Context Disclosure as a Source of Player Motivation in Human Computation Games

ABSTRACT

Human computation games (HCGs) enables players to perform real-world tasks as a by-product of playing them. Since these tasks typically require domain knowledge, game designers map them onto simpler ones that non-experts can tackle. In many HCGs, most details on how the game mechanics abstract the task are left undisclosed, leaving players oblivious to how they are truly contributing. One outstanding question is whether revealing such details can promote player motivation. This paper presents a study carried out on *Amazon's Mechanical Turk* (AMT) workers using the *MATCHMAKERS* HCG. The study examined the impact of context disclosure at four levels of granularity: from zero context, the task's significance, the description of the task, to how the task is accomplished through game mechanics. Based on the study's findings, AMT Workers were most motivated when receiving the task's description. This work aims to provide insights into how context can be used to better motivate HCG players.

CCS CONCEPTS

• Human-centered computing ~ Empirical studies in HCI

Author Keywords

Player motivation; human computation games; games with a purpose; serious games.

INTRODUCTION

By empowering non-experts, crowdsourcing systems garner success in accomplishing tasks that are otherwise intractable [11]. However, these systems commonly suffer in their ability to attract and retain high-quality workers [10,13,19]. An approach to motivating workers that has lately gained traction is to turn the tasks into fun and engaging games (coined *human computation games* or HCGs), through a process known as *gamification* [1].

Since these tasks tend to involve domain knowledge, game designers deliberately map them onto tasks that non-experts can tackle (e.g., *Pipe Jam* [8]—software programs are represented by a network of pipes. Players derive a proof of correctness by aligning them). Often designers choose not to disclose details about the task, presumably because it is thought that players participate in HCGs for entertainment [1], and such contextual details would simply be uninteresting to them. Concealing these details, however, leaves players unaware of how they are truly supporting the underlying cause, and may be antithetical to an HCG's purpose of increasing worker motivation [7].

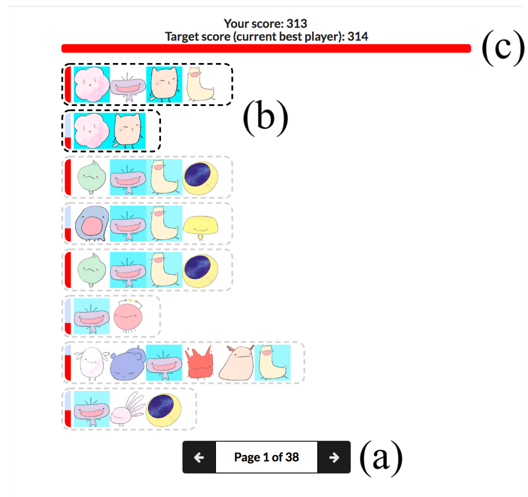


Figure 1. A screenshot of the *MATCHMAKERS* HCG. Players use the left and right arrow buttons (a) to browse through rows of characters, and group together those that are visually similar (b) to beat the target score (c).

It has been long upheld that humans are driven by their search for meaning [9]. People dedicate more time, exert more effort, and cede to less compensation in activities that are meaningful to them [16]. For example, in real world settings, prior studies have shown that knowing the underlying context of a task, such as the significance of a task, can motivate people to worker harder and longer [16]. Similar findings have been shown in crowdsourcing systems [4,15] and in HCGs [14].

Far from being the mere pleasure-seekers as previously thought [1], players do seem to care about the greater significance of their gameplay [7]. Given that players find context intrinsically rewarding, what details about the underlying task promotes player motivation? Players may not want all the details, as they can be complicated—

CONF '22, Jan 1 - Dec 31 2022, Authorberg.

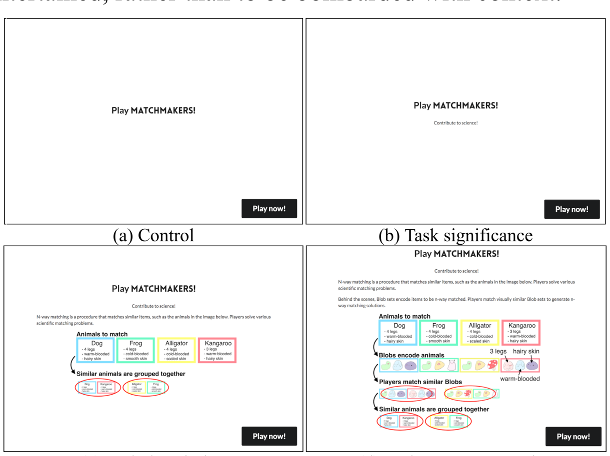
unique doi string will go here

demanding time and effort on the player to understand it, as well as on the game designer to craft together descriptions that are easy to grasp. In *Pipe Jam*, for example, it is not clear how aligning a network of pipes maps to deriving a proof of correctness [8].

