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].

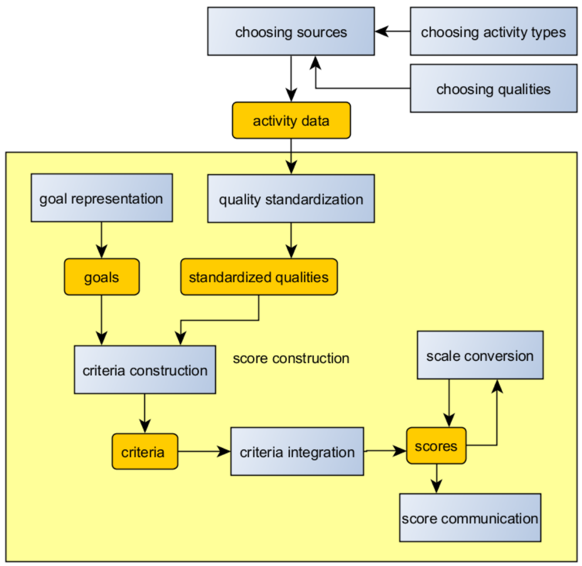

Figure 1: Process model of activity scoring. Blue boxes are processes, orange boxes are outputs.

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.

[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.

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 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.