players perform a feature generation task.
3. A simple algorithm that selects the most promising generated features for use in a system, in addition to standard features.

This workshop submission – for reasons of space – focusses upon briefly explaining the game itself (i.e. 1 and 2) We presume a sliding window approach, where sensor streams are chopped into fixed windows of N seconds (usually N =1), this is the most common approach in Human Activity Recognition [1]. The exercise of feature selection is a matter of finding functions that summarize data effectively as to aid a classifier in developing an effective decision rule.

Equation 1:

$$SF = \begin{cases} 1/|\Delta F| & \text{if} |\Delta F| \leq 2 \\ \log(1/|\Delta F|) & \text{otherwise} \end{cases}$$

$$\Delta F = |F_{min} - F_{max}|)$$

Equation 2:

$$E = \min_{S_r} \sum_{seg=1}^{n} abs\left(\frac{L_{seg}}{\sum L} - \frac{L'_{seg}}{\sum L'}\right)$$

Figure 4: Equations references in the text. Equation 1 is the scale factor used when translating data into blobs. Equation 2 is the energy function applied to a given snake with respect to each blob. Note we apply this in such a way as to be rotationally invariant.

Generating Blobs

We take a window of data, and from this a single channel of sensor data. This is then wrapped around as demonstrated in **Figure 3** in order to create a blob. The mapping here is deterministic, with minimal loss in sensor data. A dynamic scale factor (Eq1, **Figure 4**) is used to avoid signal noise being scaled into a meaningful blob size.

The Game

The player is presented with a number of blobs; on the left is a set of blobs from one class, on the right another. The goal is to draw an abstract shape which serves as a feature going forwards, namely that it will effectively assist a system in classifying – or distinguishing between – two categories. For each problem the player is presented with, the shape they draw is automatically evaluated, and the life bar adjusted, based upon the different between the left and right fitting energies (below).

Feedback and Fitting Metrics

The system uses an adaption of a snake fitting algorithm. This works by reducing the touch input into a series of points, and thus a polygon. This polygon is then deformed to fit each individual blob in turn, with the energy being the average change in segment wise length, normalized over the length of the polygon, with the formula being provided in Eq2 (**Figure 4).** The benefits of such an approach is that an identical approach can be used in the final (automatic) development of a system, as well as being efficient enough to quickly compute on mobile devices.

Conclusion

This submission has overviewed BlobSnake, a novel and exploratory approach towards engaging the public in Human Activity Recognition system development, as well as aiming towards contributing to science communication in a legal setting. This authors looks forward to discussing these concerns at the workshop itself, and beginning to advance these two agendas.

Acknowledgements

The author would like to thank those he has worked with on this submission: Patrick Olivier, Carlton Shepherd and Thomas Ploetz. This work was supported by an EPSRC DTA Award and a Google Scholarship.

References

[1] Bulling, A. et al. 2013. A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors. *ACM Computing Surveys*. (2013).

[2] Hammerla, N. et al. 2013. On Preserving Statistical Characteristics of Accelerometry Data using their Empirical Cumulative Distribution. *ISWC* (2013).

[3] Kirkham, R. et al. 2013. The Breaktime Barometer – An Exploratory System for Workplace Break-time Social Awareness. *Ubicomp 2013: Ubiquitious Computing* (2013).

[4] Meyerson, M.I. 2010. Significant Statistics : The Unwitting policy Making of Mathematically Ignorant Judges Significant Statistics : The Unwitting Policy Making of Mathematically Ignorant Judges. 37, 3 (2010).

[5] Plötz, T. et al. 2012. Automatic Assessment of Problem Behavior in Individuals with Developmental Disabilities. *Ubiquitous Computing 2012* (2012).

[6] Ward, J.A. and Gellersen, H.W. 2011. Performance Metrics for Activity Recognition. *ACM Transactions on Intelligent Systems and Technology*. 2, 1 (2011), 1–23.

Gamifying Research on Children's Understanding of Law in their Lives

Effie L-C. Law
Department of Computer Science
University of Leicester
LE1 7RH Leicester
U.K.
lcl9@le.ac.uk

Joanna P.L. Barwick
School of Law
University of Leicester
LE1 7RH Leicester
U.K.
jplb1@leicester.ac.uk

Dawn E. Watkins
School of Law
University of Leicester
LE1 7RH Leicester
U.K.
dew3@leicester.ac.uk

Elee Kirk
School of Law
University of Leicester
LE1 7RH Leicester
U.K.
ek170@leicester.ac.uk

Abstract
Based on the assumptions that games are inherently motivating and that typically children are familiar with games, a game is being developed as a research tool to capture how far children aged 7-11 years old understand law in their everyday lives. The game consists of legal scenarios in four different settings. The participatory design approach with the traditional focus group technique has been used to gather children's feedback on the draft scenarios. We also reflect on strategies, opportunities, challenges, and ethics concerning gamified research.

Author Keywords
Law; Children; Gamified research; Everyday lives;

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. See:

Introduction
Gamifying learning in the form of digital educational games (DEGs) for children is an endeavor that has already attracted many research studies in the last decade (e.g., [1], [2], [3]). The publication of Marc Prensky's work [4] at the turn of millennium has arguably fuelled enthusiasm in researchers and boosted uptake by practitioners (e.g., teachers , game

Figure 1. **Non-player character, Lex, of the LICL game.**

designers). In contrast, gamifying research with children is a more recent research effort with increasing interest. Both types of gamification activity are based on the assumption that games are inherently motivating and engaging and on the observation that digital games are an integral part of most children's lives. Games, when well-designed, can sustain children's attention, curiosity as well as cooperation in dealing with the given tasks or quests. Additionally, both activities are applied in a variety of domains, more prevalent in science and technology and relatively less in social sciences, arts and humanities.

In this position paper, we report our newly launched research project – Law in Children's Lives (LICL)[1], utilizing games as a research tool to find out how children understand law in their everyday lives. In the following we present an overview of LICL and then discuss four issues pertaining to gamifying research - strategies, opportunities, challenges, and ethics – as the focus of this workshop.

Overview of LICL

According to [5], gamification is "the use of game design elements in non-game contexts". LICL is an 18-month research project aiming to utilize the power of gamification to research with school children. The research method commonly used to elicit children's understanding of a particular topic is semi-structured interview with or without the use of props (e.g., pictorial scenarios). However, there are some known drawbacks of this conventional approach, which is: i) time-consuming; ii) prone to the social desirability or experimenter effect [6]; (iii) difficult to engage children in imagining scenarios vividly and responding accordingly. Digital games are deemed as a viable solution for these issues.

The main goals of the LICL project are to examine children's awareness of certain legal provisions that apply to them and to assess children's perception of their being empowered by these laws in their everyday lives. The earlier related work has looked into these issues but rather from a law-first approach [7], assessing the level of children's understanding of legal institutions and processes ([8], [9]) and the role of legal actors ([10]). In LICL we adopt a child-centred approach and are particularly interested in finding out how far children demonstrate legal competence in their decision-making.

Currently, the LICL project is designing an android tablet-based game app for data collection. Four settings with which children are most familiar have been identified, namely school, shop, park, and a friend's home, as micro-worlds of the game. In each of the settings, several scenarios are presented where children are expected to apply their legal competence to interpret the situations and make decisions accordingly. For instance, in a shopping scenario, after paying for an item a child unwraps its packaging and surprisingly finds that the item is faulty (e.g., a broken toy or a mouldy bar). What the child would choose to do (e.g., asking for refund, throwing away the item) and the rationale underlying the choice of action will be captured by the game through simple dialogues between the child and an alien, Lex (non-player character [11]; Figure 1). Through systematic analyses of such dialogues, the child's understanding of the relevant law can be inferred.

[1] www2.le.ac.uk/departments/law/research/law-in-childrens-lives

The target groups of the LICL project are children aged 7 to 11 years old. The game has been developed through participatory design [12] with a sample of 16 children from three primary schools, comprising two groups of dyads from each of the four school years (Year 3, 4, 5 and 6, covering the age bracket of 7-11). The children had been asked to comment on the scenarios, which were presented to them orally by a researcher. Then we adapted the scenarios.

After the pilot phase, the main study with the executable game prototype will be conducted in a representative sample of schools in summer 2015. Children will play the game individually in the classroom and some also take the tablet home to play with their parents or carers. This presents the opportunity to interpret how children's decision-making would be influenced by adults and vice versa - another innovative feature of the project.

Discussion

The workshop aims to reflect on the four aspects of gamifying research. We discuss them based on the insights hitherto gained from the planning and running of the LICL project.

Strategies:
Gamifying research, while being a recent trend, is not entirely new. The initiative "games with a purpose" (GWAP) dated back in 2006 can arguably be the pioneer work of gamified research. GWAPs[2] such as Verbosity (a game for collecting common-sense facts) and Squigl (a game in which players trace the outlines of objects in photographs) are not bound to a particular discipline or domain. Games with simple rules and graphical design could serve as GWAPs. Indeed, games as simple as tetris can be engaging. However, when a game is linked to a particular topic of a domain, it is essential to synthesize different perspectives and expertise in game design and domain-specific knowledge – a challenge proved rather tricky to resolve. Methodologically, it may *not* be adequate to rely solely on a gamified research tool for data collection. A mixed-method approach integrating the strengths of the traditional and gamified research techniques and tools is recommended. This can help triangulate empirical findings as well.

