GWAPs: Games with a Problem

Kathleen Tuite
Department of Computer Science and Engineering
University of Washington
ktuite@cs.washington.edu

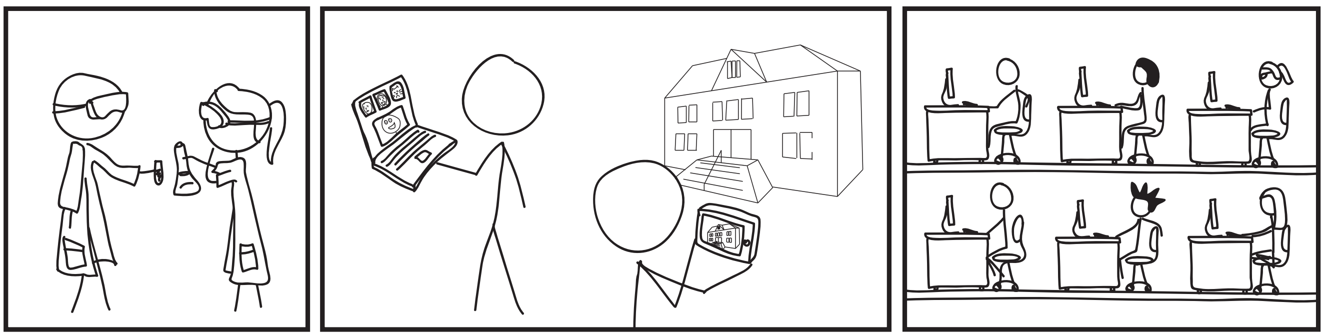

Figure 1: Games with a purpose (center) should fit between skilled professions (left) and simplistic micro-work that uses people as *human-processing units* (right). Games with a purpose should not seek to 'gamify' uninteresting tasks that are already easily crowdsourced, but to facilitate complex, collaborative, and creative and skilled work.

ABSTRACT
Games with a purpose (gwaps) are expensive and challenging to create, yet are touted as a cure-all for getting people to do uninteresting and repetitive tasks in exchange for fun and entertainment. Not only are poorly-executed gwaps ineffective for their purposes, but they can potentially undermine future gwaps by burning out players and discouraging potential gwap supporters and investors. The goal of this paper is to challenge the role of gwaps as tools for disguising boring work as entertainment. We discuss recommendations for future gwap designs that are informed both by commercial game design and by parallel developments in paid crowd work.

Keywords
games with a purpose, crowdsourcing, gamification

1. INTRODUCTION
Games with a purpose (gwaps) are a type of game in which the player's actions in the game contribute to a real-world purpose outside of the game, whether it be predicting protein structures or providing labels for images.

Many gwaps have been developed since the first "GWAP" [34], the ESP Game [33] in 2004. But gwaps don't seemed to have lived up to the initial hype of transforming millions of hours typically poured into traditional games into useful and productive work. Even after the initial success of the ESP Game, it was shut down in 2011 [15]. The standard way to accomplish the same type of work is to crowdsource the work directly using a service like Mechanical Turk [10].

It is time to reflect on how gwaps have evolved in the past decade. We believe gwaps are still too commonly patterned off a fixed notion of what gwaps should be like, one that stems from von Ahn positioning his games as a way to accomplish work by providing entertainment instead of monetary compensation. The more fun and addicting the game is, according to this notion, the more work players will be tricked into doing for free.

This notion likely stems from von Ahn's 2008 article in which he stated, "People play not because they are personally interested in solving an instance of a computational problem but because they wish to be entertained." This could be interpreted literally as an observation that ESP Game players are probably more interested in the tag-matching mechanic and guessing the same answer as another player than they are in providing labels for images. Or it could be interpreted, as mentioned above, that if you make a task entertaining, people will do it; that the sole purpose of gwaps is to find the "fun" in otherwise tedious and uninteresting tasks.

Games that aim to provide only entertainment in exchange for work risk seeming exploitative, even if that was not the designer's intentions. Even just the perception of gwaps as potentially exploitative and shallow is damaging to the field, as it can drive players away and limit possibilities for people to build new games or to do research into designing better games.