Grounding much of the research on player motivation [3] is *self-determination theory (SDT)* [18], which contends that human behavior is shaped by a need to experience autonomy, competence, and relatedness in their activities. The degree to which a game can satisfy these needs is thought to govern its success [3]. Based on SDT, we reason that concealing from the player contextual details may undermine their sense of competence, and hence lessen player motivation.

In this work, we conducted a study to address the question of what contextual details of an HCG's underlying task might promote player motivation—as measured by player performance, attraction and retention. Four levels of granularity were considered: no context, the task's significance, the description of the task, and how the task is accomplished through game mechanics. The study was carried out on *Amazon's Mechanical Turk* (AMT) using *MATCHMAKERS* [5]—a web-based HCG for performing *n-way matching*, chosen because the concept of an n-way matching and its mapping to the game mechanics is particularly complicated. Briefly, n-way matching entails finding an optimal grouping of items from a collection of sets based on a given weight function [5,17]. The game abstracts n-way matching by having players group visually similar characters that encode items to-be-matched (Figure 1) [5].

The study's results indicate that providing the description of the underlying task motivated players the most, while disclosures of lesser context (i.e., the significance of the task) and more context (i.e., how the task is accomplished through game mechanics) did not seem to serve any greater motivational function. The takeaway from this is that there is middle ground between what to hide and what to reveal: while players appreciate having context for their activities, this comes second fiddle to their main motive of being entertained, rather than to be bombarded with context.



(a) Control (b) Task significance (c) Task description (d) Task-to-game mapping

Figure 2. The study's four experimental conditions.

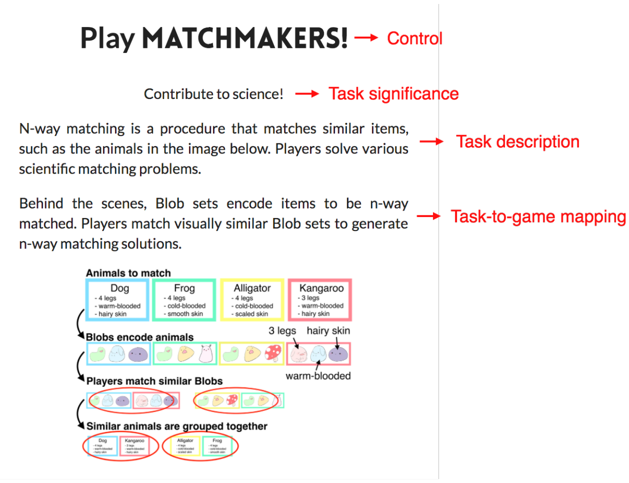


Figure 3. The text component of each experimental condition.

In sum, this work contributes to research on designing successful HCGs by providing insights into how context can be used to better engage players.

STUDY DESIGN

To address our research question, this study modulated the level of context that players received in *MATCHMAKERS*.

Experimental Conditions

The study considered the four experimental conditions below (**Error! Reference source not found.**):

Control. No context provided.
Task significance. The significance of the task is revealed, i.e., to contribute to science.
Task description. A description of the task is provided.
Task-to-game mapping. How the n-way matching task maps onto the game mechanics is revealed.

We presented the descriptions on the landing page of the game website as it is a natural place to provide introductory information. The descriptions were text-based (Figure 3**Error! Reference source not found.**) apart from the *task description* and *task-to-game mapping* conditions, both of which included a figure as the text on its own might be difficult to grasp. For consistency, all conditions were worded and presented identically; descriptions with more context simply appended additional text and figures onto those with less context (Figure 3**Error! Reference source not found.**). The descriptions underwent a pilot study and several revisions with a technical writer who had three years of experience, to help ensure that they were concise and articulate.

Participants

396 players were recruited on AMT. Of the 188 participants who self-reported their demographic information: 99 were female. The mean age was 35.3 (*SD=17.5)*. Players also reported the amount of time they spent playing digital games on a weekly basis (0-1 hours: 40, 1-2 hours: 42, 3-6 hours: 37, 7-10 hours: 31, 11-20 hours: 13, 21-40 hours: 12, 40+ hours: 13).