Opportunities:
Applying gamified applications to elicit responses from children, who may otherwise find it difficult to adopt the role depicted in the scenarios on legal and right-based issues, is deemed promising. With mobile technologies becoming more versatile, affordable and usable, the portability of the gamified research tool - the tablet-based game - is the opportunity that the LICL project taps to gather data in children's home, stretching beyond the confine of a school setting. Another opportunity enabled by a gaming environment is the capture of multimodal data: textual, audio, video, and automatic logging. Nonetheless, research with children restricts the types of data (e.g., videos) to be collected unless special arrangements are made. Furthermore, as the game-based scenarios induce some sense of realism in the participants, who, when immersed in the game, may express their emotions somewhat naturally in their audio responses. This can facilitate the researcher to understand the children's experience patterns.

[2] www.cmu.edu/homepage/computing/2008/summer/games-with-a-purpose.shtml

Project team biography

The LCIL project is funded the Economic and Social Research Council (ESRC) in the UK. It is being led by Dr Dawn Watkins from the School of Law at the University of Leicester. Dawn is a former lawyer with expertise in family law and legal education.

The other project members are:

- Dr. Effie Law, Reader, whose expertise is in HCI, focusing on user experience and usability methodologies, and in technology-enhanced learning, especially game-based learning;
- Dr Joanna Barwick, Research Associate, whose research interests are the social and cultural significance of games.
- Dr Elee Kirk, Research Associate, with expertise in using innovative research methods with young children;

Challenges:

Similar to the traditional research methods and tools, selecting representative samples for a gamified research study is critical for the validity and reliability of empirical data. Nonetheless, as games, despite their popularity, are not to everybody's liking, assuming that gamified research works regardless of individual preferences can compromise the quality of empirical findings. Developing a gamified research tool involves a number of design decisions (e.g., 2D vs. 3D graphics), which are relevant to entertainment games as well. However, a particular challenge of developing a gamified research tool is to strike a balance between gaming and directing a player's focus onto the topic in question. In other words, the game elements should not be too prominent or engaging lest they would distract the player from the main tasks.

Ethics:

Like all research studies involving human participants, in developing and deploying a gamified research tool, potential ethical and moral issues need to be taken into serious consideration. In particular, the playful guise of the game may prompt the research participants to share unwittingly more than they otherwise would. To safeguard against this potential risk of manipulation, all participants (children, parents/carers) and stakeholders (teachers) must be well-informed about the purpose of the gamified research tool, the goal of the overall study, and their participation rights. Consent forms, written in a jargon-free and easy language, must be understood and signed by the participants. In LICL, individual legal scenarios will also be debriefed shortly after the game play session.

References

[1] Gee, J.P., 2003. What videogames have to teach us about learning and literacy. Basingstoke: Palgrave Macmillan

[2] Squire, K. (2003). Video games in education. *Int. J. Intell. Games & Simulation*, *2*(1), 49-62.

[3] Law, E. L. C., & Sun, X. (2012). Evaluating user experience of adaptive digital educational games with Activity Theory. Int'l J. of Human-Computer Studies, 70(7), 478-497.

[4] Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE)*, *1*(1), 21-21.

[5] Deterding, S. et al. (2011). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 EA* (pp. 2425-2428). ACM.

[6] Muller, M.J. (2007). Participatory design: The third space in HCI. In J. Jacko and A. Sears (Eds.), *Handbook of HCI* (2nd Ed.). Erlbaum.

[7] Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, 303-315.

[8] Sarat, A., & Kearns, T.R. (1995*). Law in everyday life*. Michigan: University of Michigan Press.

[9] Flin, R., Stevenson, Y., & Davies, G. (1989). Children's knowledge of court proceedings. *British Journal of Psychology, 80*, 285-297.

[10] Peterson-Badali, M., Abramovitch, R. & Duda, J. (1997). Young children's legal knowledge and reasoning ability. *Canadian Journal of Criminology, 39*, 145-170.

[11] Powell et al. (2008). Children's perceptions of role of police: a qualitative study. *International journal of police science and management, 10* (4), pp. 464-473.

[12] Lankoski, P., & Björk, S. (2007). Gameplay design patterns for believable non-player characters. In *Proc. DiGRA* (pp. 416-423).

Experiences with a Gamified Online Questionnaire for Crowdsourcing Human Recycling Capabilities

Pascal Lessel
DFKI GmbH
Saarbrücken, Germany
pascal.lessel@dfki.de

Maximilian Altmeyer
DFKI GmbH
Saarbrücken, Germany
maximilian.altmeyer@dfki.de

CHI 2015. Workshop: Researching Gamification: Strategies, Opportunities, Challenges, Ethics

Abstract
In this paper we share our experiences with an online questionnaire which had as its main goal to crowd-source how people classify various objects for recycling. To keep people engaged to complete it, and to assess gamification elements we planned to use in a persuasive system for this task later on, we had already integrated these elements into the questionnaire. Besides positive feedback from some participants, we also learned that there are drawbacks and pitfalls with such elements that can be problematic depending on the hypotheses to be addressed with the questionnaire.

Author Keywords
Crowdsourcing, gamified online questionnaire, recycling

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction
Sorting garbage is a relevant topic as world cities in 2012 generated about 1.3 billion tonnes of solid waste per year [3]. In terms of recycling, in Germany for example, four (sometimes five) different trash bins for households are available that are designated to hold only a specific kind of trash. If the separation of garbage is done

Figure 1: Classification question

No feedback: Participants only see their score after all classification task (without feedback on their decision) together with the high score list.

Ground truth feedback (GTF): Participants always see whether they decided correctly and what would be the correct answer. The gamification elements available are: seeing own points at th top, how many points are necessary for the next position on the high score list and their current placement on it (both shown after a classification).

GTF with explanation: Same as GTF. In addition, an explanation of how the ground truth decision was given by providing a short statement, and a reference to an official document was shown.

GTF with same crowd decision: Same as GTF. In addition, they see how many people decided in the same way, by seeing a percentage.

GTF with crowd decisions: Same as GTF. In addition, they see how the crowd decided, by seeing a percentage per classification option.

Text box 1: Overview of conditions

properly, it has a positive effect in terms of environmental protection, e.g. by greatly increasing the recovery rate of domestic waste [2]. Nevertheless, not everyone seems to do it properly [1]. Reasons for that might be that the rules on what belongs in which of these trash bins makes it difficult to do properly, or people are simply not motivated enough to do it. In HCI, the topic of encouraging people to reflect on their recycling behavior (e.g. [8]) has been under investigation for a few years now. As discussed in [5] we found the work on the BinCam (e.g. [7]) a relevant and useful approach to encourage people to recycle better. This approach, which uses gamification, relied on the performance of a crowd recruited via Amazon Mechanical Turk (AMT)[1]. This crowd had the task to classify pictures taken by a camera inside the kitchen trash bin, to decide whether all objects were sorted correctly. The performance of the crowd was not good: in a random picture sample, 15 of 20 classifications were wrong [7]. The work on the BinCam did not investigate in detail why this was the case, i.e. whether this is a systematic problem or only one based on the nature of AMT.

To gain insights into this topic, we decided to analyze human capabilities in recycling and whether the *Wisdom of Crowds* [6] can produce better results than individuals in this domain (for details see [5]). We conducted this with the help of a gamified online questionnaire. If the performance of a crowd produced reliable classification results (unlike to the BinCam reports), we would have a big opportunity: If an underlying system design encouraged people to classify pictures of waste, we could not only use this as feedback for intelligent systems, but would also have the chance to educate people participating in the classification process and receiving feedback as well.

[1] https://www.mturk.com, last accessed on 21/02/2015

Gamified Online Questionnaire

To gain insights into this topic, we decided to use an online questionnaire, in which participation was voluntary and without any monetary compensation, to reduce the chance of random answers to earn money faster [4], which could have been an issue at AMT. Participants had the task to classify 40 objects in terms of how they would recycle them in Germany. As we envisioned a game later on that would encourage people to classify such pictures on demand without payment, we also integrated game elements into the questionnaire to assess them a priori. Another motivation for using these was that we hoped to influence the dropout rate positively and to spark more interest in the questionnaire. Figure 1 shows the questionnaire interface. Here, participants needed to decide how they would dispose of waste and were asked to state how confident they were in their decision. To assess our hypotheses (see [5] for more details) it was necessary to use multiple conditions, in which we varied the game elements and feedback types. We had a control group, which had not received any feedback, and the only game element was that participants knew they would receive a score in the end and could compare it to other participants' scores. We also had four feedback groups which were also accompanied with gamification elements; for an overview see Text box 1. The feedback was provided together with a happy or sad emoticon and additional information, depending on the conditionm, which were equally distributed (based on completed runs). We explained in the beginning that points are given for correct answers and subtracted otherwise. Moreover, we told the participants that they can gain bonus points by answering quickly (a precondition later on in our game setting). Participants were recruited via social media and we requested that they had lived at least three years in Germany, to ensure that they were familiar with local

recycling rules. Besides the classification task, we also asked questions about their waste sorting behavior and how they (would) assess the game/feedback elements (if they were in a group without them). After finishing these tasks, participants had the chance to provide us with their e-mail address for a follow-up study. This took place one week later, in which we only showed objects they had classified wrongly in their first run. We wanted to assess whether we could educate people even if they did not know that they would be re-tested.