We believe that games with a purpose still have potential to change the world, but we need to pause and reflect on what has made only certain gwaps successful and effective thus far.

The *problem* gwaps face is a potential future where they are interpreted too narrowly, leading to shallow and deceptive experiences that provide little meaningful alternative to the kind of paid micro-work that is being explored (and recently questioned) in the realm of crowdsourcing.

Crowdsourcing experts remain suspicious of gwaps: "While dressing up such tasks as games may reduce boredom, entertainment represents a fairly superficial form of work satisfaction." [22] If it is a task such as labeling images, then yes, it would be better to directly pay someone to do the work. But if it is a complex task requiring skill and a good understanding of the situation (crowdsourcing tasks rarely provide this context) then a game approach could be more appropriate. Additionally, it is tricky to use money to incentivize and reward things like creative problem solving and ingenuity, so a game could be an ideal platform.

The goal of this paper is to challenge the role of gwaps. Instead of compensating boring work with with entertainment, they should be used to structure complex tasks, facilitate skilled work, and allow players to make meaningful and recognizable contributions. Figure 1 illustrates the direction in which gwaps should evolve: away from micro-work and towards rich, complex, skilled tasks. In the following sections, we will make recommendations (based on observations of deployed gwaps and human computation systems) that we hope will secure a future for effective, deep, and engaging gwaps.

2. BACKGROUND

Since games with a purpose are a form of crowdsourcing, this section will cover the original gwaps and what tasks they set out to crowdsource, as well as related ideas from crowdsourcing.

2.1 The first Games with a Purpose

Luis von Ahn and Laura Dabbish kicked off the genre of "games with a purpose" [34] in 2004 with the first GWAP, the ESP Game [33]. The ESP Game paired two anonymous players online, asked them to provide labels for an image they both saw, and rewarded both players for agreeing on labels. The purpose of the game was to provide labels for images, labels such as type of scene or what objects are present, exactly the kinds of common and recognizable attributes that a player would expect their anonymous partner to also recognize. The ESP Game was a specific example of an *output-agreement game*, one of three game templates also including *input-agreement games* and *inversion-problem games* proposed by von Ahn in 2008 [34]. All three templates describe games that produce tags or annotations for images or songs through creative combination of humans coming up with and verifying each other's tags. This same article also stated that, "People play not because they are personally interested in solving an instance of a computational problem but because they wish to be entertained."

While these games demonstrated that player effort could be harnessed and used to accomplish useful work, they also began perpetuating two limiting and potentially harmful notions about gwaps: first, that gwaps are about making boring tasks entertaining, and second, that gwaps are about labeling generic data, data in which the player has no personal investment. These notions quash the range of purposes where people imagine gwaps to be effective, and emphasize designing games around intrinsically uninteresting tasks.

In 2008, the protein folding game Foldit[1] officially launched. In Foldit, players manipulate the 3D structure of a protein and try to find the most physically plausible (lowest energy) way to pack the protein together. Like with von Ahn's games, players perform a task that is not solved through automation. Unlike von Ahn's games, the task is a spatial reasoning one instead of common sense knowledge, and does not fit any templates such as output-agreement. Foldit players also have the potential to be personally invested in the purpose, in learning about biochemistry, or in simply mastering the skill of packing proteins. Foldit players have found solutions on par with and better than automated techniques [20].

In his dissertation [7], Foldit creator Seth Cooper proposed a *framework for scientific discovery games*, informed by his work on Foldit. The framework suggests that a game should make synergistic use of computational strengths and human player strengths in a way that is more effective than either computer or human working alone. Additionally, the game should be playable by non-experts, and the game should coevolve with the players to become a better tool as they gain more experience. Finally, the game should encourage long-term player retention through competitive and collaborative reward structures.

While these are reasonable suggestions, they are presented in the context of a single successful game. We hope to expand upon this framework and make our recommendations in light of a wider variety of gwaps.

