

Evaluation of Game Designs for Human Computation

Julie Carranza, Markus Krause

University of California Santa Cruz, University of Bremen
jcarranz@ucsc.edu, phaeton@tzi.de

Abstract

In recent years various games have been developed to generate useful data for scientific and commercial purposes. Current human computation games are tailored around a task they aim to solve, adding game mechanics to conceal monotonous workflows. These gamification approaches, although providing valuable gaming experience, do not cover the wide range of experiences seen in digital games today. This work presents a new use for design concepts for human computation games and an evaluation of player experiences.

Introduction

Current approaches seen in human computation games are designed around a specific task. Designing from the perspective of the task to solve which leads to the game design, has recently been termed “gamification”. This approach results in games that offer a specific gaming experience, most often described as puzzle games by the participants. These games show the potential of the paradigm, but are still limited in terms of player experience. Compared to other current digital games their mechanics and dynamics are basic. Their design already applies interesting aesthetics but is still homogenous in terms of experience and emotional depth. Yet, to take advantage of a substantial fraction of the millions of hours spent playing, it is necessary to broaden the experiences that human computation games can offer. This would allow these games to reach new player audiences and may lead to the integration of human computation tasks into existing games or game concepts.

This paper introduces a new design concept for human computation games. The human computation game *OnToGalaxy* is drawn upon to illustrate the use of this concept. This game forms a contrast to other human computa-

tion games as an action oriented space shooter comparable to asteroid clones like *Starscape* (Moonpod, 2004). The design method presented will be explained along examples from the development of *OnToGalaxy*. A survey was conducted to compare *OnToGalaxy* and *ESP* in order to highlight the differences between more traditional human computation games and games that can be developed using the proposed design concept.

Related Work

Human computation games or *Games with a Purpose (GWAP)* exist with a range of designs. They can be puzzles, multi-player environments, or virtual worlds. They are mostly casual games, but more complex games do exist. Common tasks for human computation games are relation learning or resource labeling. A well-known example is the *GWAP* series (von Ahn, 2010). It consists of puzzle games for different purposes. *ESP* (von Ahn & Dabbish, 2004) for instance, aims at labeling images. The game pairs two users over the internet. It shows both players the same picture and lets them enter labels for the current image. If both players agree on a keyword, they both score and the next picture is shown. Other games of this series are *Verbosity* (von Ahn, Kedia, & Blum, 2006) that aims at collecting commonsense knowledge about words and *Squigle* (Law & von Ahn, 2009) or *Peekaboom* (von Ahn, Liu, & Blum, 2006), which both let players, identify parts of an image. All these games have common features such as pairing two players to verify the validity of the input and their puzzle like character.

Similar approaches are used to enhance web search engines. An instance for this task is *Page Hunt* (Ma, Chandrasekar, Quirk, & Gupta, 2009). The game aims to learn mappings from web pages to congruent queries. In particular, *Page Hunt* tries to elicit data from players about web pages to improve web search results. *Webparady* (Aras, Krause, Haller, & Malaka, 2010) is another game for website annotation. It aims at gathering natural language ques-

tions about given resources from a web page. The game is similar to the popular *Jeopardy* quiz. Other approaches rate search results. *Thumbs Up* (Dasdan, Drome, & Kolay, 2009), for example, utilizes digital game mechanics to improve result ordering. Another application domain of human computation is natural language processing (NLP). Different games try to use human computation games to extend NLP systems. *Phrase Detectives* (Chamberlain, Poesio, & Kruschwitz, 2008) for instance, aims at collecting anamorphic annotated corpora through a web game. Other projects use human computation games to build ontology's like *OntoGame* (Siorpaes & Hepp, 2007), or to test natural language processing systems as in the example of *Cyc Factory* (Cyc FACTory, 2010).