Experiences

Besides findings related to our different research hypotheses (which we discuss in depth in [5]), we also received some insights on the use of gamification and feedback elements in an online survey:

- **Positive feedback**: We did not ask for an overall impression of the "game" during the questionnaire, but as we promoted it over social media, we had the chance to collect reactions, providing at least anecdotal evidence. Most comments on links and e-mails that reached us illustrated that the competition was perceived positively, and 64% also entered a nickname onto a high score list. Some users posted their position on this list as a comment under the link to the survey and tried to mock other players for doing worse. People also tried to match nicknames to other participants ("wild guesses" under the comments). Both raised ethical questions as the anonymity breaks in this case. In two cases a discussion about specific pictures arose. It is questionable whether the same "meta-talk" would have happened in a "standard" questionnaire. We are keen to investigate such aspects in the next gamified online questionnaire directly. Our questions on the feedback/game elements also showed that they were perceived positively, overall.
- **Number of dropouts**: 66 runs were not completed (26.4% of all runs). Considering the dropout rate of participants who had done at least one classification, the no-feedback condition produced the lowest rate (4 of 49), followed by the feedback condition in which the distribution of all answers was shown together with the ground truth (7 of 49). The other conditions produced 14 (showing how many people have decided the same) and the two remaining 12 dropouts. It is currently unclear, whether the feedback or the gamification elements caused this, but it is an issue that needs to be kept in mind in later online questionnaires: Showing feedback might demotivate people, especially if they disagree with the ground truth (as recycling rules might also differ within a country).
- **Conversion rates for follow-up study**: The number of participants in the follow-up run after one week was lower than expected. Of the 184 participants who completed the first questionnaire, only 36 (19.6%) participated. Besides the aforementioned reason, it indicates that the game elements were not rewarding enough to consider participating again. People in the no-feedback condition were more likely to do this follow-up (33%), while in the four feedback conditions only between 8% to 23% took part. The worst result came from the explanation group. In general, again, it seems that the feedback discouraged participants from continuing. Another explanation could be that we only allowed participation once (using technical countermeasures and a description in the introduction stating this). This is counterintuitive in

games[2] and might have led to a lower rate of participation, as people who performed badly initially might have lost interest in doing the follow-up study.

- **Timing issues**: Even though we could show that people in the feedback conditions produced better results over time than in the non-feedback condition, we were not able to find any significant differences between the feedback conditions. We assume that the information that faster decisions produced bonus points led to participants assuming that the clock was also running during a feedback cycle and they did not read thoroughly through the conveyed feedback, but only glimpsed the correct answer (which was shown in all feedback conditions). Hence, with this setup, we learned that it is crucial to clarify more precisely aspects that do not harm the game score. We also found that a significant number of people chose only extreme confidence values, which also seems related to this timing issue.

Discussion

Even though we could find answers to most of our hypotheses stated in [5], we learned that a gamified questionnaire could also have drawbacks that needed to be tackled beforehand. These pitfalls also showed that the usage of game elements might not always be advisable and certain questions might be answered more clearly if no game elements or feedback potentially introduce a new source of bias. Nevertheless, our identified pitfalls here could also have been introduced by simply relying on the wrong set of gamification elements for our demographic (the qualitative questions indicated otherwise though [5]) or that the overall impression of our gamified survey was still more a survey than a game (hence reducing the effectiveness of the chosen elements). For the case at hand it seems quite interesting to see what happens if the main design is a game instead of a survey and a comparison of the results here, with the results made in this game might be worthwhile to investigate.

[2] see also: http://goo.gl/G3Fi5w last accessed on 21/02/2015

References

[1] Environment Bureau Hong Kong. Blueprint for Sustainable Use of Resources. http://goo.gl/JfSUZW, May 2013. [last accessed 21/02/2015].

[2] Environmental Protection Department. Programme on Source Separation of Domestic Waste - Annual Update 2010. http://goo.gl/c1lmjH, May 2010. [last accessed 21/02/2015].

[3] Hoornweg, D., and Bhada-Tata, P. What a Waste: A Global Review of Solid Waste Management. *Urban Development Series*, 15 (March 2012), 1–116.

[4] Ipeirotis, P. G., Provost, F., and Wang, J. Quality Management on Amazon Mechanical Turk. In *Proc. of HCOMP 2010*, HCOMP '10, ACM (2010), 64–67.

[5] Lessel, P., Altmeyer, M., and Krüger, A. Analysis of Recycling Capabilities of Individuals and Crowds to Encourage and Educate People to Separate Their Garbage Playfully. In *Proc. CHI 2015 (to appear)*, ACM (2015).

[6] Surowiecki, J. *The Wisdom of Crowds*. Anchor, 2005.

[7] Thieme, A., Comber, R., Miebach, J., Weeden, J., Krämer, N., Lawson, S., and Olivier, P. "We've Bin Watching You": Designing for Reflection and Social Persuasion to Promote Sustainable Lifestyles. In *Proc. CHI 2012*, ACM (2012), 2337–2346.

[8] Zlatow, M., and Kelliher, A. Increasing Recycling Behaviors Through User-Centered Design. In *Proc. DUX 2007*, ACM (2007), 27:1–27:1.

Biography

Pascal Lessel studied Computer Science at Saarland University. In 2012 he has started to work as Researcher at the German Research Center for Artificial Intelligence focusing on digitizing of paper-based artifacts and persuasive technologies. Early in 2014, he identified the combination of Gamification and Crowd computing as interesting research direction.

IT'S COMPLICATED: The ethics of gamified labor

Marigo Raftopoulos
GEElab, School of Media and Communication
RMIT University
Melbourne, Australia
marigo.raftopoulos@rmit.edu.au

Steffen P. Walz
GEElab, School of Media and Communication
RMIT University
Melbourne, Australia
steffen.walz@rmit.edu.au

CHI 2015, Seoul, South Korea.
ACM 978-1-XXXX-XXXX-X/XX/XX

Abstract
An increasing number of applications are using gamification in research and participatory problem solving, however several ethical issues are beginning to emerge that may compromise their integrity. Our paper highlights the ethical issues of using gamification to extract unpaid labor, and the use of persuasive gamification design practices that can potentially be considered exploitative. We conclude by suggesting the collaborative development of an industry framework based on a value-sensitive design to overcome these issues.

Author Keywords
Gamification, labor, ethics, innovation, problem solving, value-sensitive design, games with a purpose.

ACM Classification Keywords
H.5.1. Multimedia information systems

Introduction
The digital economy has produced new and pervasive forms of engagement and participation in research, problem solving and value creation. At this point in time we are seeing a confluence of approaches such as crowdsourcing, collaboration and gamification that are rapidly being adopted by organisations to access data, accumulate cognitive resources or solve problems far more cost-efficiently than at any other time in history.

The immediate private and social benefits of this phenomenon are significant, and this has been the key reason that the rate of adoption of gamification has been able to successfully spread across many industries and domains. There are however significant ethical issues that have been overlooked during this hyper-growth period. The confluence of these pervasive technologies has socialized us into a system of gamified labor [1] or digital labor [2] where these new systems and applications have created a new cognitive working class. Today's peer-to-peer values of openness, participation, co-creation, creative-commons orientation [3] and fun [4] have created a new kind of work, and much of it is unpaid in economic terms on the assumption that the public are receiving intrinsic benefits derived from participation. A more critical view is that this phenomenon commodifies cognitive or intellectual labor while capital accumulation remains with those who own the digital assets. In addition to this, it is questioned whether human interaction with these gamified systems is facilitated by persuasive technologies to encourage participation.

In this position paper we will discuss the challenges associated with gamified labor and research, and finish with a call to action for participants in this CHI15 workshop. One of the key challenges we have as researchers is how we can shape the gamification of research to democratize labor processes rather than use it to prop up existing economic constructs that facilitate exploitation or an uneven distribution of economic reward. Here is where the complications set in: Gamifying research may lead to productivity and innovation, however, ethical considerations challenge the nature of design decisions, the investment decisions, and how profits are distributed, placing any potential gains at risk. There are no mechanisms in place to manage this tension fairly and equitably under current constructs.

Our research into gamified systems

There are a wide range of gamified systems and applications used across many different domains. Our database of over 300 enterprise gamification examples, accumulated as part of our doctoral research, shows that 14% were identified as cases where gamification was used for collaborative problem solving or innovation [20]. In terms of the technologies used in these examples, 12% were digital games or simulations, 58% were gamified platforms, websites or applications, and 30% were playful experiences that involved physical interaction by users with a gamified digital application. These examples also showed that there was a wide spread of target audience for the application, i.e., 23% were targeted to internal staff, 30% were direct to customers, clients or patients, 16% were targeted to a specific industry or community and 30% were targeted at the general public. Key features of these gamified applications included crowdsourcing, collaboration and data capture to extract the resources required to solve the project objectives. However on closer investigation of the design and investment decisions that were made in this sample of projects, there was no common or established best practice on identifying or managing potential ethical issues.