2.2 Crowdsourcing

Games with a purpose are a form of crowdsourcing and human computation. They share the goal of organizing human effort on tasks that cannot be fully automated. The online crowdsourcing platform Mechanical Turk launched in 2005 [18], shortly after the ESP Game was launch. These two related fields are roughly the same age and crowdsourcing is evolving along a trajectory similar to gwaps, from being used for simple, isolated tasks such as image labeling, to more complex and creative tasks that require high levels of skill.

A common approach in crowdsourcing is to think of work-

[1] `http://fold.it`

ers as *human processing units* [9] (HPUs) and to design crowd work flows inspired by computational work flows in distributed computing. Seen this way, crowd workers form a specialized computation device with complementary strengths to CPUs and GPUs. Implicitly, HPUs are presumed to share properties with other computational devices such as working at a steady rate, repeatedly producing equivalent outputs for identical inputs, and requiring only a small amount of software (instructions) to perform a self-contained task.

Also implied is that the HPU is a stand-in for an automated processor, that a human is being used in place of a computer for a simple task until computers become capable of doing the work themselves. We believe that the HPU model is a particularly bad area of crowdsourcing for gwaps to tackle, and the discontinuation of the ESP Game (in favor of paid crowd work) versus the continuing scientific impact of the more complex game Foldit reflects this.

The Future of Crowd Work [22] (FoCW), a report created by a group of crowdsourcing experts, outlines a framework for making crowd work complex, collaborative, and sustainable. Examples of complex tasks include document editing [1], taxonomy creation [5], iterative document writing [23], and customized itinerary planning [36]. The authors of FoCW are concerned that crowdsourcing itself will also be seen as exploitative and with limited applications unless it can reliably be used for complex, creative, and collaborative work. Their suggestions bias heavily towards computing-inspired architectures, such as the algorithmic toolkit TurKit [26], that still use individual workers as isolated processing units performing relatively simple tasks inside of a complex automated system.

We believe more complex games could be used to crowdsource more complex and creative tasks. Crowd work flows could give workers more responsibility to understand and navigate a complex system instead of carefully isolating them with their own individual units of work. Games are themselves systems, and the popularity of games demonstrate that players are comfortable, and even eager, to approach complicated new systems and figure out how to master them. Games with a purpose, a focused application of games in general, could provide an answer to how to crowdsource complex, creative, and collaborative tasks.

3. RECOMMENDATIONS

We would like to see a new era of effective, deep, and engaging gwaps emerge, instead of continuing the view that gwaps are for disguising work as play and providing superficial satisfaction. To this end, we make a series of recommendations based on gwaps and general crowdsourcing techniques that have proved effective (or ineffective) over the past decade.

3.1 Bias for transparency

Von Ahn said players were not interested in the problem, but Foldit players specifically cite the possibility to contribute to science as a motivating factor, including "to propel humanity into new directions."[2] Other scientific discovery games, including EteRNA and Phylo, share this property. PhotoCity players also cited the purpose "to reconstruct the world in 3D" as a reason for trying the game [32].

[2]https://fold.it/portal/node/991072

In the Verigames[3] [12] suite of gwaps centered on software verification, players are never told what software they are verifying by playing games. Even though the current iteration of the games apply verification to interesting open source software like BIND[4](which provides DNS service for much of the Internet) and Hadoop[5] (a distributed computation engine that serves many institutions around the world), players are left to speculate about the games' purpose given only the clue that the project is sponsored by the United State Defense Advanced Research Projects Agency (DARPA).

By making the purpose of the game explicit, it provides opportunities for the player to personally identify with the purpose or to be intrinsically motivated by their attachment to the purpose. Games provide players with an experience, usually an opportunity to make decisions that have impact in the game, but gwaps can provide an even richer experience and give players an opportunity to make an impact in the real world. Obfuscating the purpose can have the reverse effect of limiting the potential excitement of the game, or worse, making the player suspicious of what is being hidden from them.