Most approaches to human computation games follow the gamification idea. *HeardIt* (Barrington, O'Malley, Turnbull, & Lanckriet, 2009) stresses user centered design as its core idea. This makes it different from other human computation games that are mostly designed around a specific task as their core element. *HeardIt* also allows for direct text interaction between players during the game session, which is commonly prohibited in other games in order to prevent cheating (von Ahn & Dabbish, 2004). *KissKissBan* (Ho, Chang, Lee, Hsu, & Chen, 2009) is another game with special elements. This game involves a direct conflict between players. One player tries to prevent the "kissing" of two other players by labeling images. Another game that has special design elements is *Plummings* (Terry et al., 2009). This game aims at reducing the critical path length of field programmable gate arrays (FPGA). Unlike other games, the task is separate from the game mechanics. The game is about a colony of so-called *Plummings* who need adequate air supply. By keeping the length of the air tubes as short as possible the player saves the colony from suffocation. Other games that are similar in nature are *Pebble It* (Cusack, 2010) and *Wildfire Wally* (Peck, Riolo, & Cusack, 2007). Unlike other human computation games, these are single player games because the quality of the input can be evaluated by an artificial system. Another interesting game is *FoldIt* (Bonetta, 2009). It is a single player game that presents simplified three-dimensional protein chains to the player, and provides a score according to the predicted quality of the folding done by the player. *FoldIt* is special because all actions by the player are performed in a three dimensional virtual world. Furthermore, the game is not as casual as other human computation games as it requires a lot training to solve open protein-puzzles.

To enhance current design concepts we propose to combine strengths of two existing frameworks. Hunicke, LeBlanc, and Zubek describe a formal framework of *Mechanics*, *Dynamics*, and *Aesthetics* (*MDA*) to provide guidelines on the design of digital games. *Mechanics* are the particu-

lar components of a game, at the level of data representation and algorithms. *Dynamics* define the run-time behavior of the *Mechanics* acting on player inputs and outputs over time. *Aesthetics* describe the desirable emotional responses evoked in the player, when he or she interacts with the game system (Hunicke, LeBlanc, & Zubek, 2004).

The *IEOM* framework (Krause & Smeddinck, 2012) describes core aspects of systems with a human in the loop. The four relevant aspects are *Identification*, *Observation*, *Evaluation*, and *Motivation*. *Identification* means finding a task or subtask which is easy for humans but hard for computers. *Observation* describes strategies to collect valuable data from the contributors by observing their actions. *Evaluation* addresses issues related to the validation of the collected data by introducing human evaluators or algorithmic strategies. The aspect of *Motivation* is concerned with the question of how a contributor can be motivated to invest cognitive effort. When games are the motivational factor for the system this aspect can be connected to *MDA*. The aspects from *IOME* give general parameters that have to be taken into account for *MDA*. These parameters can be expressed as *Mechanics*. The next Section gives an example how to use this combined framework.

OnToGalaxy

Designing a game for human computation is a complex task. To give an impression what such a design can look like, this section will present *OnToGalaxy*. *OnToGalaxy* is a fast-paced, action-oriented, science fiction game comparable to games like *Asteroids* (Atari, 1979) or *Starscape* (Moonpod, 2004). The aim of *OnToGalaxy* is to illustrate the potential of human computation games and to demonstrate the possibility of integrating human computation tasks into a variety of digital games. The section will describe the human computation game design of *OnToGalaxy*. The main human computation task in *OnToGalaxy* is relation mining. A relation is considered to be a triplet. This triplet consists of a subject, a predicate, and an object for instance: "go" (subject) is a "Synonym" (predicate) for "walk" (object). Relations of this form can be expressed in *RDF* (Resource Description Framework) notation. Many human computation tasks can be described in this form. Labeling tasks, like in the game *ESP* for instance, are easy to note in this way: "Image URI" (subject) "Tagged" (predicate) with "dog" (object).

Conceptual Framework

Since *OnToGalaxy* is a space shooter, it is not obvious how to integrate such a task. The previously described framework can be used to structure this process. The gamification approach is most common for human computation games. This approach starts with *identifying* a task to

solve, for instance labeling images. A possible way of *observation* would be to present images to contributors and let them enter the corresponding label. The *evaluation* method could be to pair two players and let them score if they type the same label. This process results in mechanics that are similar to those seen in ESP. Taking these mechanics *MDA* can reveal matching *Dynamics* and *Aesthetics* to find applicable game designs. An example of *Mechanic/Dynamic* pairing is “Friend or Foe” detection. The player has to destroy other spaceships only if they fit a certain criteria. The criterion in this case is the label that either describes the presented image or not. Using the two frameworks in this way, it is possible to describe requirements for the human computation task using *IEOM* and refine the design with *MDA*. Figure 1 shows these frameworks linked through the element of *Mechanics*.