Citizen science games like 'FoldIt'' [5] and Games with a Purpose (GWAP) show how online games can be successfully used to solve large-scale problems [6] [7] [8]. Using games and game-like environments to solve problems has received wide attention in the popular media [9] [10] [11] and this attention has raised public awareness and willingness to using games and gamification experimentally in non-entertainment contexts. Gamification is known for its engagement and fun, however this obscures the nature of these games as work [12] and their potential exploitation.

Research into GWAP shows us that the key motivation for people to play a game was not driven by the fact

that they will solve a problem, but to be entertained [6] [7] [8]. The implication is that in designing a game with a purpose we need to primarily design for engagement, as the intrinsic motivator is not sufficient on its own. This opens the way for designers to adopt persuasive game design techniques that may compromise project integrity. Questionable persuasive technologies include:

- Using persuasive technologies or captology where human emotions, actions and behaviors are shaped and reinforced through technologies such as surveillance, conditioning and channeling.
- Using gamification as an operant conditioning type of persuasion tool where technology shapes human behavior through a predetermined schedule of reward and punishment [13] [14]
- Undertaking data collection that can potentially compromise individual privacy through performance monitoring, surveillance and data 'leakage' in gamified enterprise applications, which are issues that form part of the wider human–computer interaction discourse on data, privacy and ethics of persuasive technologies [15] [16] [17] [18]

A key method that may assist in overcoming the potential ethical issues raised in this paper is to utilize the key methodologies available to us in software design, for example, value-sensitive design (VSD).

Value-sensitive design and gamification

VSD is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner [19]. The utilization of VSD can be used to ensure a more ethical approach to gamified research design. Key elements that can be integrated into the design process includes the VSD tripartite model of conceptual, empirical and technology to identify key stakeholders, the values that are implicated, how value is created and appropriated in the application, and how technology design can support stated values. Where this method can benefit in the design of gamified research is the consideration of 13 individual VSD 'values' as part of the consultation and participatory design process, which includes mechanisms that highlight human welfare, ownership, property, privacy, informed consent, trust and identity.

There are many pragmatic challenges with developing and implementing VSD, however the potential benefits provide an important counterbalance to the potential tensions that can be caused by unethical gamified research design and practices. Discourse on values in game design already has a rich history. For example Flanagan's work on 'values at play' in designing for values in socially-oriented game design [21] is a good base that can help inform how we can apply ethics to gamified research.

Conclusions

The challenge to the research community is to come together to evaluate how we can utilize these existing tools to address the ethical challenges we are facing in gamified research. This will also provide the potential to build improved systems that do not rely on the exploitation of labor, but enables the full potential of human creativity and innovation.

Citations

[1] Dewinter, J. and Kokurek, C. Games, Gamification and Labour Politics, *Journal of Gaming and Virtual Worlds*, 6, 2 (2014) 103-107.

[2] Bauwens, M., Thesis on Digital Labor in an Emerging P2P Economy. In Digital Labor: The Internet as Playground and Factory (ed) Scholz, T., Routledge, New York, 2013.

[3] Ibid.

[4] 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: Tampere, Finland (2011) 9-15.

[5] Khatib, F., Cooper, S., Tyka, MD., Xu, M., Makedon, I., Popović, Z., Baker, D., and Foldit Players, Algorithm discovery by protein folding game players, University of Washington, *Proceedings of the National Academy of Sciences of the United States of America*, 108, 7 (2011).

[6] Von Ahn, L. Games with a Purpose, Invisible Computing, Intel Research, June 2006 http://www.cs.cmu.edu/~biglou/ieee-gwap.pdf

[7] Von Ahn, L. and Dabbish, L. Designing Games With A Purpose. Communications of the ACM (contributed articles) *Association for Computing Machinery*, 51, 8 (2008).

[8] Siorpaes, K., and Hep, M. Games with a Purpose for the Semantic Web, IEEE Intelligent Systems, *IEEE Computer Society* (2008).

[9] McGonigal, J. Reality Is Broken: *Why Games Make Us Better and How They Can Change the World. Penguin*, London, 2011.

[10] Reeves, B. and Read, J.L. *Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete*. Harvard Business School Press, Boston, 2009.

[11] Zichermann, G. and Linder, J. *Game-Based Marketing: Inspire Customer Loyalty Through Rewards, Challenges, and Contests*. Wiley, Hoboken, NJ, 2010.

[12] Sutton-Smith, B. *The Ambiguity of Play*, Harvard University Press Cambridge, MA, 2001.

[13] Llagostera, E., 'On gamification and persuasion', *Brazilian Symposium on Computer Games and Digital Entertainment*, 2-4 November, Brazil, (2012) 12-21.

[14] Schell, J., *The Pleasure Revolution*, http://venturebeat.com/2011/12/05/jesse-schell-talks-about-the-pleasure-revolution/. Accessed 14 May 2014.

[15] Albrechtslund, A., 'Ethics and technology design', *Ethics and Information Technology*, 9,1 (2007) 63-72.

[16] Ball, K., Categorizing the workers: Electronic surveillance and social ordering in the call center, in D. Lyon (ed.), *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination*, Routledge, London (2003) 201–25.

[17] Berdichevsky, D., and Neuenschwander, E., Towards an ethics of persuasive technology, *Communications of the ACM*, 42,5 (1999) 51–58.

[18] Carroll, J. M., Human-computer interaction: Psychology as a science of design, *International Journal of Human-Computer Studies*, 46,4 (1997) 501–22.

[19] Friedman, B., Kahn, P. H. and Borning, A., Value sensitive design and information systems, in K. E. Himma and H. Tavani (eds), *The Handbook of Information and Computer Ethics*, John Wiley and Sons Inc, New Jersey (2008) 69-101.

[20] Raftopoulos, M., Waltz, S., and Greuter, S., How Organisations Play: Towards a taxonomy for enterprise gamification (under consideration)

[21] Flanagan, M., Howe, D., and Nissenbaum, H. Values at play: Design tradeoffs in socially-oriented game design. CHI 2005, ACM Press (2005), 751-760.

Score design for meaningful gamification

Simon Scheider
ETH Zurich
Stefano-Franscini-Platz 5
8093 Zurich, Switzerland
sscheider@ethz.ch

Martin Raubal
ETH Zurich
Stefano-Franscini-Platz 5
8093 Zurich, Switzerland
mraubal@ethz.ch

Peter Kiefer
ETH Zurich
Stefano-Franscini-Platz 5
8093 Zurich, Switzerland
pekiefer@ethz.ch

Christian Sailer
ETH Zurich
Stefano-Franscini-Platz 5
8093 Zurich, Switzerland
csailer@ethz.ch

Paul Weiser
ETH Zurich
Stefano-Franscini-Platz 5
8093 Zurich, Switzerland
pweiser@ethz.ch

CHI'15, April 18–23, 2015, Seoul, Korea.
Gamifying Research: Strategies, Opportunities, Challenges and Ethics

Abstract
In this paper, we provide an overview on the design of scores that can be used in gamification and sketch how user behavior can be influenced by design and communication.

Author Keywords
Scoring methods, meaningful gamification

ACM Classification Keywords
H.5.2 [Information interfaces and presentation (HCI)]: User-centered design.

Introduction
The effectiveness of gamification relies on feedback loops which influence user behavior. Such feedback loops involve (1) measuring behavior, (2) relating it to other behaviors or norms (relevance), (3) "illuminating the path ahead" (consequence) and (4) action[1] [2]. Scores, i.e., quantitative evaluations of behavior in a game, play an essential role in the second stage of this loop. The challenge lies in designing scores which allow users to internalize externally intended behavior, and thus enable *meaningful gamification* [6].

[1] Thomas Goetz: Harnessing the Power of Feedback Loops, http://www.wired.com/2011/06/ff_feedbackloop/

In a preliminary walk through, we analyze relevant dimensions of score design and their role in the whole process, as summarized in Figure 1. In order to illustrate our analysis, we use the example of gamified apps which support people to act in an environmentally sustainable way, e.g., UbiGreen [1] or GoEco! [8].

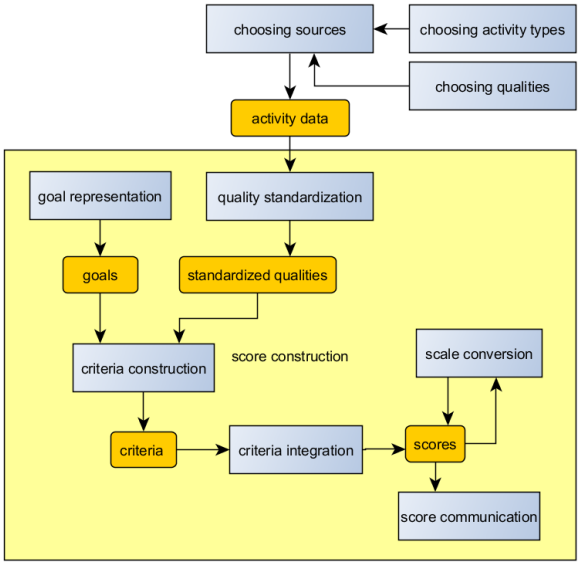

Figure 1: Process model of activity scoring. Blue boxes are processes, orange boxes are outputs.

