Motivation in human computation task design

Anne M. Postma

Delft University of Technology
a.m.postma-1@student.tudelft.nl,

Abstract. There are no real systematic studies on motivational design in human computation tasks. However, motivating people to do tasks is essential for the success of your human computation project. This survey paper gives overview of different motivation aspects for human computation by reviewing motivational design done in the past. Furthermore, it hopes to give the reader insight into the effect of motivational aspects on task design by providing heuristics. Finally, it gives suggestions for further research in this field.

Keywords: human computation, task design, motivation, gamification

1 Introduction

In a world were computers are taking over an increasing amount of tasks, there are still tasks that remain too difficult or impossible for computers (e.g. identifying image content). People are often much better in solving these types of tasks. Human computation projects can facilitate this human computer interaction. In theory, everybody that owns a computer is able to do these tasks, however these tasks are often not directly beneficial for workers. Motivating people is therefore very important for the success of a human computation project.

Different papers have suggested that motivational design is becoming an important factor to indicate if a project is successful. [3, 14, 15, 19, 20]. Even so, Malone, Laubacher and Dellarocas [14] claim there are no real systematic studies on this issue. Therefore this survey paper gives an overview of different motivational solutions for human computation tasks that have been used in the past. Additionally, the paper hopes to give the reader insights into the effect of different motivational factors in such a way that the right ones can be chosen for implementation.

To begin with, this paper will go into detail of how a human computation task can be defined in Section 2. Section 3 presents different aspects of motivation that can be applied in human computation task design. Additionally, the issue that different types of workers will have different motivation is covered in Section 4. Finally, Section 5 shows future outlooks on human computation tasks and section 6 completes the paper with a conclusion and how this work relates to other disciplines.

2 What is a human computation task

Before going into motivational aspects within human computation task design, it is necessary to get an idea of what a human computation task might be.

There are many different definitions for human computation and no clear boundary of what is not. Quinn and Bederson[17] define it as *"systems of computers and large numbers of humans that work together in order to solve problems that could not be solved by either computers or humans alone"*. While Law and Ahn [12] call it *"... idea of using human effort to perform tasks that computers cannot yet perform, usually in an enjoyable manner"*. This already shows a difference in focus, one being on collaboration between a large number of people and the other on humans enjoying to support computers.

Relation with other fields. Quinn and Bederson [18], while building a taxonomy of the field, related human computation to other subjects. Namely, *collective intelligence*, *crowdsourcing* and *social computing*.

- *Crowdsourcing* is very much related to human computation in the sense that both ask a group of people to contribute to a certain task. They differ in a sense that *crowdsourcing* replaces work that is usually done by traditional human workers while human computation replaces work that cannot be done by computers.
- *Social computing* relates to human computation in a way that multiple people contribute to a certain job, however it only facilitates normal human behavior and does not aim towards specific human computation problems.
- The overarching notion of *collective intelligence* describes that large groups working together can accomplish great things. Human computation is one way to do this, but is not necessarily performed by a group of people, however involving a large group of people is still common within human computation.

Building blocks. Malone et al. [14] developed building blocks for collective intelligence system which is based on two pairs of questions *(1) Who is performing the task? Why are they doing it? (2) What is being accomplished? How is it being done?*. Within this paper the building block of *Why* people are doing a human computation task takes a central role. However the questions *Who*, *What* and *How* are very much related to this as these influence *Why* people will do a task.

Boundaries of human computation. A traditional human computation task is a relatively small task, e.g. tagging an image. This implies that its a computation a computer might be able to solve someday however is not able to do so yet [18]. The tasks are distributed over a large number of people and the human participation is directed by the computational system [18]. According to Reeves and Sherwood [19], the boundaries of what constitutes a human computation task is still fuzzy and new tasks and definitions are stretching those boundaries (e.g. citizen journalism).

Human computation platforms. To publish a human computation project a platform or your own environment can be used. Kittur et al. [10] recognizes a number of popular crowd work platforms for general-purpose markets and markets for specific expertise. The majority of the research done on traditional human computation tasks is done with Mechanical Turk (MTurk)[1]. Here, a requester can relatively easy publish human computation tasks to be done by workers of MTurk for a small amount of money.