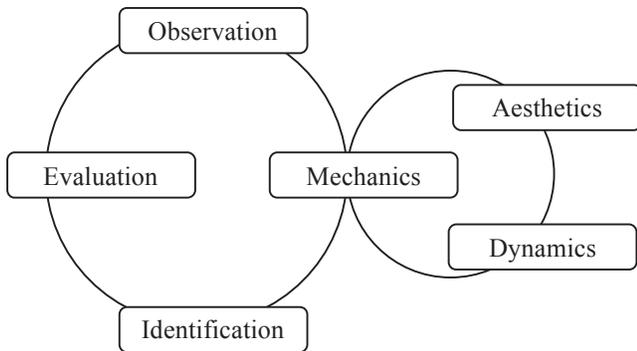


Figure 1: The combined workflow of *MDA* and *IEOM*. The aspects in *IEOM* define general parameters expressed in *Mechanics* to be considered in the game design process shaped by the *MDA* framework.

OnToGalaxy

Using the proposed concept, *OnToGalaxy* was designed as illustrated by the example above. He or she receives missions from his or her headquarters, represented by an artificial agent, which takes the role of a copilot. Alongside the storyline, these missions are presenting different tasks to the player like “Collect all freighter ships that have a callsign that is a synonym for the verb X.” A callsign is a text label attached to a freighter. If a player collects a ship according to the mission description, he or she agrees on the given relation. In the example this means, the collected label is a synonym for X. Three different tasks were explored with *OnToGalaxy*. All integrated tasks deal with contextual reasoning as described above. Every task was implemented in a different version of *OnToGalaxy*.

Ontology Population

The first task populates the *DOLCE* (Descriptive Ontology for Linguistic and Cognitive Engineering) with com-

mon words of the English language. The test set uses 5 base categories of the *DOLCE D18* (Masolo et al., 2001) and a corpus of 2189 common words of the English language. For each category 6 gold relations were marked by hand (3 valid and 3 invalid) to allow the evaluation algorithm to detect unwanted user behavior. The task performed by the player was to associate word A to a certain category C of the *DOLCE*. For instance the mission description could ask the following question: is “House” a “Physical Object,” where “House” is an element of the corpus and “Physical Object” a category of *D18*. The mission description included the category of *DOLCE* in this case “Physical Object”. The freighters labels were randomly chosen from the corpus of words in this case “House”. To test player behavior a set of gold data was generated for each category. A typical mission could be “Collect all freighters with a label that is a (Physical Object)”. Figure 2 shows a screenshot of this first version of *OnToGalaxy*. 32 players played a total amount of 26 hours under lab conditions, classifying 552 words with 1393 votes. Thus, each word required 2.8 minutes of game play to be classified.



Figure 2: Ontology population version of *OnToGalaxy*. The blue ship on the right is a labeled freighter. The red ship in the middle is controlled by the player. On the lower right is the mission description. It tells the player to collect freighters labeled with words of “touchable objects” such as “hammer”.

Synonym Extraction

The second task was designed to aggregate synonyms for German verbs. The gold standard was extracted from the *Duden* synonym dictionary (Gardt, 2008). The test was composed of three different groups of verbs. The first group A was chosen randomly. The second group B was selected by hand and consisted of synonyms for group A. Thus each word in A had a synonym in B. The third group was composed of words that were not synonym to any

word in A or B to validate false negatives as well as false positives. Gold relations were added automatically by pairing every verb with itself. In the mission description one verb from of the corpus of A, B, and C was chosen randomly. The mission description tells the player to collect all freighters with a label being a synonym for the selected verb and to destroy the remaining ships. The freighters were labeled with random verbs from the corpus. Figure 3 shows a screenshot of this second version of *OnToGalaxy*.