Domains, qualities and sources of scoring

The question of *what* should be scored involves two steps. First, selecting the *activity types* to be scored. In our example, we are mainly interested in personal mobility, e.g., travel and daily mobility. Second, deciding which *qualities* of these activities should be taken into account. There are qualities of the activities themselves as well as qualities of outcomes [9]. In our case, the former involve velocity, cost or CO_2 emissions. The latter include whether traveling takes you to your destination on time. The same domain can be scored based on different sources: mobile technology offers new ways of *measuring* user activities by sensors. In our case, we can use acceleration and position sensors to determine travel modes [11]. However, there is also the possibility of scoring based on human ratings [9]. For example, users can self-rate their mobility performance and peers can tag each others' mobility behavior with "Likes".

Score construction

This section addresses *how* should be scored. Once qualities are obtained, they need to be evaluated with respect to *goals* which turns qualities into *criteria*. The latter can be used to obtain *scores*, i.e., evaluations of behavior.

Goal representation

Since scoring criteria need to be defined with respect to goals, an important decision concerns which goals are to be taken into account. User goals and goals of a developer need not coincide. Therefore, external goals need to be associated with user goals in order to facilitate internalizing externally promoted behavior [6] and, furthermore, to assure usability [7]. Taking *many goals* into account increases the chance that users can find their own relevant goals reflected in a gamified application. Another option is to let users define their own goals which avoids *technology paternalism* [4].

Quality standardization

Qualities can be *standardized* in different ways, by comparing them to: (1) the past, to measure individual change (e.g., reduction of CO_2 emissions); (2) the

behavior of others, based on leaderboards, or other types of statistics; (3) established norms, e.g., CO_2 emission contingents can be based on the 2^oC standard of temperature rise[2]; (4) a set of *conceivable alternatives*. In the latter case, we standardize qualities of activities with respect to what a user might have done instead, given his or her own goals. For example, staying at home instead of commuting to work would reduce CO_2 emissions but is not a feasible alternative for most users. Using public transport instead of a car might be an alternative for urban dwellers but not for people living in rural areas. Standardization by conceivable alternatives allows us to embed external goals into a user's context. However, it requires a rather detailed user model[3] and fixed limits on what is considered conceivable.

Criteria construction
Once standardized, qualities can enter *criteria construction*. That is, it needs to be evaluated how far values of qualities contribute to a goal. This requires comparing values to favored states implied by a goal. For example, keeping CO_2 emissions within internationally established contingents might be considered a favored state with respect to climate protection. However, it might make more sense to choose a personalized standard as a favored state, such as generating CO_2 savings with respect to one's own past or in competition with others, in order to keep motivation alive.

Criteria integration
Once criteria are established, they can be turned into a single score. Several strategies to integrate multiple criteria can be used, ranging from compensatory to non-compensatory multi-criteria decision making techniques [5]. For example, criteria for green mobility (e.g., reducing CO_2 emissions) need to be integrated with others (e.g., ensuring food supply), because both can contradict each other (e.g., carrying bulky shopping items in trains is difficult). For most people, daily necessities can not be compensated with long term goals. Similar to the standardization by conceivable alternatives (see above), multi-criteria decision support allows embedding external goals into a user's context. However, this time, different goals can enter the score continuously, e.g., in terms of a weighted sum.

Scale levels
Scores can have different scale levels, ranging from nominal over ordinal, interval to ratio scaled [12]. For example, a *badge* received for green mobility behavior in the past means that a ratio scale (CO_2 emission sums) was turned into a nominal scale (according to a minimal amount of CO_2 savings). An example for an ordinal scale are the narrative progression icons of the UbiGreen app [1] which reflect individual mobility behavior during a week. Choosing a scale level affects how much meaningful information people can extract out of a score and whether behavior can be assessed as critical.

Score communication and choice suggestion

The influence of scores on behavior depends on the scores themselves, as well as on their presentation. For example, feedback on energy consumption of households based on scores standardized with respect to means of household neighborhoods causes a boomerang effect for those households below the mean. They tend towards this mean [10]. This effect can be avoided if information is accompanied by smileys indicating approval of behavior. A particular design challenge is to make people aware of

[2] "Copenhagen Accord". U.N. Framework Convention on Climate Change. United Nations. 18 December 2009.

[3] Which, in turn, further increases the need for privacy protection.

concrete choices in a situation. Hassenzahl and Laschke's *pleasurable troublemakers* [3] are objects which embody alternatives to default behavior through what they afford, such as lamps that require deliberate actions to be kept on.

Conclusion

In this paper, we explored the design choices regarding construction of scores in gamification. The main challenge is representing the situated user context in order to internalize external goals. Scoring can take account of this in several ways, namely through quality standardization w.r.t. conceivable alternative behavior, as well as multi-criteria integration. A further challenge is to communicate scores and alternative behavior choices in a decision situation. Future work should more deeply connect our model to measurement and design theories and investigate its generalization over further use cases.

Acknowledgements

This research was supported by the Swiss Competence Center for Energy Research (SCCER) Efficient Technologies and Systems for Mobility and the Commission for Technology and Innovation (CTI).

References

[1] Froehlich, J., Dillahunt, T., Klasnja, P., Mankoff, J., Consolvo, S., Harrison, B., and Landay, J. A. Ubigreen: investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2009), 1043–1052.

[2] Goetz, T. *The decision tree: taking control of your health in the new era of personalized medicine*. Rodale, 2010.

[3] Hassenzahl, M., and Laschke, M. Pleasurable troublemakers. *The Gameful World: Approaches, Issues, Applications* (2015), 167–195.

[4] Huber, M. Z., and Hilty, L. M. Gamification and sustainable consumption: overcoming the limitations of persuasive technologies. In *ICT Innovations for Sustainability*. Springer, 2015, 367–385.

[5] Jankowski, P. Integrating geographical information systems and multiple criteria decision-making methods. *International journal of geographical information systems 9*, 3 (1995), 251–273.

[6] Nicholson, S. A user-centered theoretical framework for meaningful gamification. In *Games+ Learning+ Society 8* (2012).

[7] Norman, D. Cognitive engineering. In *User Centered System Design*, D. Norman and D. S., Eds. Lawrence Erlbaum Assoc., 1986, 31–61.

[8] Rudel, R., Cellina, F., Weiser, P., Raubal, M., and Scheider, S. Goeco! a community based eco-feedback approach to promote sustainable personal mobility styles. In *SCCER-Mobility 1st Annual Conference at ETH Zurich* (2014).

[9] Scekic, O., Truong, H.-L., and Dustdar, S. Incentives and rewarding in social computing. *Commun. ACM 56*, 6 (June 2013), 72–82.

[10] Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. The constructive, destructive, and reconstructive power of social norms. *Psychological science 18*, 5 (2007), 429–434.

[11] Shin, D., Aliaga, D., Tunçer, B., Arisona, S. M., Kim, S., Zünd, D., and Schmitt, G. Urban sensing: Using smartphones for transportation mode classification. *Computers, Environment and Urban Systems* (2014).

[12] Stevens, S. On the theory of measurement. *Science*, 103, 677–680.

Family Support to Improve Physical Activity in COPD: Game as a Research Tool

Noopur Singh
Health Psychology Section
University Medical Centre
Groningen, 9700 AD
The Netherlands
n.singh@umcg.nl

Robbert Sanderman
Health Psychology Section
University Medical Centre
Groningen, 9700 AD
The Netherlands
r.sanderman@umcg.nl

Hylke W. van Dijk
Lectoraat Serious Gaming,
NHL University of Applied
Sciences, 8900CB, Leeuwarden,
The Netherlands
H.W.vanDijk@nhl.nl

Adelita V. Ranchor
Health Psychology Section
University Medical Centre
Groningen, 9700 AD
The Netherlands
a.v.ranchor@umcg.nl

CHI'15 'Gamifying Research: Strategies, Opportunities, Challenges and Ethics' Workshop, April 19, 2015, Seoul, Republic of Korea.

Abstract
Games provide an effective research tool to investigate the interactions and responses among research subjects in a prospective situation. This research investigates the scope of involving family in goal setting, feedback and persuasion to improve physical activity of COPD patients by enhancing their self-efficacy. In this paper, we present a research context that can benefit from using game as a research tool.

Author Keywords
COPD; Self-Efficacy; Physical Activity; Family; Game

ACM Classification Keywords
H.5.m [Information interfaces and presentation]: Miscellaneous; K.8.0 [General]: Games.

Introduction
Chronic Obstructive Pulmonary Disease (COPD) is a group of respiratory diseases characterized by breathing problems like dyspnea (shortness of breath) due to inflammation of respiratory tract and limited airflow. According to World Health Organization[1], COPD is the third leading cause of death in the world. Behavior change is an important element in daily self-management of COPD. Family is often recognized as an important influence that supports or impedes behavior change but

there is insufficient evidence for utilizing family support in interventions for behavior change. This research focuses on COPD patients in the Northern Netherlands. It is worthwhile to note that the Dutch government plans to involve the family in chronic healthcare to tackle the economic burden of ageing and chronic diseases on healthcare system.

Although COPD is irreversible, recommended physical activity improves symptoms and quality of life of the patient. Unfortunately, COPD patients, majorly elderly and less educated, have a sedentary lifestyle. Lack of physical activity creates a loop of worsening symptoms and further reduces physical activity level of the patient. Apart from clinical characteristics, lower self-efficacy is an important psychological barrier to adhere to physical activity routine[7]. Self-efficacy refers to 'a persons belief regarding whether or not they feel they can successfully execute particular behaviors in order to produce certain outcomes'[4].