3 Motivational aspects

There are different frameworks for human computation tasks design with regards to motivation. Quinn and Bederson[18] distinguish motivation as one dimension with five different sample values namely *Pay*, *Altruism*, *Enjoyment*, *Reputation* and *Implicit work*. While the *Why* building block for collective intelligence of Malone et al. [14] consists of three types of incentives: money, love and glory. Also many researchers [3, 6, 8, 10, 13–16] use the deviation of intrinsic and extrinsic motivation.

Based on the different frameworks and for the purpose of creating heuristics for future task design, a new differentiation is presented here. Namely *Financial payment*, *Gamification* and *Task goal*. The next chapters include a more detailed description of the three heuristics. To conclude, table 2 will provide overview of these heuristics.

3.1 Financial payment

The most common method to increase motivation is by making use of financial incentives, which the majority of the economic theories are based on. Mason and Watts [15] show that increased financial incentives led to an increase in quantity of tasks done. Next to that, Gneezy and Rustichini [7] found that giving more money also results in better performance. Shaw, Horton and Chen [20] found out that two types of financial incentives led to non-experts producing better results, the Bayesian Truth Serum (BTS) and the Punishment Agreement. Shaw et al [20] beliefs this can be caused by a combination of confusion and increase in cognitive demand within the task. Mason and Watts [15] also found a difference in results based on type of compensation scheme, a quota system resulted in better quality work than a piece-rate system, which in turn has been proven to be better than an hourly rate.

Still, Mason and Watts [15] did not find an increase in quality because of the anchoring effect. They even go so far as advising to pay as little as possible if you can make up for the decrease in quantity. But keep in mind that workers can feel insulted by the mere pennies they get [10].

Multiple researchers [7, 15] recommend using non-financial rewards as the quality of work will be as good or better than using financial rewards. Furthermore, Gneezy and Rustichini [7] claim that money can also have a negative

[1] www.mturk.com

effect on other intrinsic motivation as a participant will adjust their social contract when they receive money. Additionally, Quinn and Bederson [18] claim money will give more incentives to cheat the system which in turn results bad quality work.

Furthermore, Mason and Watts [15] propose the possibilities to develop a pay system which is tied to the quality that the participants delivers, similar to conventional work-related payment systems.

3.2 Gamification

"Gamification is an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience and user engagement."[5]. Deterding et al [5] claim gamification can be very useful in non-gaming contexts as well, in for example human computation tasks. However, it has become especially popular to turn a human computation task into a game in itself.

Games With A Purpose. The popular Games With A Purpose (GWAP) made their entry after von Ahn and Dabbish [1] developed their successful ESP game. Here, participants get points when they describe an image in a similar manner as their partner meanwhile giving the requestor descriptions of images. Goh and Lee [8] also indicated that games foster intention to use and can be equally motivating in collaboration games or competitive games.

Von Ahn and Dabbish [1] claim people play not because they are personally interested in solving an instance of a computational problem but because they wish to be entertained. They thereby fueled the discussion about the importance of enjoyment and fun in human computation task design [1].

Von Ahn and Dabbish [1] recognizes three types of human computation games, output- agreement games, inversion-problem games, and input-agreement games. However, they argue that it is important to not just stick a game-like interface on the task but to create an integrated system where game interaction and work coincide. According to Eickhoff et al. [6] this made current GWAPs highly tailored towards a certain (often niche) problem at hand and do not lend themselves for application across domains.

Malone [13] designed heuristics for enjoyable user interfaces including a Challenge (clear goal, uncertainty of reaching the goal), Fantasy (emotionally appealing, metaphors to relate to) and curiosity (informational complexity, knowledge structure) that could be applicable to human computation games in general. Game-elements used in human computation games are summarized in table 1.

Feedback systems. Feedback systems are often used in the form of reputation systems. However, Mumm and Mutlu [16] also researched subjective feedback (praise). They found out that both subjective (praise) and objective (leaderboard) feedback has a good influence of motivation. The two types of feedback even interact in a sense that intrinsic motivation was higher when subjective feedback was given, but only if objective feedback was not included and vice versa [16].