3.2 Bias against HPU model

In crowdsourcing, the structure of the Mechanical Turk platform favors *micro-tasks*, jobs that can be done quickly by any worker for a small, pre-determined quantity of money. The assumption is that a thousand different workers would be able to do a thousand different jobs independently and in parallel, because each worker is an interchangeable HPU. Games like the ESP Game also focus on small, simple tasks, where the player's ability to contribute on their first task is about as good as it is on the thousandth task.

The model of treating workers or players as interchangeable human processors capable of small, isolated units of work works well with platforms such as Mechanical Turk, but makes for simplistic games.

Researchers in computer vision, the original domain gwaps were designed for, predominantly use Mechanical Turk for their HPU-style tasks. The few exceptions (beyond The ESP Game) include games and tools for *generating* data or observing how *human decision-making* works. PhotoCity [32] is a game for gathering thousands of new photos of a location in order to reconstruct a 3D model, and Bubbles [11] is a game for identifying discriminative patches of birds, released on Mechanical Turk as a paid task.

As the field of crowdsourcing has tried to advance beyond the HPU model, it has struggled with how to accomplish more complex and creative work. It is difficult to expect expertise from a short-term worker but also it is difficult to decompose certain tasks into small chunks. Games, as self-contained systems, might be the most appropriate platform for complex and creative tasks because they can simultane-

[3]http://www.verigames.com/

[4]https://www.isc.org/downloads/bind/

[5]https://hadoop.apache.org/

ously provided sustained interaction, train their players, and provide context and an environment for the task. (Figure 1)

Foldit was developed because automatically predicting protein structures is still a hard problem computationally. When the biochemists at the Baker Lab tried harvesting more processing power through the Rosetta@Home screensaver, they found that volunteers who watched and had begun to understand the screensaver were writing the lab suggesting tweaks [2]. A system of processing and visualizing the proteins was in place, they just needed to make it interactive so that human players could guide the computation. Foldit players are not human processors inside of an automated system, but drivers of the system.

Gwaps should try to avoid HPU-style tasks for two reasons: First, paid crowdsourcing is already effective for this purpose, and second, gwaps have the potential to make use of players' abilities to learn, understand, and control complex systems as a driver of that system instead of an isolated processing unit within that system.

3.3 Give players something to learn

It is hard to design for learning in a typical crowdsourcing setting. Platforms like Mechanical Turk do not distinguish between users with different skill levels or different amounts of experience, except for binary qualification requirements (pre-tests, essentially) that workers can be required to satisfy before working on a real task [6].

Games, on the other hand, provide huge opportunities to test a player's skill, and for players to learn and advance their skills. In his book *A Theory of Fun* [24], Raph Koster points out that learning the rules of a game and coming to master the mechanics are what makes games fun. Repetitive tasks that present no new challenges can make games boring.

Simplistic gwaps like the ESP Game do not provide an opportunity to learn, since each image labeling round in the game is similar in difficulty to every other image labeling round. Games like Foldit, PhotoCity, and Duolinguo do allow for or directly promote learning.

In the game PhotoCity, players walk around taking photos in order to reconstruct 3D models of buildings on their school campus. Through repeated interaction with the reconstruction pipeline and specific feedback for each photo, players developed their own strategies for what types of buildings to photograph and how to grow models around challenging obstacles such as corners. These insights were the same insights a domain expert in computer vision would have, but the players learned them through their own hands on experience [32].

Similarly in Foldit, when players were given the tools do design their own recipes, they invented patterns to make their protein folding more efficient. The patterns they invented turned out to be similar to those used by scientists [7, 19]. These players taught themselves to be better at the game, and as a result, were even more effective at the purpose.

[6] `http://docs.aws.amazon.com/AWSMechTurk/latest/AWSMechanicalTurkRequester/Concepts_QualificationsArticle.html`

In his latest game, Duolinguo [7], von Ahn addresses the need for bilingual translation experts by using the game to directly train these experts and give any willing participant the required skills. The collective purpose is to translate text on the web, but the game attracts and retains players by functioning as a personal language tutor.