According to Bandura[3], family support affects self-efficacy but there is limited research about the scope of family in improving self-efficacy or behavior change among COPD patients. In context of other chronic illnesses, results indicate mixed outcomes (support or hindrance to self-management activities) for family involvement in patient's daily self-management activities[11]. Therefore, there is an acute need to explore the scope of family support in COPD population for behavior change and self-efficacy improvement. Hence, this research investigates the scope of involving family in goal setting, feedback and persuasion through gaming technology to improve physical activity and self-efficacy among COPD patients in the Northern Netherlands.

Scope of Family support

Family in Goal-setting
Goal setting is a well-known strategy to achieve behavioral changes in combination with feedback[13]. Recent research in chronic disease management encourages family engagement in goal setting for self-management behaviors and achievement of those goals[12]. There is lack of knowledge about role of family in goal setting for physical activity among COPD patients.

Family in Self-efficacy
According to Bandura[3], self-efficacy can be achieved through various sources. Family could play a role in two of them:

- By providing motivational feedback for 'mastery experiences'. Mastery experience refers to experiences followed by successful completion of a task. Multiple aspects of the motivational feedback, such as, timing, frequency, presentation of the feedback by family to support physical activity requires further research.
- By verbal persuasion, that refers to encouraging someone about his/her capability to perform a particular task. The technique of persuasion in this context requires further research as it sometimes leads to negative outcomes[2].

Aim and Key Questions

The aim of this study is to investigate, if and how family could be involved in goal setting, motivational feedback and persuasion through gaming technology to improve physical activity and self-efficacy among COPD patients. Refer to Figure 1. The key questions this research investigates:

- What is the influence of family involvement on patient's physical activity and associated self-efficacy?
- To what extent family can be involved in goal setting, motivational feedback and persuasion to achieve physical activity targets?
- What are the patient-family relationship factors that influence the achievement of physical activity targets through this collaboration?
- How do variations in timing, frequency and presentation of motivational feedback and persuasion affect patient's self-efficacy and goal achievement?

Figure 1: Research Overview

Research Plan

This research uses game technology as a research tool in combination with traditional research methods to investigate the scope and influence of engaging family on daily physical activity among COPD patients. For instance, we could make use of existing physical activity games that use goal setting, motivational feedback and persuasion. 'Ubifit'[5] and 'Fish'n'steps'[8] that use visual metaphors of a garden and a fish respectively are examples of such games. In the game, the patient and family member would collaboratively play and use persuasion and feedback to achieve the physical activity goal, for example- number of steps/day set in consultation with the healthcare provider.

Participants (COPD patients and their family member) would be invited for pre-game, game-play and post-game sessions. During the pre-game session, information about how to play and access the game will be provided to the players. Additionally, a pre-game questionnaire will be used to collect the information about demographics and relevant relationship factors between the patient and the family member. The pre-game session will be followed by game-play at home. The patient and the family member will play the game in their home settings. The game-play will incorporate variations in motivational feedback and persuasion within the game. For instance, different timings, frequencies and presentations of feedback and motivation would be incorporated in the game to be used by the family member. The logged information during the game play will provide information about the physical activity level and the interactions between the patient and family. The post-game session will consist of an interview session and a post-game questionnaire. The post-game session would evaluate the experience of engaging family in physical activity improvement and player feedback on different variations of goal setting, feedback and persuasion used in the game. This research process would help in investigating the key research questions (mentioned earlier).

Related Works

There is very limited research that use game as a research tool to investigate behavior change among COPD patients. Our approach gathers support from research in different research domains like spatial planning [9] and Information interface and presentation[6] that have

implemented games as a research and a data-gathering tool. In context of COPD, there are few serious game interventions like 'mCOPD'[14] and 'Airflow'[10] that support and train for breathing exercises.

Discussion

Previous research has attempted to integrate healthcare provider and peer support through technology to facilitate individual behavior change. There is a lack of research that investigates the same through family support. Use of game as a research tool is a very recent approach that opens new possibilities in research. Digital game is an excellent tool for data collection and could be structured to provide unobtrusive observation of participant behavior. Game provides a safe environment to investigate scenarios by eliciting human behavior, like, interactions and responses. In this research, game would create a reference experience to reflect upon interactions and responses in a prospective situation where family is involved through game technology to facilitate behavior change.

Conclusion

This position paper presented a research context, which utilizes game as a research tool. This research investigates the scope of involving family in goal setting, feedback and persuasion to improve physical activity and self-efficacy among COPD patients. This research approach would provide a greater insight into our key research questions.

Acknowledgement

This work is part of the research program of University Campus Fryslân (UCF), which is financed by the Province of Fryslân.

References

[1] WHO — The top 10 causes of death. http://www.who.int/mediacentre/factsheets/fs310/en/.

[2] Ann, M., Stephens, P., Franks, M. M., Rook, K. S., Iida, M., Hemphill, R. C., and Salem, J. K. Spouses Attempts to Regulate Day-to-Day Dietary Adherence Among Patients With Type 2 Diabetes. *Health Psychology 32*, 10 (2013), 1029–1037.

[3] Bandura, A. *Self-efficacy : the exercise of control*, 1 ed. Worth Publishers, New York, 1997.

[4] Bourbeau, J., Nault, D., and Dang-Tan, T. Self-management and behaviour modification in COPD . *Patient education and counseling 52*, 3 (Mar. 2004), 271–7.

[5] Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., and Landay, J. a. Activity sensing in the wild: a field trial of ubifit garden. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)* (2008), 1797–1806.

[6] Cosley, D., Ludford, P., and Terveen, L. Studying the effect of similarity in online task-focused interactions. *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work - GROUP '03* (2003), 321–329.

[7] Hartman, J. E., Boezen, H. M., de Greef, M. H., and Ten Hacken, N. H. Physical and psychosocial factors associated with physical activity in patients with chronic obstructive pulmonary disease. *Archives of physical medicine and rehabilitation 94*, 12 (Dec. 2013), 2396–2402.e7.

[8] Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., and Strub, H. B. Fish n Steps : Encouraging Physical Activity with an Interactive Computer Game. In *UbiComp 2006: Ubiquitous Computing* (2006), 261–278.

[9] Mayer, I. S., Carton, L., de Jong, M., Leijten, M., and Dammers, E. Gaming the future of an urban network. *Futures 36*, 3 (Apr. 2004), 311–333.

[10] Qin, Y., Vincent, C., Nadia, B., and Shi, Y. AirFlow : Designing Immersive Breathing Training Games for COPD. In *CHI* (2014), 2419–2424.

[11] Rosland, A.-M., Kieffer, E., Israel, B., Cofield, M., Palmisano, G., Sinco, B., Spencer, M., and Heisler, M. When is social support important? The association of family support and professional support with specific diabetes self-management behaviors. *Journal of general internal medicine 23*, 12 (Dec. 2008), 1992–9.

[12] Rosland, A.-M., and Piette, J. D. Emerging models for mobilizing family support for chronic disease management: a structured review. *Chronic illness 6*, 1 (Mar. 2010), 7–21.

[13] Strecher, V., Seijts, G., Kok, G., Latham, G., Glasgow, R., DeVellis, B., Meertens, R., and Bulger, D. Goal Setting as a Strategy for Health Behavior Change.pdf. *Health Education Quarterly 22*, 2 (1995), 190–200.

[14] Xu, W., Huang, M.-c., Liu, J. J., Ren, F., Shen, X., Liu, X., and Sarrafzadeh, M. mCOPD: Mobile Phone Based Lung Function Diagnosis and Exercise System for COPD. *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments - PETRA '13* (2013), 1–8.

Measuring User Engagement in an Enterprise Gamified System

Ayoung Suh
School of Creative Media and Dep.
of Information Systems
City University of Hong Kong
18 Tat Chee Avenue, Kowloon,
Hong Kong
ahysuh@cityu.edu.hk

Abstract
The main purpose of enterprise gamification is to increase employees' engagement in work-related activities, such as knowledge sharing, sales performance, idea competition, and training and education, by using game design elements. Currently, researchers are calling for systematic examination of how gamification ideas are executed in the workplace. Despite increasing scholarly and practical attention to the effectiveness of gamified systems in organizations, the question of how to measure the user engagement within an enterprise gamified system remains unclear. This study raises a methodological issue regarding different approaches to operationalizing user engagement. By testing the proposed model that explains the relationships between game dynamics and user engagement with empirical data collected from 128 users of an enterprise gamified system, this study shows how the effects of different game dynamics on user engagement vary depending on the operationalization of user engagement.