25 players classified 360 (not including gold relations) possible synonym relations under lab conditions. The players played 14 hours in total. Each synonym took 2.3 minutes of game play to be verified. Even though the overall precision was high (0.97) the game had a high rate of false positives (0.25). A reason for that can be a misconception in the game design. The number of labels that did not fulfill the relation presented to the player was occasionally very high - up to 9 out of 10. Informal observations of players indicate that this leads to higher error rates than more equal distributions. Uneven distributions of valid and invalid labels were also reported as unpleasant by players. Again informal observations point to a distribution of 1 out of 4 to be acceptable for the players.



Figure 3: The mission in the lower mid tells the player to collect all freighters (blue ships) with a label being a synonym for “knicken” (to break). The female character on the bottom gives necessary instructions. The player is controlling the red ship in the center of the screen.

Finding Evocations

The third task that was integrated into *OnToGalaxy* was retrieving arbitrary associations sometimes called evocation as used by Boyd-Graber (Boyd-Graber, Fellbaum, Osherson, & Schapire, 2006). An evocation is a relation between two concepts and whether one concept brings to mind the other. For instance the term “green” can bring to

mind the term “grass”. The game presents two terms from a collection of 1000 words to the player to decide whether these terms have such a relation or not. Therefore a total of one million relations have to be checked. This collection was not preprocessed to collect human judgment for every relation. All possible evocations are presented to players, as there are no other mechanisms involved to exclude certain combinations. The game attracted around 500 players in the first 10 hours of its release. Figure 4 shows a screenshot of this third version of *OnToGalaxy*.



Figure 4: The mission in the lower center tells the player to collect the highlighted vessel if the label “pollution” is related to the term “safari”.

Experiment

As a pre-study the third version of *OnToGalaxy* was compared to *ESP*. *ESP* was chosen because many human computation games are using *ESP* as a basis for their own design. The 6 participants of the study were watching a video sequence of the game play of *OnToGalaxy* and *ESP*. The video showed both games at the same time and the participants were allowed to play, stop and rewind the video as desired. Watching the video the participants gave their impression on the differences between both games. During these sessions 69 comments on both games were collected. *ESP* was commented 24 times (0.39) and *OnToGalaxy* 45 times (0.61). This pre-study already indicated the differences in aesthetics. The final study was conducted with a further refined version of *OnToGalaxy*. Again a group of 20 participants watched game play videos of *OnToGalaxy* and *ESP*. As we were most concerned with perception of aesthetics, we chose not to alter the pre-study structure and have participants watch videos of game play for each of the games.

Using the *MDA* framework as a guideline, we set about creating a quantitative method of measuring the aesthetic

	Social	Objective	Realistic	Self-discovery	Imaginative	Challenging	Emotional	Entertaining	Exploration	Story	Visual
min	-3.61	-2.69	-1.99	-1.57	-0.83	-0.52	-0.67	-0.20	0.65	0.36	0.98
max	-1.99	-0.21	-0.11	0.07	1.23	0.92	1.07	1.90	2.05	2.44	2.58
mean	-2.8	-1.45	-1.05	-0.75	0.2	0.2	0.2	0.85	1.35	1.4	1.85

Table 1: This table shows the confidence intervals from the mean differences of each question. The abbreviations used are explained in Table 2. Negative values indicate a stronger association with ESP positive values with *OnToGalaxy*.

qualities of a game. Specifically we wanted to examine what aspects of aesthetics set *OnToGalaxy* apart from *ESP*. Using common psychological research methods, we created a survey that examined more closely the sensory pleasures, social aspects, exploration, challenge, and game fulfillment of these two games. Hunnicke et al. put forth eight taxonomies of *Aesthetics*— *Sensation*, *Fantasy*, *Narrative*, *Challenge*, *Fellowship*, *Discovery*, *Expression*, and *Submission* - that were used as a guideline in stimuli creation. Table 2 lists resulting questions, as well as a shorthand reference to these. “Visual” was our only question relevant to *Sensation*. As participants watching video clips of gameplay that did not have sound, the questions on sensory pleasure were related only to visual components. For some of the more subjective *Aesthetics*, we chose to use different questions to get at the specificity of a particular dimension – for example *Fantasy*. A game can be seen as having elements of both fantasy and reality. Asking whether a game is imaginative or realistic can give a clearer picture of what components of fantasy are present in the game. With respect to *Narrative*, we chose to see whether participants felt there was a story being followed in either game, as well as a clearly defined objective. *Challenge* and *Discovery* (examined by the “Challenge” and “Exploration” questions respectively) were more straightforward in our particular experiment as *OnToGalaxy* and *ESP* are both casual games. *Fellowship* was examined by the question of how “Social” the game was perceived to be. *Expression* was addressed through the question of “Self Discovery.” Finally we chose to look at *Submission* (articulated as “game as pastime” by Hunnicke) using the questions “Emotional” and “Entertaining.” Our thoughts for this decision came from our conception of what makes something a pastime for a person.