Table 1. common game-elements used in human computation games

Game element	Elaboration
Time	Generic timers for rounds of games are used. [1, 8, 9, 12]
Levels	Use levels to keep players curious. However, unlimited levels force players to finish the game by losing or giving up which resulted in less return to the game and less quality of work. [1, 6]
Scores	Scoring based on agreement with partners or the majority of players is used. [1, 6, 8, 9, 12]
Bonuses	There can be bonus rounds or additional scores when participants continue to give the right (agrees with other) answers. [2, 6, 12]
Collaboration	Is used to find agreement on a specific task among players. [1, 6, 8, 9, 12]
Competitiveness	Is used to motivate people directly and also forces participants to create more detailed input. [8, 9]
Taboo outputs	Forces participants to create more detailed input. [1, 8]
Reputation-systems	Allows users to compare their results with others. Eickhoff et al.suggests that integration with social networks will lead to even more compelling incentives and that multiplayer games should be researched. [1, 4, 6, 8, 12]
Randomness	To make the player curious and continue playing. [1, 6]

Mumm and Mutlu [16] argue that praise from a humanlike embodiment, even a rudimentary one, is better than none at all. Were false praise does not work in a classroom setting it is acceptable to use in the solitary nature of human computation [16]. However, they also acknowledge that praise could have a negative effect on intrinsic motivation when people start to rely on it too much [16].

Quality. In general there are a lot of success-stories of GWAP and other human computation games. The ESP game [2] resulted in 6 descriptions per image of which less than two percent had nothing to do with the image. Additionally a lot of participants returned to play the game after the first time. Also Pagehunt [6] proved to give better quality (accuracy), efficiency (time per task), incentive (continue without paying), consistency (agreement amongst players) and robustness (less cheating). However goh and lee [8] found that, while participants reported liking the collaborative and competitive human computation games over the control application, those using the latter seemed to generate better quality tags. Therefore quality checking remains very important in gamification of human computation.

Cheating is also a natural but negative side-effect of gamification as users try to score as many points as possible. Cheating should be avoided in the core of the game design. For example, in TagATune [12] players have to derive a ground truth from each-other which makes sure there is a natural incentive to be truthful which in turn resulted in more detailed descriptions. KissKissBan [9] provides

cheating-proof mechanism in the sense that players identify cheaters themselves which also led to more diverse labeling. Von Ahn and Dabbish [1] suggests to prevent cheating with random matching of players, player testing, repetition of task and using taboo outputs. Also Quinn and Bederson [18] suggest designing the task in such a way that it is not easier to cheat than to do the task.

3.3 Task goal

There seems to be a paradox between alienating task-goal from task or emphasizing the goal to increase motivation. Eickhoff et al.[6] noticed that it was important to split the game from the aftertaste of work which is also confirmed by the OnToGalaxy game by Krause et al.[11]. Von Ahn and Dabbish [1] and Ho et al. [9] even consider the human computation results generated as a side effect of the game play.

However, altruism as a motivational aspect [18] could not exist if the goal of the task is not known. An example of an altruistic task is helpfindjom.com where thousands of online volunteers tried to find Jim Gray is over 500,000 satellite images. Unfortunately Jim Gray was not found but it still demonstrated the power of altruistic motivation. Krause et al. [11] also argue preventing task identification hinders participants from questioning the use and effect of their actions which could result in less good results.

The size of the task plays a role in the sense that not noticeable work can generate very useful computation and dual purpose[18]. For example, reCAPTCHA[2] forces a user to solve a small task but also checks whether or not the user is human before they can continue on a website. Also participants playing games that do not need too much time commitment will result into playing it just one more time [1].

A task name can be confusing in itself and distract from the task at hand. Reeves and Sherwood [19] noticed that calling their human computation game eyespy resulted in people thinking it was related to a children's game where riddles instead of clear tags should be used as descriptions. von Ahn and Dabbish [1] also note that tasks that are both well-specified and challenging lead to higher levels of effort and task performance than goals that are too easy or vague.

Next to the task itself, there could be extras included for the participants not very much related to the task. For example TagATune [12] allowed for a get it functionality so that participants could download a song when they liked it, giving the task a dual purpose.