Author Keywords
Gamification; user engagement; behavioral modeling; measurement; game dynamics; enterprise gamified system

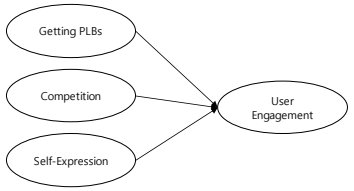

Figure 1. Baseline Model

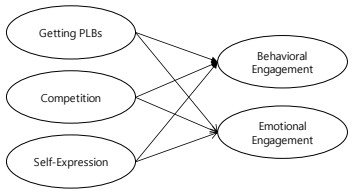

Figure 2. Model Elaboration (Behavioral vs. Emotional Engagement)

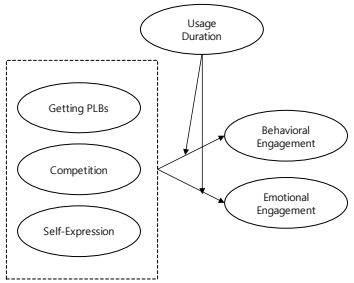

Figure 3. Model Elaboration (The moderating effect of usage duration)

Introduction

Many organizations strategically adopt gamification ideas to enhance their employees' engagement in, for example, knowledge sharing, idea competition, and sales performance (Kapp et al. 2013). Points, levels, and badges (PLBs); leaderboards; and virtual goods are the most commonly used game mechanics in a gamified system. Using these mechanics, organizations are trying to make work-related activities that are not inherently enjoyable game-like and interesting, thus increasing employees' engagement.

While managers acknowledge the potential benefits of gamification ideas, they also point out that it is difficult to maintain sustained user engagement within a gamified system over time. Some researchers have pointed out that the perceived beneficial effects of the game elements used (e.g., getting PLBs) may be short-lived because such extrinsic rewards potentially undermine users' intrinsic motivation (Kankanhalli et al. 2012). For this reason, it has been suggested that a systematic and granular understanding of how to maintain user engagement through game dynamics within an enterprise gamified systems is needed (Deterding, 2014). Understanding the antecedents and motivation of user engagement can assist in the design of appropriate game mechanisms and techniques to enhance their impact (Kankanhali et al. 2012). Of importance is how to operationalize and measure the research variables of interest. What should be measured and how should it be measured in order to determine user engagement?

Theoretical Background

Users are engaged in a system when it "holds their attention and they are attracted to it for intrinsic rewards" (Jacques et al. 1999, p. 58). Figure 1 shows the baseline model that explains the relationship between game dynamics and user engagement. This study identifies three aspects to be considered when operationalizing user engagement.

Behavioral vs. Emotional Engagement

Literature suggests that engagement can be categorized into two perspectives: behavioral and emotional. Behavioral engagement refers to the extent to which an individual is physically involved in doing particular activities during work-role performance (Truss et al. 2013, p. 2659). Emotional engagement refers to a positive state of mind, that is, "an individual's involvement and satisfaction with, as well as enthusiasm for, work" (Harter et al. 2002, p. 269). Behavioral engagement ("doing" engagement) within an enterprise gamified system can be captured by the levels of participation in activities. For example, in the case of a gamified system for knowledge sharing, a user's behavioral engagement can be measured by the number of posts, comments, and knowledge contributions. By contrast, emotional engagement (being engaged) can be captured by a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication and absorption (Schaufeli et al. 2002). Here, the question is: *Do the game dynamics influence user engagement in a different manner depending on the different operationalizations of user engagement?*

Short-term vs. Long-term Engagement

One of the methodological issues regarding the measurement of user engagement is how the effects of game dynamics on user engagement vary over time. From the perspective of technology use, researchers

Game Dynamics	BE	EE
Getting PLBs	.32***	.29***
Competition	.22**	.19**
Self-expression	.17**	.33***

*<.05; **<.01; *** <.001

BE: Behavioral Engagement

EE: Emotional Engagement

Table 1. Summary of PLS analysis at the individual level

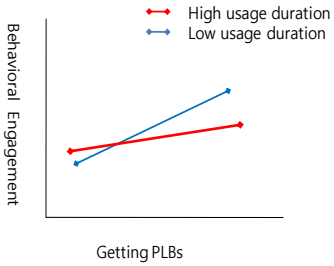

Figure 4. Moderating effect of usage duration

have found that the positive effects of extrinsic rewards on the use of technology tend to decrease over time (Magni et al. 2010). As users repeatedly interact with others and become used to the reward-based mechanisms within a gamified system, the users' sense of novelty and curiosity decreases, thus reducing the level of needs satisfaction (a main driver leading to intrinsic motivation). As time passes, users may develop a better understanding of the novelty, and the effects of reward-based mechanisms of gamification on needs satisfaction may diminish. Here, the question is: *Do the positive effects of game dynamics on user engagement decrease as usage duration increases?*

Individual-level vs. aggregated-level Engagement

While many enterprise gamified systems have been designed for facilitating employees' work-related activities, such as knowledge contribution, sales performance, and participatory activities in training program, research can measure user engagement at the individual and the aggregated levels. Here, aggregated level refers to group, collective, or organization. Even in case of group activities, such as group-based idea competition, assembly-based production, and crowd-based prediction, game dynamics on user engagement at the individual level can contribute to a better understanding of why and how game dynamics induce individual or group engagement. Therefore, multilevel theorizing cutting across individual and aggregate levels is important. The question is:

How different are the relationships between game dynamics and user engagement at the individual and aggregated levels?

Methods

To collect empirical data, this study has conducted a survey in a large global IT consulting company located in Seoul, Korea. The firm has adopted gamification system for the purpose of stimulating employees' motivation for knowledge sharing by transforming the electronic repository-based knowledge management system into an online community-based knowledge sharing system. The survey ended after 131 valid responses were gathered. After removing responses that contained unanswered items, 128 responses were used for the final analysis. In the entire data set, 69% of respondents were male, and 31% were female.

Results

The data analysis technique of partial least squares (PLS) was used for the analysis.

Behavioral vs. Emotional Engagement

The PLS analysis revealed that the relationships between game dynamics and user engagement were different depending on whether behavioral or emotional engagement was measured behavioral and emotional engagement. When behavioral engagement was used as a dependent variable, the game dynamics accounted for 34% of the variance existing in user engagement, and the getting PLBs was the most salient in increasing user engagement, followed by competition and self-expression. By contrast, when emotional engagement was used as a dependent variable, the game dynamics accounted for 48% of the variance of user engagement, and self-expression was the most salient factor in increasing user engagement, followed by getting PLBs and competition. Table 1 summarizes the results of PLS analyses.

Game Dynamics	BE	EE
Getting PLBs	.11*	.09*
Competition	.34***	.33***
Self-expression	.31***	.26***

*<.05; **<.01; *** <.001

BE: Behavioral Engagement

EE: Emotional Engagement

Table 2. Summary of PLS analysis at the team level

Short-term vs. Long-term Engagement

This study added the usage duration into the baseline model to examine how the effects of game dynamics on user engagement vary as time passes. The results show that the positive effect of getting PLBs on behavioral engagement decrease as usage duration increases. By contrast, the results show that self-expression and competition are not moderated by usage duration.

Individual-level vs. aggregated-level Engagement

The results demonstrate that the game dynamics exert different influences on user engagement at different levels. Table 2 shows the results of the PLS analysis at the aggregated level (the survey data was aggregated at the functional team). The results indicate that competition exerts the most salient influence on user engagement at the team level.

Discussion and Implications

To create theory-based knowledge regarding the effects of game dynamics on user engagement in the workplace, researchers need to operationalize and measure variables and explain how and why game dynamics X increases user behavior Y. To do so, operationalization of user engagement (as a dependent variable) is important in order for research to become more granular. The results of analyses show the importance of determining how to operationalize user engagement. The concept of user engagement is complex and contains many aspects. If researchers ignore the different facets of user experience, the empirical results of analyses may produce a partial or biased picture of game dynamics within enterprise gamified systems.

References

[1] Deterding, S. Gamification Absolved? (2014) http://gamification-research.org/2014/08/gamification-absolved-2/#more-918

[2] Harter, J. K., Schmidt, F. L., & Hayes, T. L. Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. Journal of Applied Psychology, 87, 2 (2002), 268-279.

[3] Kankanhalli, A., Taher, M., Cavusoglu, H., and Kim, S.H. Gamifciation: A new paradigm for online user engagement. International Conference on Information Systems. 2012.

[4] Kapp, K., Blair, L., and Mesch, R. The gamification of learning and instruction field book: Ideas into practice. Wiley, CA. 2013.

[5] Jacques, R., Preece, J., and Carey, T. Engagement as a design concept for multimedia. *Canadian Journal of Educational Communication*, 24, 1(1995), 49–59.

[6] Magni, M., Taylor, M. S., and Venkatesh, V. To play or not to play: A cross-temporal investigation using hedonic and instrumental perspective to explain user intentions to explore a technology. *International Journal of Human-Computer Studies*, 68, 9 (2010), 672-588.

[7] Schaufeli, W.B., Martínez, I.M., Pinto, A.M., Salanova, M., and Bakker, A.B. Burnout and engagement in university students: A cross-national study. *Journal of Cross-cultural Psychology*, 33,5 (2002), 464-481.

[8] Truss C, Alfes K, Delbridge R, Shantz A, Soane E. Employee Engagement in Theory and Practice, London: Routledge, 2013.

Moving Beyond the Effectiveness of Gamification

Rob van Roy
CUO | Social Spaces,
iMinds - KU Leuven
Parkstraat 45, bus 3605
3000 Leuven, Belgium
Rob.vanRoy@soc.kuleuven.be

Bieke Zaman
CUO | Social Spaces,
iMinds - KU Leuven
Parkstraat 45, bus 3605
3000 Leuven, Belgium
Bieke.Zaman@soc.kuleuven.be

Abstract
The recent booming popularity of gamification has incited researchers to investigate the effectiveness of this technique. However, by identifying different possible implications for both user and context, this paper wants to move beyond effectiveness and to elaborate on different ethical ramifications of the use of gamification. The paper concludes with formulating some guidelines for future research.