Every game involves learning: learning the rules of the game, learning the skills that are necessary to play, and learning strategies to play well. By giving a gwap depth, players are able to improve their skills, develop strategies, and become better able to contribute to the game's purpose. Tying back to *A Theory of Fun*, obtaining a level of proficiency and mastery can be satisfying to players in its own right.

3.4 Match mechanics to purpose

In any game, the *core mechanic* is defined as the actions the player repeats most often while striving to achieve the game's overall goal [16, p. 188]. A solid core mechanic is crucial for any type of game, but especially for gwaps requiring that players' actions are useful to the purpose of the game. Games about collecting or labeling data tend to require mechanics that enforce the quality and reliability of the data. Games that require the player to build something or search for a solution need a way to evaluate the quality of the player's work. How effective a gwap can be depends on the effectiveness of the mechanics.

We define the term *orthogonal game mechanics* to refer to a mechanic in a gwap that does not serve the purpose, that distracts from or even gets in the way of the player's efforts to achieve the purpose. One example is OntoGalaxy [25], a space shooter game where the player navigates a spaceship and collects objects floating in space that have certain ontological properties. The purpose is to populate an ontology, but the mechanics necessitate that the player also be good at navigating the ship and quickly deciding if a moving object is part of the ontology. This limits the type of person who can contribute effectively and incentivizes players to gain mastery of a skill that does not contribute to the purpose. If the designers want to limit the reach of the game in order to very strongly appeal to a certain type of player, this may work in their favor.

Another example of games with orthogonal mechanics are the Landspotting [30] games, which introduce four separate games for land-cover labeling that each map on to an existing genre of game. The genres include a tagging game, a strategy game, a tower-defense game, and a tile game. The goal was to find a mechanic that would fit with the purpose, but each of the borrowed mechanics wound up emphasizing skills that were unrelated to the task of providing correct and accurate land-cover data.

Some gwaps thrive from borrowing mechanics from other games, especially if those mechanics align with and support the purpose. The game Open Trumps [3] borrows mechanics from the card game Top Trumps, in which players compare statistics and numerical properties for different entities such as films, vehicles, or athletes. The purpose of the Open Trumps game is to have players learn about about data by

[7] `http://www.duolingo.com/`

interacting with it through a game.

The game PhotoCity used the metaphor of ‘capture the flag’ to succinctly explain and visualize the territory-ownership mechanic. Virtual flags would be placed on real world buildings that players could capture by taking photos. The flags also spawned automatically [31] as the models grew to show the players where to go next.

The flags in PhotoCity were a borrowed mechanic, but the scoring system was invented based on the 3D reconstruction pipeline itself. PhotoCity reconstructed 3D point clouds and each new photo added new *3D points* to the model. These *3D points* were used as the game points. [31]

When existing mechanics do not fit the game’s purpose, gwap designers must invent their own, such as the points in PhotoCity or the agreement mechanisms in von Ahn’s games (ESP Game, Peek-a-boom, and Tag-a-tune).

In order for a gwap to be effective, the game mechanics should support the purpose, not try to disguise it or detract from it. Game mechanics should allow players to be efficient and effective, thus making a better contribution towards the purpose.

3.5 Provide timely and informative feedback

Feedback is what enables learning and assures the player they are performing the correct behavior. Appropriate feedback takes the form of a player taking an action in the game and the game responding with information about whether that was a good or bad action to take, and information about how the player’s action changed the state of the game (this exactly Crawford’s Listen-Think-Speak loop [8]).

Not having feedback makes it difficult for a player to know how well he or she is doing, which is especially important when game actions impact the real world and the players would like to know if they are actually helping to solve a problem.

The game EyeWire [8] lets players mark the shape of a neuron in 3D, but the correct shape is not known in advance so it is difficult to tell the player if her solution is correct. In Foldit, the correct shape of the protein is not known in advance either, but the Rosetta biochemistry software provided a starting point for designing the game interface and scoring mechanism [6]. For every move a player makes, the Rosetta energy function computes a score, providing immediate feedback to the player about how well their protein is folded.