As the context given in the study's promotional materials can influence player motivation and interfere with the context that is disclosed in the game, it was crucial for promotional materials to be non-specific. If deployed elsewhere, such as on social media, a vague recruitment post might be treated as spam and fail to elicit enough participants (this was the case in our pilot study when it was launched on Facebook with the following post: "Play *MATCHMAKERS*: [link to game]!"). In contrast, recruitment posts on AMT can be non-specific since there is no expectation for *Human Intelligence Task* (HIT) requesters to provide context. To recruit players, we simply informed AMT workers that the HIT involved playing a game.

To lessen the impact of monetary incentives, we employed a payment mechanism commonly used by other studies on worker motivation [2,12]. AMT workers were given a base pay of \$0.45 for completing one round of the game and answering a post-study questionnaire, and a bonus of \$0.01 for each subsequent round. By keeping the bonus marginally low and fixed, motivation to continue playing is unlikely to be a result of monetary incentives.

Procedure

Players were randomly assigned to one of the four experimental conditions upon entering the game's landing page. We ensured that there were roughly an equal number of participants in each condition.

On the landing page, participants could choose to leave by exiting the website, or to play the game by clicking the "Play now!" button placed at the lower right corner of page. After pressing "Play now!", participants were presented with their first round of the game and a tutorial explaining the game's rules and objectives. After completing the tutorial, participants began playing the game (Figure 1). In each round of the game, players were to form a group of characters that would beat a specified target score, by matching characters that were visually similar. Upon beating the target score, a popup menu appeared providing players with a choice to quit or play another round of the game, either by continuing to improve the score of their current group of characters or to start afresh. If participants chose to quit the game, they were thanked for their participation and asked to fill out a post-study questionnaire which collected demographic information and an open-ended response about each player's motivation for playing the game.

Measures

Player motivation was evaluated quantitatively and qualitatively. For each player, the quantitative measures considered the following:

Number of rounds. The number of rounds completed.
Number of recompleted rounds. The number of times that the player attempted to improve their score on a round that they've that already completed, rather than starting afresh or quitting. This indicated whether players cared about improving their score on an existing round.
Total time. The total time spent playing the game.
Mean time per round. Mean time spent in each round.
Induced to play. Whether at least one round was attempted.
Induced to stay. Whether more than one round was attempted; that is, the player stayed beyond the minimum number of rounds required to complete the HIT.

To qualitatively assess player motivation, the post-study questionnaire that asked each player to give their reason for playing the game.

Analysis

Our analysis only included players who attempted at least one round of the game, apart from the i*nduced to play* measure.

Since the data was non-parametric, Kruskal-Wallis tests were run on all numerical measures to examine differences among conditions (*number of rounds, total time, number of recompleted rounds, mean time per round*) and Chi-squared tests were run on all Boolean measures (*induced to play, induced to stay)*. Where appropriate, all numerical measures were assessed using Bonferroni-corrected Wilcoxon Rank-Sum tests for post-hoc testing, and all Boolean measures used Bonferroni-corrected pairwise chi-square tests.

Moreover, the open-ended response as to why each player played the game was parsed automatically for keywords based on common themes that emerged.

Results

A total of 395 players visited the game's website over the course of the study. Of those 395 players, 16 players were removed from the analysis—fifteen due to inadvertently erasing a portion of the data records and one player whom we suspected was non-human because that player had visited the game's instruction page an abnormally large number of times (>100). In total, 379 players were analyzed. 95 players were placed in the *control*, *task description*, and *task-to-*

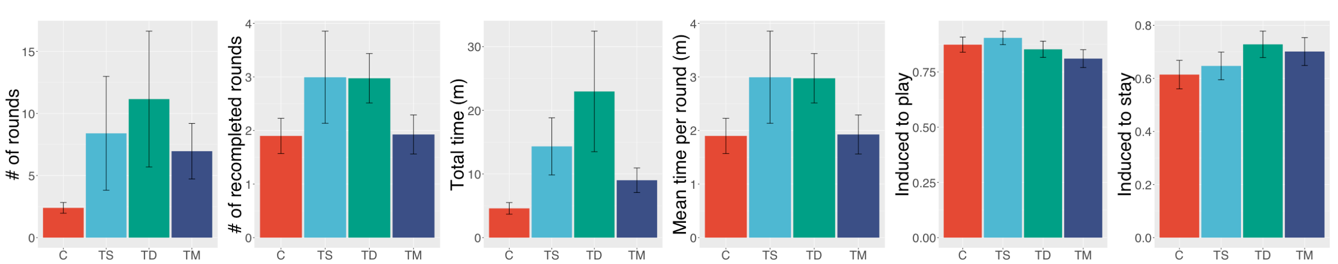


Figure 4. Mean and standard deviation graphs for each quantitative measure (C: *control*, TS: *task significance*, TD: *task description*, TM: *task-to-game mapping*).