Author Keywords
Gamification; User; Context; Ethics; Motivation

ACM Classification Keywords
H.5.m [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces - User-Centered Design; K.4.1 [Computers and Society]: Public Policy Issues - Ethics; K.8.0 [Personal Computing]: General - Games

Introduction
The central tenet of gamification, "the use of game design elements in non-game contexts" (p. 10) [5], is the enhancement of people's motivation and engagement [15]. Therefore, gamification has become popular across different sectors ranging from marketing and informatics to politics, education and health [9,14]. Gartner has estimated that by the end of 2015 over 50% of the businesses will use gamification [6], leading to a projected gamification industry revenue of $1.5 billion [12]. In academia, research on gamification has also recently boomed, mainly with the goal to judge the effectiveness of gamification.

In this paper, we will argue that in order for the research field on gamification to mature, it is timely to move beyond the study of its effectiveness and reflect on how the implementation of gamification, in turn, can shape the broader context it is implemented in and the users involved. Though these implications for user and context are to date underexplored, they bring about important questions and ethical ramifications in a world where gamification can become omnipresent. By complementing existing criticisms of gamification with research findings that originate in other research fields, we will pinpoint concerns related to both the gamification context (including society) and its users. By articulating – sometimes provocative and far

stretched – statements, we want to draw attention to the need to reframe existing research questions, keeping the potential implications in mind, and to use new and innovative research designs in the study of gamification.

Implications for the context

Firstly, not every context is suited to play in. Starting from a very young age, people are socialized about the way they should act in different settings according to their cultural norms [7]. Nonetheless, by implementing gamification, users are asked to start playing, even in contexts where playing is culturally inappropriate. Deterding talks in this respect about the 'embarrassment' (p. 311) this mandatory play may bring forth [4]. Yet, in a gamified world, the discrepancy between the expectations of gamification and those on the basis of people's culture can have more far reaching implications. Notably, people might adapt their expectations of how to behave, transforming the existing culture into one of play and games. Furthermore, when, for example, the typical 'trial & error'-behavior of games [1] becomes part of a society's culture, people may also start to believe that they can try again when they failed on their first attempt, which is likely to become problematic in sectors like healthcare or jurisdiction.

Secondly, competitiveness, one of the most used features of games [18] in gamification [8], stimulates a struggle to be the best, if necessary even by cheating. Moreover, as winning automatically implies someone else's loss, it may promote a more selfishness-centered society [16] and discourage admirable characteristics like, for example, volunteer work or doing good for people.

Furthermore, this competitive context can harm the future of 'bad gamers', illustrating the interrelatedness of context and user. Although losers may set higher goals for the future, research has revealed that eventually they will perform systematically worse than their 'winning' counterparts, regardless of their pre-existing competences [2]. This example shows that gamification in this way can overshoot its goal, having far-reaching negative effects on the worst performing, and by extension the least motivated.

Implications for the user

As for the implications for the users involved in gamification, we will first discuss the implications for their intrinsic and extrinsic motivation. Intrinsic motivation is the human specific motivation that is inherent to and driven by the activity itself, whereas extrinsic motivation is caused by external factors independent from the activity such as rewards [17]. Intrinsic motivation has been found to outperform extrinsic motivation, having a more long-lasting influence on performance and leading to autonomous self-regulation [14]. Although implementing extrinsic motivation can stimulate people on the short term, it also has the potential to demolish existing intrinsic motivation [3,17], learning the users they should only perform the activity when rewarded [13]. Most gamified systems, however, rely on extrinsic motivational cues, by rewarding activities with badges or by encouraging competition. Consequently, by replacing the existing higher order intrinsic motivation with its extrinsic counterpart, gamification can potentially harm highly motivated people [10]. Furthermore, in the case of removal of the gamification elements and hereby also the corresponding extrinsic

motivation cues, we risk leaving unmotivated people behind.

Secondly, when gamification becomes omnipresent, questions arise about its utility. Different scholars have already claimed that the found positive effects of gamification can be attributed to a 'novelty effect' (e.g. [11]), stating that the effects are just temporarily, caused by the newness and accompanying excitement of the implemented system [10]. Furthermore, the omnipresence of gamification can speed up this process, transforming the newness of this technique into generality, removing the initial excitement. As a consequence, users will turn their backs on the by then boring gamification, resulting in the opposite of what it was implemented for.

Toward long-term, user- & context-centered research

In this paper, we pinpointed some of the possible negative implications gamification can have in society. However, although these considerations have a strong academic basis, specific research about these possible consequences lacks. Therefore, we evoke researchers to not only reframe their research questions paying attention to these implications, but also to use other research designs to tap in this underexposed subfield of gamification. Although detailed methodological guidelines go beyond the scope of this paper, we would like to highlight two important points:

Firstly, gamification research should broaden up, investigating beyond the basic outcome measures such as effectiveness. By looking at the influence of gamification on its users and context, interesting insights can originate, leading to a more nuanced understanding of the mechanics of this technique. Secondly, the use of a long-term perspective in this field becomes increasingly indispensable. In this way the possibility of a novelty effect can be scrutinized and the potential long-term implications on society listed in this paper can be examined.

Conclusion

To summarize, we argue that for the research field to mature, scholars should go beyond the effectiveness of gamification and explore other possible implications on both user and context by reframing their research questions and by using other research methods. This way, researchers can actively participate in the debate as to whether gamification should be used in multiple contexts, making it omnipresent, or should be limited to specific contexts for specific reasons only.

References

[1] Buckley, P. and Doyle, E. Gamification and Student Motivation. *Interactive Learning Environments 22*, 6 (2014), 1–14.

[2] Buser, T. *The Impact of Losing in a Competition on the Willingness to Seek Further Challenges.* Social Science Research Network, Rochester, NY, 2014.

[3] Deci, E.L., Koestner, R., and Ryan, R.M. Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research 71*, 1 (2001), 1–27.

[4] Deterding, S. Eudaimonic Design, or: Six Invitations to Rethink Gamification. In M. Fuchs, S. Fizek, P. Ruffino and N. Schrape, eds., *Rethinking Gamification*. Meson Press, Lüneburg, 2014, 305–323.

[5] Deterding, S., Dixon, D., Khaled, R., and Nacke, L. From Game Design Elements to Gamefulness: Defining "Gamification." *Proceedings of the 15th International Academic MindTrek Conference:*

Envisioning Future Media Environments, ACM (2011), 9–15.

[6] Gartner Group. Gartner Says By 2015, More Than 50 Percent of Organizations That Manage Innovation Processes Will Gamify Those Processes. http://www.gartner.com/newsroom/id/1629214.

[7] Giddens, A., Appelbaum, R.P., Duneier, M., and Carr, D. *Essentials of sociology.* W. W. Norton & Company, New York, 2013.

[8] Glover, I. Play as you learn: gamification as a technique for motivating learners. In J. Herrington, A. Couros and V. Irvine, eds., *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2013*. AACE, Chesapeake, VA, 2013, 1999–2008.

[9] Hamari, J., Koivisto, J., and Sarsa, H. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. *Proceedings of the 47th Hawaii International Conference on System Sciences*, (2014), 3025–3034.

[10] Hanus, M.D. and Fox, J. Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education 80*, (2015), 152–161.

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

[12] Konrad, A. Inside the Gamification Gold Rush. *Fortune*, 2011. http://fortune.com/2011/10/17/inside-the-gamification-gold-rush-2/.

[13] Muntean, C.I. Raising Engagement in E-learning Through Gamification. *Proceedings of the 6th International Conference on Virtual Learning ICVL*, ICVL (2011), 323–329.

[14] Richter, G., Raban, D.R., and Rafaeli, S. Studying Gamification: The Effect of Rewards and Incentives on Motivation. In T. Reiners and L.C. Wood, eds., *Gamification in Education and Business*. Springer International Publishing, 2015, 21–46.

[15] Rojas, D., Kapralos, B., and Dubrowski, A. The Missing Piece in the Gamification Puzzle. *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, ACM (2013), 135–138.

[16] Simon, R.L., Torres, C.R., and Hager, P.F. *Fair Play: The Ethics of Sport.* Westview Press, 2014.

[17] Tohidi, H. and Jabbari, M.M. The effects of motivation in education. *Procedia - Social and Behavioral Sciences 31*, (2012), 820–824.

[18] Vorderer, P., Hartmann, T., and Klimmt, C. Explaining the Enjoyment of Playing Video Games: The Role of Competition. *Proceedings of the Second International Conference on Entertainment Computing*, Carnegie Mellon University (2003), 1–9.

About the Authors

Rob van Roy is a **researcher** at CUO | Social Spaces, **iMinds – KU Leuven** (Belgium). Having graduated as a master in Communication Science at KU Leuven in 2013, he is now pursuing a PhD degree. Rob's research interests include new media, persuasive design, **education** and **gamification**.

Bieke Zaman is assistant **professor** in Human-Computer Interaction at the **KU Leuven - iMinds, Belgium**. She mainly coordinates research on **value-sensitive design, interaction design and children,** user experience evaluations and educational **games** design. Bieke is lecturing master courses on 'Human-computer Interaction', 'Usability Design', Qualitative Research, and Media Research & Innovation.