In the game Happy Moths [29] from Citizen Sort, the game is to categorize moths by their shape and color by dragging each moth into a different bin. The game does not know the correct answers for all moths ahead of time, so it gives the player a score based on how well they did on the secret ‘happy moth’, or moth for which the correct answers were already known. This feedback is not enough to tell the user if they are making the right categorizations, which may cause the player to lack confidence in their work and give up on the game.

Gwaps should provide feedback to convey to their players whether or not they are achieving their goals and contributing effectively to the purpose.

3.6 Contribute in a way that adds up

This recommendation unifies previous recommendations, including explicitly stating the purpose, designing good mechanics, and providing informative feedback. Players should be able to contribute effectively to the purpose, but also understand and measure the specific contribution of their own actions. Additionally, players (and external audiences) should be able to understand and measure the impact of the collective contributions of all players. A player may be drawn to a game because they have a personal investment in the purpose, and a gwap turns their interest into *agency*, a way for them to have an effect on the purpose.

With a game like the ESP Game, it is hard for players to see the impact of their work. Are the labels a person provides new and informative? How do the resulting labels get used? Where does the way in which the data was collected provide added value, and where does it fall short?

In contrast, games like Foldit have lead directly to scientific discoveries and publications, which include the Foldit players’ teams on the author list. Two recent contributions include finding the solution of the crystal structure of the Mason-Pfizer Monkey Virus Retroviral Protease [21] and designing a novel synthetic enzyme for the Diels-Alder reaction [13].

Games with a purpose should give people access to important and challenging problems that they do actually care about solving, not because they simply “wish to be entertained.”

4. FUTURE RESEARCH

In order for gwaps to reach their potential, we need to have a better understanding of what makes them fun, engaging, and successful. We now suggest a few areas of future research, which is different than the work of making specific games for specific purposes, but it should help future gwap designers in their specific endeavors. Gwaps are expensive and challenging to develop, and tools are needed to make the production of successful gwaps less risky. Researchers, whose aim is to produce generalizable knowledge, are in the right position with the right resources to build these tools and associate knowledge. We propose four directions of research, but other directions of research are worth pursuing as well.

4.1 Cooperation and competition

Emmerich and Masuch’s work [14] explore the relationship between collaboration and competition in traditional games. Gwaps warrant their own investigation: What does it mean to use competition as a mechanic when, given the broader context, the players are really cooperating towards the purpose of the game?

Cooper’s framework for scientific discovery games includes a

[8] `www.eyewire.org`

statement about collaborative and competitive reward structures that encourage long-term involvement. Foldit uses both collaboration (teams) and competition (players versus players, teams versus teams) but what are the effects of these different reward and organization structures on players? Do they effectively encourage long term involvement or do people stick with Foldit for other reasons?

KissKissBan [17] is an image-tagging game that is similar to the cooperative ESP Game, but with a third adversarial player. They found that their emchanic increased label diversity and prevented player collusion. Designers looking to use either or both mechanic in their gwap should understand the effects each mechanic can have and how to use them effectively.

4.2 Intrinsic and extrinsic motivation

A big risk for gwaps is that a game will be developed that players do not want to play, so understanding what causes (or kills) motivation is crucial in developing an effective gwap. Motivation in online games in general has many facets, including achievements, social incentives, and immersion [35]. In gwaps, the same motivations can apply, but the purpose, or the impact of the game in the real world, can also be a strong motivator. In this paper, we recommended that designers expose the purpose because of evidence that it motivated players. Our question is how can in-game motivation and real-world motivation be combined to yield the most effective gwaps?

4.3 The phenomenon of unequal contribution

Games like the ESP Game assume that all contributions from all players are roughly equal. But *how much* a player contributes can vary drastically from player to player, and the *quality* of a player's contribution can vary as well. In games like Farmville (Zynga 2009), a player who plays (and spends) exponentially more than the average player is called a *whale*. According to Zynga, these whale's are the company's lifeline: "We rely on a small percentage of our players for nearly all of our revenue". [27] Farmville is not a gwap (the underlying purpose is to make money) but even in gwaps, there are likely to be players who contribute far more than the average player. For the sake of the purpose, the game should be able to make effective use of their work. A research question would be how to balance making the game an effective tool for power-users with how to make the game appealing and accessible to a wide audience of novice users (who have the potential to become power-users).