Measure	Omnibus
Number of rounds	$H(3) = 10.06, p < .05$
Number of recompleted rounds	$H(3) = 9.82, p < .05$
Total time	$H(3) = 13.35, p < .005$
Mean time per round	$H(3) = 9.82, p < .05$
Induced to play	$\chi^2 = 3.65, n.s.$
Induced to stay	$\chi^2 = 2.95, n.s.$

Table 1. Results of the quantitative measures.

game mapping conditions, whereas 94 players were placed in the *task significance* condition.

Quantitative Results

Mean and standard deviation plots for each measure are presented in Figure 4 and a summary of the test results are presented in Table 1.

We observed a significant main effect for *number of rounds, number of recompleted rounds, total time,* and *mean time per round*. All significant differences were between the *control* and *task description* condition, with players in the *task description* condition outperforming those in the *control* condition. There were no significant effects found for the *induced to play* or *induced to stay* measures.

For *number of rounds*, players in the *task description* condition, on average, completed about four times more rounds of the game than those than those in the *control* condition (p<.01). The *number of recompleted rounds* for players in the *task description* was also significantly higher than those in the *control* condition (p<.05). Furthermore, these players also spent an average of four times longer playing the game than those in the *control* condition, based on *total time* (p<.005). For *mean time per round,* those in the *task description* condition spent longer within each round of the game, compared to those in the *control* condition (p<*.05*).

Qualitative Results

Of the 188 players who reported on why they played the game—via an open-ended question—most responses were themed around entertainment purposes (103/229, 45%), curiosity about the game (55/229, 24%), and monetary incentives (49/229, 21%). None of the respondents mentioned supporting the game's cause as a reason for playing. (Note that the total counts for the emergent themes are higher than the number of respondents since some players specified multiple reasons for playing).

DISCUSSION

Based on the findings, players were most motivated when they received the description of the task. That is, the *task description* condition was the only condition whose players significantly outperformed those in the *control* condition. We reason that this might be because the *task significance* condition was too vague to provide sufficient context regarding the cause that players were supporting, whereas the *task-to-mapping* condition contained so much detail that it might have been overwhelming. The *task description* condition likely served as a middle ground that resulted in the best player motivation.

On the surface, this seemingly contradicts previous findings that people derive pleasure from knowing the context of their activities. However, as the players of our study have reported, players chiefly play HCGs for entertainment (and monetary rewards, with respect to AMT), and are not inclined to grasp complicated concepts that are not directly pertinent to the gameplay.

LIMITATIONS AND FUTURE WORK

Some discretionary caution should be taken when interpreting these results. One shortcoming is that the study evaluated the impact of context on a single game. Given that all games are nuanced in ways that may motivate players differently (e.g., the entertainment factor, task significance, reward mechanisms, challenge, etc.), it is uncertain whether these results will still hold in other games. Moreover, our study framed context in a single dimension. Alternative presentation styles (e.g., a video clip) and approaches to framing context may influence players differently. Future work should examine how such factors might interplay with context disclosure.

Our final point on limitations is that the study was conducted only on AMT workers, who were all compensated for playing. Although our payment scheme was designed to lessen the impact of monetary rewards [2,12], whether these findings still pertain to other populations, such as unpaid players, remains to be explored. A practical issue in designing such a study how to recruit voluntary players while providing minimal context, since the framing of the promotional materials can potentially influence players and interfere with the context given in the game.

CONCLUSION

This paper examined how varying the level of context disclosed, from zero context to the task-to-game mapping, would influence player motivation. Given that designing successful HCGs is presently a challenging endeavor, a question of critical importance is how one can better engage players [6,20].