3.4 Overview

To provide better heuristics to implement during human computation task design, an overview is presented in table 2.

[2] blog.recaptcha.net

Table 2. motivational heuristics for human computation task design

Heuristic	**In regards to other frameworks**			**Conclusion**
	Quinn and Bederson [18]	Malone et al. [14]	Intrinsic and extrinsic	
Financial payment	Pay	Money	Extrinsic	Using no financial incentives produces similar quality as with financial incentives.
				More and specific payment could increase result.
				Money can diminish intrinsic motivation and can lead to cheating.
Gamification	Enjoyment	Love	Intrinsic	Gamification can work in non-gaming contexts as well.
				Gamification can be applied to turn a task into a game however game interaction and work should coincide.
				Malone [13] developed heuristics that could be applied on human computation games.
				Cheating should be avoided from the core of the human computation task design.
	Reputation	Glory	Extrinsic	Each form of feedback increases motivation even if it is false feedback.
				Praise from a humanlike embodiment is better than none at all.
Task goal	Altruism	Love	Intrinsic	Emphasizing a human computation tasks altruistic value works well.
	Implicit work	n/a	extrinsic	Splitting a human computation game from the aftertaste of work increases motivation.
	n/a (other aspects)	n/a	n/a	A task name and explanation should be clear.
				A human computation task could include extras.

4 Worker types

Within human computation task designs, the target audience should also be taken into account. Eickhoff et al. [6] distinguishes money-driven or entertainment-driven participants. Goh and lee's [8] results suggest that players with different motivations have different preferences for the types of human computation games they play. Eickhoff et al. [6] suggest to get better insight in workertypes to tailor a task design in a similar way advertisements do.

One way of researching the motivations of your target group is by doing a survey. Antin and Shaw [3] do suggest using surveys that are less vulnerable to social desirability bias. Namely, in their own research they found North-American citizens to over-reporting their motivation. Furthermore, Indian participants claimed to be motivated by sense of purpose but are more motivated by fun and killing time.

Additionally, the target groups knowledge can be taken into account to serve as a resource of knowledge to exploit [19]. For example a local dialect can be easiest translated by a person that speaks that specific dialect. This also motivates participants as they become aware of their added value.

An important thing to keep in mind is that workers are not the reason for poor quality work and that they are not lazy, stupid or deceitful as popular myths claim [10].

However, as mentioned before, cheating can still occur in human computation games, therefore malicious workers should be excluded. Quinn and Bederson [17] claim that the necessary means of discouraging cheating usually depends on the means of motivation used in the tasks design, figuring out this relation could prevent malicious workers on future human computation games.

5 Future of human computation tasks

There has been research on a future in which human computation competes with conventional work resulting in the task to become much more complex. Kittur et al. [10] emphasize more work needs to be done on designing tasks before this can be realized. They ask human computation platforms to make it easier and faster to create effective tasks in which they educate requestors on job design and the effects on quality. They also suggest more communication means with workers (e.g. trial run, direct communication to support synchronous collaboration). Also Kittur et al. [10] lay out research challenges in twelve major areas where motivation is still one of them.

In a popular TED talk [3] about incentives in the 21th century, Dan Pink claims that money will not be the most important incentive. Therefore it is important to research how motivation for human computation could work for more complex tasks. Von Ahn and Dabbish [1] also admit that the game templates for GWAP might not be optimal for creative tasks. They believe developing templates for such tasks to be an interesting research area.

[3] www.ted.com/talks/dan_pink_on_motivation.html

Within the field of Industrial Design Engineering, there have been design methods developed to structure a complex creative (design) problem in different smaller pieces, similar to human computation tasks. An example of such a design methods is the morphological map[4] where a product's parameters are listed and solved individually to each other and multiple times. This allows a designer to match different solutions together to create new products.

However, this evolution of human computation tasks also opens up the discussion of what a task should entail to be considered computation.