4.4 Design patterns for gwaps

Most gwaps fall into one of two genres, one of casual data-labeling games and one of scientific games such as Phylo, EteRNA, and EyeWire. It is possible that this split is because templates and frameworks have been written by von Ahn and by Cooper for only these types of games. Gwaps outside of these genres exist, but their patterns are not as well understood. Researchers like Celino [4] are just beginning to define a methodology for *location-based gwaps* such as PhotoCity and CityExplorer [28]. Many more games and genres exist, but the space of games has not yet been fully explored. By developing design patterns and templates for additional genres of gwaps, researchers will help future gwap designers avoid common pitfalls and create more effective and compelling games. In the same way crowdsourcing researchers are experimenting with and designing new crowd work flows, games researchers must innovate on game design specifically for games with a purpose.

5. CONCLUSION

The power of gwaps comes not from disguising *work* as *play*, but in providing the public with access to important and challenging problems facing science and society. In an effort to understand how the assumptions of early gwap designers have shaped the trajectory to present day and to envision how to move forward, we discussed a number of deployed gwaps and similar projects in human computation / crowdsourcing and made design recommendations and research recommendations that we hope will secure a future for interesting, responsive, deep, and engaging gwaps.

6. REFERENCES

[1] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and K. Panovich. Soylent: A word processor with a crowd inside. In *UIST '10*, UIST '10, pages 313–322, New York, NY, USA, 2010. ACM.

[2] A. Burke. Games that solve real problems: Crowdsourcing biochemistry. `http://www.forbes.com/sites/techonomy/2011/10/27/games-that-solve-real-problems-crowdsourcing-biochemistry/`, 2011.

[3] A. B. Cardona, A. W. Hansen, J. Togelius, and M. G. Friberger. Open Trumps, a Data Game. In *FDG '14: Foundations of Digital Games*, Apr. 2014.

[4] I. Celino. Location-based games for citizen computation. In P. Michelucci, editor, *Handbook of Human Computation*, pages 297–316. Springer New York, 2013.

[5] L. B. Chilton, G. Little, D. Edge, D. S. Weld, and J. A. Landay. Cascade: crowdsourcing taxonomy creation. In *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Request Permissions, Apr. 2013.

[6] S. Cooper, F. Khatib, and D. Baker. Increasing Public Involvement in Structural Biology. *Structure/Folding and Design*, 21(9):1482–1484, Sept. 2013.

[7] S. Cooper and Z. Popovic. A framework for scientific discovery through video games. 2011.

[8] C. Crawford. *The Art of Interactive Design.* A Euphonious and Illuminating Guide to Building Successful Software. No Starch Press, 2002.

[9] J. Davis, J. Arderiu, H. Lin, Z. Nevins, S. Schuon, O. Gallo, and M.-H. Yang. The hpu. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 9–16. IEEE, 2010.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255, 2009.

[11] J. Deng, J. Krause, and L. Fei-Fei. Fine-Grained

Crowdsourcing for Fine-Grained Recognition. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 580–587. IEEE Computer Society, 2013.

[12] W. Dietl, S. Dietzel, M. D. Ernst, N. Mote, B. Walker, S. Cooper, T. Pavlik, and Z. Popović. Verification games: making verification fun. In *FTfJP '12: Proceedings of the 14th Workshop on Formal Techniques for Java-like Programs*. ACM Request Permissions, June 2012.

[13] C. B. Eiben, J. B. Siegel, J. B. Bale, S. Cooper, F. Khatib, B. W. Shen, F. Players, B. L. Stoddard, Z. Popovic, and D. Baker. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nature Biotechnology*, 30(2):190–192, Jan. 2012.

