
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

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 the 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 condition, 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

games² 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])

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

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.

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/c11mjH>, May 2010. [last accessed 21/02/2015].
- [3] Hoorweg, 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.