After creating questions that addressed the eight taxonomies of *Aesthetics*, we chose to use a seven-point Likert scale to give numerical value to these different dimensions. On the Likert scale a 1 corresponded to “Strongly Disagree,” a 4 corresponded to “Neutral,” and a 7 corresponded to “Strongly Agree”. The scales use seven points for two reasons. The first is that there is a larger spread of values, making the scales more sensitive to the participants’ responses; the second reason was to omit forced choices. Participants were either filling out an online survey or sending in the questionnaire via e-mail. The survey con-

sisted of an open-ended question (“Please describe the differences you can observe between these two games”) and two sets of the Likert questions - one for *OnToGalaxy* and one for *ESP*. The online version did randomize the order of the questions. From the previous studies, a relatively large effect size was expected, therefore only a small number of participants were necessary. A power analysis gave a minimum sample size of 13 given an expected Cohen’s *d* value of 1.0 and an alpha value of 0.05.

Question	Abbreviation
I felt the game was visually pleasing	Visual
I felt this game was imaginative	Imaginative
I felt this game was very realistic	Realistic
I felt this game followed a story	Story
I felt this game had a clear objective	Objective
I felt this game was challenging	Challenge
I felt this game had a social element	Social
I felt this game creates emotional investment for the player	Emotional
I felt this game allowed the player to explore the digital world	Explorative
I felt this game allowed for self-discovery	Self-Discovery
I felt this game was entertaining	Entertaining

Table 2: This table shows the questions from the questionnaire and its related abbreviation.

Results

Dependent two-tailed t-tests with an alpha level of 0.05 were conducted to evaluate whether one game differed significantly in any category. For five out of ten questions the average performance (score out of 7) are significantly different, higher or lower. Three categories are more closely associated to *OnToGalaxy*. For the question “Visual” *OnToGalaxy* has a significantly higher average score (M=5.1, SD=1.26) than *ESP* (M=3.25, SD=1.41), with $t(19)=4.56$, $p<0.001$, $d=1.39$. A 95% CI [0.98, 2.58] was calculated for the mean difference between the two games.

For the question “Story” the results for *OnToGalaxy* (M=3.35, SD=1.89) and *ESP* (M=1.95, SD=1.31) are $t(19)=2.61, p=0.016, d=0.83, 95\% \text{ CI } [0.36, 2.44]$. For the question “Exploration” the results for *OnToGalaxy* (M=3.75, SD=1.68) and *ESP* (M=2.4, SD=1.27) are $t(19)=3.77, p=0.001, d=0.88, 95\% \text{ CI } [0.65, 2.05]$.

Some categories are more associated to *ESP*. For the question “Social” *ESP* has a significantly higher average score (M=4.85, SD=1.09) than *OnToGalaxy* (M=2.05, SD=1.23), with $t(19)=6.76, p<0.001, d=2.47$. A 95% CI [3.61, 1.99] was calculated for the mean difference between the two games. For the question “Objective” the results for *ESP* (M=5.10, SD=1.87) and *OnToGalaxy* (M=3.65, SD=1.56) are $t(19)=2.28, p=0.03, d=0.85, 95\% \text{ CI } [2.69, 0.21]$. Table 1 shows all measured confidence intervals including those not reported in detail. The values of the question “Realistic” are not reported in detail as at least 5 participants pointed out they misunderstood the question.