6 Conclusion

To conclude, there is still a lot of research needed and possible in relation to motivational design for human computation tasks. For example, the intrinsic *love* motivation of Malone et al. [14] is hard to grip into a scheme as it is often related to the individual worker and to the work that needs to be done and cannot be designed separately. Still, *love* motivation has proven its added value in the shape of enjoyment and altruism, thus making research in this field very valuable.

Additionally, there are still a lot of relating fields that are not often included in research about motivational design for human computation. More links could be drawn between (graphical) design knowledge. For example, certain color combinations tire the eyes of viewers more than others which could influence their time spend doing the task and their return to the task. Or personas, commonly used in industrial design, could help identify worker types.

Also, multiple researchers have indicated that more research should also be done outside the MTurk environment as this can lead to unrepresentative results of possible workers. Furthermore, workers are aware of the possibility of financial payment as motivational incentive which could also influence motivational research.

Still, this paper hopes to have provided overview of current motivational designs for human computation tasks and allows future requesters to design proper human computation tasks that are able to motivate their workers adequately.

References

1. Ahn, L.V., Dabbish, L.: Designing games with a purpose. ;Commun. ACM(2008)58-67
2. Ahn, L.V., Dabbish, L.: Labeling images with a computer game. ;In CHI(2004)319-326
3. Antin, J., Shaw, A.D.: Social desirability bias and self-reports of motivation: a study of amazon mechanical turk in the US and India. ;In CHI(2012)2925-2934
4. Dellarocas, C.: Online Reputation Systems: How to Design One That Does What You Need. ;MITSloan-51-3(spring 2010)33-38

[4] wikid.eu/index.php/Morphological_chart

5. Deterding, S., Sicart, M., Nacke, L., O'Hara, K., Dixon, D.: Gamification. using game-design elements in non-gaming contexts. In CHI Extended Abstracts(2011)2425-2428
6. Eickhoff, C., Harris, C.G., Vries, A.P.D., Srinivasan, P.: Quality through flow and immersion: gamifying crowdsourced relevance assessments. ;In SIGIR(2012)871-880
7. Gneezy, U., Rustichini, A.: Pay Enough or Don't Pay at All. ;The Quarterly Journal of Economics-115(2000)791-810
8. Goh, D.H., Lee, C.S.: Perceptions, quality and motivational needs in image tagging human computation games. ;J. Information Science(2011)515-531
9. Ho, C., Chang, T., Lee, J., Hsu, J.Y., Chen, K.: KissKissBan: a competitive human computation game for image annotation. ;In KDD Workshop on Human Computation(2009)11-14
10. Kittur, A., Nickerson, J.V., Bernstein, M.S., Gerber, E.M., Shaw, A., Zimmerman, J., Lease, M., Horton, J.J.: The future of Crowd Work. ;CSCW(2013)1301-1317
11. Krause, M., Takhtamysheva, A., Wittstock, M., Malaka, R.: Frontiers of a Paradigm - Exploring Human Computation with Digital Games. ;KDD-HCOMP(2010)22-25
12. Law, E., Ahn, L.V.: Input-agreement: a new mechanism for collecting data using human computation games. ;In CHI(2009)1197-1206
13. Malone, T.W.: Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games. ;CHI(1982)63-68
14. Malone, T.W., Laubacher, R., Dellarocas, C.N.: Harnessing Crowds: Mapping the Genome of Collective Intelligence. ;MIT Sloan Research Paper No. 4732-09 (2009). Available at http://ssrn.com/abstract=1381502
15. Mason, W., Watts, D.J.: Financial Incentives and the "Performance of Crowds". ;KDD-HCOMP(2009)77-85
16. Mumm, J., Mutlu, B.: Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. Computers in Human Behavior(2011)1643-1650
17. Quinn, A., Bederson, B.B.: A Taxonomy of Distributed Human Computation.: Tech. Rep. HCIL-2009-23, University of Maryland, (2009).
18. Quinn, A.J., Bederson, B.B.: Human computation: a survey and taxonomy of a growing field. ;In CHI(2011)1403-1412
19. Reeves, S., Sherwood, S.: Five design challenges for human computation. ;In NordiCHI(2010)383-392
20. Shaw, A.D., Horton, J.J., Chen, D.L.: Designing Incentives for Inexpert Human Raters. ;CSCW(2011)275-284