The insights here suggest that while players do appreciate receiving context for their gameplay, more context does not equate to being better. Rather, there is a happy medium between what to reveal and hide: too few details might lack sufficient context, and too many might overwhelm. Ultimately, the context provided by games should aim to be sweet and simple wherever possible, while still delivering enough detail to fulfil the human's search for meaning.

Finally, these findings are only representative of AMT workers. Whether they are still valid for other populations, such as unpaid players, remains to be explored.

REFERENCES

1. Luis von Ahn and Laura Dabbish. 2008. Designing Games With A Purpose. *Communications of the ACM* 51, 8: 58–67. https://doi.org/10.1145/1378704.1378719
2. Dan Ariely, Emir Kamenica, and Dražen Prelec. 2008. Man's search for meaning: The case of Legos. *Journal of Economic Behavior & Organization* 67: 671–677. https://doi.org/10.1016/j.jebo.2008.01.004

3. Florian Bruhlmann. 2013. Gamification From the Perspective of Self-Determination Theory and Flow. Open Science Framework. https://doi.org/none

4. Dana Chandler and Adam Kapelner. 2013. Breaking Monotony with Meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization* 90: 123–133. https://doi.org/10.1016/j.jebo.2013.03.003

5. Christina Chung, Asako Matsuoka, Yueti Yang, Julia Rubin, and Marsha Chechik. 2016. Serious games for NP-hard problems. In *Proceedings of the 5th International Workshop on Games and Software Engineering - GAS '16*, 29–32. https://doi.org/10.1145/2896958.2896963

6. Seth Cooper, David Baker, Zoran Popović, Adrien Treuille, Janos Barbero, Andrew Leaver-Fay, Kathleen Tuite, Firas Khatib, Alex Cho Snyder, Michael Beenen, and David Salesin. 2010. The challenge of designing scientific discovery games. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10*, 40–47. https://doi.org/10.1145/1822348.1822354

7. Sebastian Deterding. 2011. *Meaningful Play: getting "gamification" right*. Google Tech Talk.

8. Werner Dietl, Stephanie Dietzel, Michael D. Ernst, Nathaniel Mote, Brian Walker, Seth Cooper, Timothy Pavlik, and Zoran Popović. 2012. Verification games. In *Proceedings of the 14th Workshop on Formal Techniques for Java-like Programs - FTfJP '12*, 42–49. https://doi.org/10.1145/2318202.2318210

9. Viktor E. (Viktor Emil) Frankl. 2006. *Man's search for meaning*. Boston Beacon Press (1962).

10. Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13*, 1301. https://doi.org/10.1145/2441776.2441923

11. Edith Law and Luis von Ahn. 2011. Human Computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 5, 3: 1–121. https://doi.org/10.2200/S00371ED1V01Y201107AIM013

12. Edith Law, Ming Yin, Joslin Goh, Kevin Chen, Michael A. Terry, and Krzysztof Z. Gajos. 2016. Curiosity Killed the Cat, but Makes Crowdwork Better. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 4098–4110. https://doi.org/10.1145/2858036.2858144

13. Andrew Mao, Ece Kamar, and Eric Horvitz. 2013. Why Stop Now? Predicting Worker Engagement in Online Crowdsourcing. In *Proceedings of the 1st Conference on Human Computation and Crowdsourcing*.

14. Elisa D. Mekler, Florian Brühlmann, Klaus Opwis, and Alexandre N. Tuch. 2013. Disassembling gamification: The Effects of Points and Meaning on th User Motivation and Performance. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, 1137–1142. https://doi.org/10.1145/2468356.2468559

15. Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. 2011. An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 321–328.

16. Brent D. Rosso, Kathryn H. Dekas, and Amy Wrzesniewski. 2010. On the meaning of work: A theoretical integration and review. *Research in Organizational Behavior* 30: 91–127. https://doi.org/10.1016/J.RIOB.2010.09.001

17. Julia Rubin and Marsha Chechik. 2013. N-way model merging. In *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering - ESEC/FSE 2013*, 301–311. https://doi.org/10.1145/2491411.2491446

18. Richard M. Ryan and Edward L. Deci. 2000. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25, 1: 54–67. https://doi.org/10.1006/CEPS.1999.1020

19. Henry Sauermann and Chiara Franzoni. 2015. Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences of the United States of America* 112, 3: 679–84. https://doi.org/10.1073/pnas.1408907112

20. Kristin Siu and Mark O. Riedl. 2016. Reward Systems in Human Computation Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '16*, 266–275. https://doi.org/10.1145/2967934.2968083