[14] K. Emmerich and M. Masuch. Helping Friends or Fighting Foes: The Influence of Collaboration and Competition on Player Experience. 2013.

[15] A. Eustace. A fall spring-clean. `http://googleblog.blogspot.com/2011/09/fall-spring-clean.html`, 2011.

[16] T. Fullerton, C. Swain, and S. Hoffman. *Game Design Workshop*. Designing, Prototyping, & Playtesting Games. CRC Press, Jan. 2004.

[17] C.-J. Ho, T.-H. Chang, J.-C. Lee, J. Y.-j. Hsu, Chen, and Kuan-Ta. KissKissBan: A Competitive Human Computation Game for Image Annotation. pages 1–4, May 2009.

[18] J. Howe. The rise of crowdsourcing. *Wired magazine*, 2006.

[19] F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popović, D. Baker, and F. Players. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences*, 108(47):18949–18953, Nov. 2011.

[20] F. Khatib, F. DiMaio, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popović, et al. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature structural & molecular biology*, 18(10):1175–1177, 2011.

[21] F. Khatib, F. DiMaio, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popović, M. Jaskolski, and D. Baker. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology*, 18(10):1175–1177, Sept. 2011.

[22] A. Kittur, J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton. The future of crowd work. In *CSCW '13: Proceedings of the 2013 conference on Computer supported cooperative work*. ACM Request Permissions, Feb. 2013.

[23] A. Kittur, B. Smus, S. Khamkar, and R. E. Kraut. CrowdForge: crowdsourcing complex work. In *UIST '11*. ACM Request Permissions, Oct. 2011.

[24] R. Koster and W. Wright. *A Theory of Fun for Game Design*. Paraglyph Press, 2004.

[25] M. Krause, A. Takhtamysheva, M. Wittstock, and R. Malaka. Frontiers of a paradigm: Exploring human computation with digital games. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 22–25, New York, NY, USA, 2010. ACM.

[26] G. Little, L. B. Chilton, M. Goldman, and R. C. Miller. TurKit: human computation algorithms on mechanical turk. In *UIST '10*. ACM Request Permissions, Oct. 2010.

[27] D. MacMillan and B. Stone. Zynga's quest for big-spending whales. `http://www.businessweek.com/magazine/zyngas-quest-for-bigspending-whales-07072011.html`, 2011.

[28] S. Matyas, C. Matyas, C. Schlieder, P. Kiefer, H. Mitarai, and M. Kamata. Designing location-based mobile games with a purpose: collecting geospatial data with cityexplorer. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*, pages 244–247. ACM, 2008.

[29] N. Prestopnik and K. Crowston. Exploring collective intelligence games with design science: A citizen science design case. In *ACM Group Conference*, 2012.

[30] T. Sturn, M. Wimmer, P. Purgathofer, and S. Fritz. Landspotting - games for improving global land cover. In *Proceedings of Foundations of Digital Games Conference 2013 (FDG 2013)*, pages 117–125, May 2013.

[31] K. Tuite, N. Snavely, D.-Y. Hsiao, A. M. Smith, and Z. Popović. Reconstructing the world in 3D: bringing games with a purpose outdoors. In *FDG '10: Foundations of Digital Games*. ACM Request Permissions, June 2010.

[32] K. Tuite, N. Snavely, D.-Y. Hsiao, N. Tabing, and Z. Popović. PhotoCity: training experts at large-scale image acquisition through a competitive game. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Request Permissions, May 2011.

[33] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Request Permissions, Apr. 2004.

[34] L. von Ahn and L. Dabbish. Designing games with a purpose. *Commun. ACM*, 51(8):58–67, Aug. 2008.

[35] N. Yee, N. Ducheneaut, and L. Nelson. Online gaming motivations scale: development and validation. In *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Request Permissions, May 2012.

[36] H. Zhang, E. Law, R. Miller, K. Gajos, D. Parkes, and E. Horvitz. Human computation tasks with global constraints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 217–226, New York, NY, USA, 2012. ACM.