□

Predicting Game Rated from Questionnaire Answers

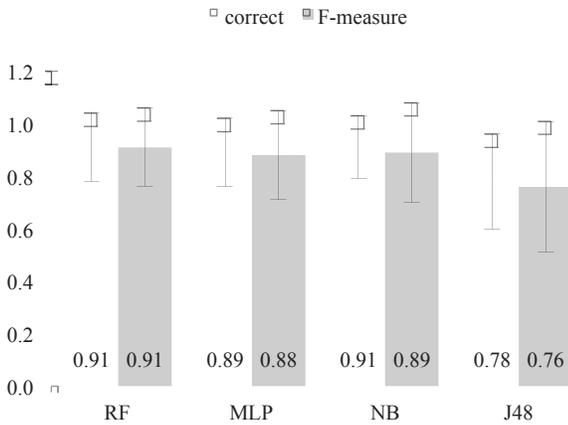


Figure 5: Results of predicting which game was rated given a vector of the Likert scale part of the questionnaire. Based on the F-measure and correctness values, the random forest (RF) performed best, the feed forward neuronal network (MLP) and Naïve Bayes (NB) achieved similar quality. The decision tree (J48) performed not as well but not significantly, given an alpha value of 0.05.

Non-standardized questionnaires, as the one used in the experiment, cannot be assumed to give reliable results. As statistical measures only detect differences between means it is still unclear whether individual inputs are characteristic given a certain game. To investigate this issue another experiment was conducted. If the questionnaire is indeed suitable to identify differences between the two games, a machine learning system should be able to predict which game the participant rated given the answers to the ques-

tionnaire. To test this hypothesis a data set from all questionnaires was generated. The data set contains the answers from 20 participants. Each participant rated both games giving a total of 40 feature vectors in the data set. Each vector consisted of 11 values ranging from 1 to 7 corresponding to the questions from Table 2. Each vector had associated the game which was rated as its respective class.

Four common machine learning algorithms were chosen to predict which game was rated. The following implementations of the algorithms were used as part of the evaluation process: Naïve Bayes (NB), a decision tree (C4.5 algorithm) (J45), a feedforward neuronal network (Multi-Layer Perceptron) (MLP) and a random forest as described by (Breiman, 2001) (RF). Standard parameters¹ were used for all algorithms. Figure 5 shows the correctness rates and F-measures achieved. The results indicate that it is indeed possible to identify differences between the two games as the correctness rates and F-measures are high. Given a similar dataset we also tried to predict from a given feature vector which game a participants would prefer. None of the algorithms had acceptable classification rates for this experiment.

Conclusion

The diversity of digital games is, in their entirety, not reflected in current human computation games - though they provide valuable gaming experience. As this paper demonstrates it is possible to integrate human computation into games differently from current designs. As the results of the presented surveys indicate, *OnToGalaxy* is perceived significantly different from “classic” human computation game designs, and performs equally well in gathering data for human computation. The presented design concept using *IEOM* and *MDA* is able to support the design process of human computation games. The concept allows designing games in a way similar to other modern games still keeping the necessary task as a vital element. The paper also gives indication that using the proposed design concepts significantly different games can be built. This allows for more versatile games that can reach more varied player groups than those of other human computation games.

References

Aras, H., Krause, M., Haller, A., & Malaka, R. (2010). Webpardy: Harvesting QA by HC. *HComp'10 Proceedings of the ACM SIGKDD Workshop on Human Computation* (p. 4). Washington, DC, USA: ACM Press.

¹ The standard parameters were derived from Weka Explorer 3.7 (Frank et al., 2010)

- Atari. (1979). Asteroids. Atari.
- Barrington, L., O'Malley, D., Turnbull, D., & Lanckriet, G. (2009). User-centered design of a social game to tag music. *HComp'09 Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 7-10). Paris, France: ACM Press.
- Bonetta, L. (2009). New Citizens for the Life Sciences. *Cell*, 138(6), 1043-1045. Elsevier.
- Boyd-Graber, J., Fellbaum, C., Osherson, D., & Schapire, R. (2006). Adding dense, weighted connections to WordNet. *Proceedings of the Third International WordNet Conference* (pp. 29-36). Citeseer.
- Breiman, L. (2001). Random Forests. *Machine learning*, 45(1), 5-32.
- Chamberlain, J., Poesio, M., & Kruschwitz, U. (2008). Phrase Detectives: A Web-based Collaborative Annotation Game. *I-Semantics'08 Proceedings of the International Conference on Semantic Systems*. Graz: ACM Press.
- Cusack, C. A. (2010). Pebble It. Retrieved from
- Cyc FACTory. (2010). game.cyc.com. Retrieved from
- Dasdan, A., Drome, C., & Kolay, S. (2009). Thumbs-up: A game for playing to rank search results. *WWW '09 Proceedings of the 18th international conference on World wide web* (pp. 1071-1072). New York, NY, USA: ACM Press.
- Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., & Trigg, L. (2010). Weka-A Machine Learning Workbench for Data Mining. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 1269-1277). Boston, MA: Springer US. doi:10.1007/978-0-387-09823-4_66
- Gardt, A. (2008). *Duden 08. Das Synonymwörterbuch* (p. 1104). Mannheim, Germany: Bibliographisches Institut.
- Ho, C.-J., Chang, T.-H., Lee, J.-C., Hsu, J. Y.-jen, & Chen, K.-T. (2009). KissKissBan: a competitive human computation game for image annotation. *HComp'09 Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 11-14). Paris, France: ACM Press.
- Hunicke, R., LeBlanc, M., and Zubek, R. 2004. MDA: A formal approach to game design and game research. In *Proceedings of the Challenges in Game AI Workshop, 19th National Conference on Artificial Intelligence (AAAI '04, San Jose, CA)*, AAAI Press.
- Krause, M., & Smeddinck, J. (2012). *Handbook of Research on Serious Games as Educational, Business and Research Tools*. (M. M. Cruz-Cunha, Ed.). IGI Global. doi:10.4018/978-1-4666-0149-9
- Law, E., & von Ahn, L. (2009). Input-agreement: a new mechanism for collecting data using human computation games. *CHI '09 Proceedings of the 27th international conference on Human factors in computing systems* (pp. 1197-1206). Boston, MA, USA: ACM Press.
- Ma, H., Chandrasekar, R., Quirk, C., & Gupta, A. (2009). Page Hunt: using human computation games to improve web search. *HComp'09 Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 27-28). Paris, France: ACM Press.
- Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., & Horrocks, I. (2001). *WonderWeb Deliverable D18 Ontology Library(final)*. Communities (p. 343). Trento, Italy.
- Moonpod. (2004). Starscape. Moonpod.
- Peck, E., Riolo, M., & Cusack, C. A. (2007). Wildfire Wally: A Volunteer Computing Game. *FuturePlay 2007* (pp. 241-242).
- Siorpaes, K., & Hepp, M. (2007). OntoGame: Towards overcoming the incentive bottleneck in ontology building. *OTM'07 On the Move to Meaningful Internet Systems 2007: Workshops* (pp. 1222-1232). Springer.
- Terry, L., Roitch, V., Tufail, S., Singh, K., Taraq, O., Luk, W., & Jamieson, P. (2009). Harnessing Human Computation Cycles for the FPGA Placement Problem. In T. P. Plaks (Ed.), *ERSA* (pp. 188-194). CSREA Press.
- von Ahn, L. (2010). www.gwap.com (Games with a Purpose). Retrieved from
- von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. *CHI '04 Proceedings of the 22nd international conference on Human factors in computing systems* (pp. 319-326). Vienna, Austria: ACM Press.
- von Ahn, L., Kedia, M., & Blum, M. (2006). Verbosity: a Game for Collecting Common-Sense Facts. *CHI '06 Proceedings of the 24th international conference on Human factors in computing systems* (p. 78). Montréal, Québec, Canada: ACM Press.
- von Ahn, L., Liu, R., & Blum, M. (2006). Peekaboom: a game for locating objects in images. *CHI '06 Proceedings of the 24th international conference on Human factors in computing systems* (pp. 64-73). Montréal, Québec, Canada: ACM Press.