The Role of Game Theory in Human Computation Systems

Shaili Jain
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138 USA
shailij@eecs.harvard.edu

David C. Parkes
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138 USA
parkes@eecs.harvard.edu

ABSTRACT

The paradigm of "human computation" seeks to harness human abilities to solve computational problems or otherwise perform distributed work that is beyond the scope of current AI technologies. One aspect of human computation has become known as "games with a purpose" and seeks to elicit useful computational work in fun (typically) multi-player games. Human computation also encompasses distributed work (or "peer production") systems such as Wikipedia and Question and Answer forums. In this short paper, we survey existing game-theoretic models for various human computation designs, and outline research challenges in advancing a theory that can enable better design.

1. INTRODUCTION

Over the last few years, there has been a great deal of progress in human computation and more specifically "Games with a Purpose" (GWAP). As of yet, this field has not borrowed much from game theory and mechanism design, which models the behavior of rational agents in situations of strategic interdependence and designs optimal protocols given such a model, respectively. We believe that game theory has considerable promise in guiding the design of human computation systems when coupled with an appropriate model of human motivations. In addition to summarizing some of the existing game-theoretic models for human computation, our goal in this paper is to suggest a number of directions for future work.

The role of GWAP is to get humans to do useful work for free, in tasks that are easy for humans but hard for computers. The first GWAP was the ESP game [26], a two-player game for labeling images on the web. The development of the ESP game was followed by the development of Peekaboom [30], a game for locating objects within an image, Verbosity [29], a game for collecting common sense facts, and Phetch [28], a game for collecting descriptions for images on the web. Subsequent work by von Ahn and colleagues include the development of TagATune [19], a game to gather tags for music clips, Squigl, a game for gathering segmentation data for images, and Matchin [9], a game to elicit user preferences. The design of each of these games shows fantastic imagination and ingenuity and the success of GWAP such as the ESP game in generating useful computational work is impressive. One of the most important challenges in the design of GWAP is to make them fun to play. Here, von Ahn and Dabbish [27] describe some features that have been identified as salient in this respect, such as time pressure, a point system and high-score list.

While it seems unlikely that game theory can be used for *de novo* design, we do believe that game theory has a role in perturbing designs, and optimizing from within an existing design space. For example, we can ask the question: *could the equilibrium of a particular game be improved by small changes to the details of the game?* For example in the ESP game, perhaps it would make sense to try to leverage an entire sequence of words entered by both players, rather than taking the first match and discarding all other words [26]? Perhaps we could consider taking the k^{th} match for some $k > 1$, rather than introducing "Taboo words," or otherwise changing the scoring function? Indeed, we do not currently understand how far from optimal each game is, when measured in terms of eliciting the most efficient work possible while maintaining (or increasing) levels of participation. For some problems of human computation even satisfactorily formalizing the design objective will be a challenge.

GWAP fit into the larger realm of peer production and systems to elicit user-generated content. Examples of such systems are Wikipedia, Question and Answer sites (such as Yahoo! Answers), and YouTube. In these systems, a distributed user base contributes to, and gets value from, the system, but generally without any actual transfer of money. Users may participate in the system for many different reasons, including recognition, for personal enjoyment, and for the pleasure of working on things that they can do well [3]. One way in which GWAP differ from other peer production systems is that the former seem to more effectively conceal productive output from users and thus must strive to promote a fun user experience. With Wikipedia, Question and Answer forums, and YouTube, the ultimate purpose (e.g. the work) is more obvious to the user and thus the pleasure of working, contributing, and from achieving greater visibility can be reward in and of itself. Indeed with GWAP, it is a challenging task to state what the goal of each task is in quantifiable terms, although this is a necessary task to assess the optimality of the game design.

In looking to the role of game theory in driving and understanding design decisions, it is instructive to consider a parallel with the development of peer-to-peer (P2P) file sharing protocols, which have benefited from an explicit consideration of incentive properties. Early P2P protocols suffered from free-riding and it became necessary to provide incentives for people to upload files in addition to downloading them [1]. BitTorrent partially remedied this problem by introducing a tit-for-tat protocol to differentially provide upload resources to peers that reciprocate [6]. Still BitTorrent is not without its faults; e.g., users have incentive to intelligently

KDD-HCOMP'09, June 28, 2009, Paris, France
Copyright 2009 ACM 978-1-60558-672-4 ...$5.00.

under-report what pieces of the file they have to their neighbors [21] and the bilateral exchanges of BitTorrent are inefficient compared with multilateral exchanges [2]. Thinking more broadly about distributed work systems, game theory has been employed in studying the equilibrium properties of *scrip systems* that account for work contributed and consumed by users [8]. While GWAP may be different from peer to peer systems, in that fun and enjoyment are a large part of the success of GWAP, game-theoretic models can still be adopted when coupled with appropriately modeled utility functions. Moreover, human computation more broadly, in encompassing peer production and other systems, resembles the salient aspects of P2P file sharing problems in regard to opportunity for free-riding and the need to promote user contribution.

2. EXISTING GAME-THEORETIC MODELS

In what we believe to be the first work to consider game theory in GWAP, Ho et al. [11] develop and study the PhotoSlap game, a game for semantic annotation of images and provide a game-theoretic analysis of their game. PhotoSlap is based on the card game Snap, and the goal is to have players "slap" on two consecutive images that contain the same object, location, or person and not slap otherwise. They show that the target strategy is a subgame-perfect Nash equilibrium, meaning that the outcome is obtained in equilibrium for a particular model of user utility functions. In a more recent paper, Ho and Chen [12] seek to identify general themes across GWAP. They point out that many of the games involve either *simultaneous verification*, such as the ESP game, or *sequential verification*, such as Peekaboom. The equilibria in the simultaneous verification game are established by modeling these games as coordination games, while these authors adopt the refinement of *sequential equilibrium* [17] in studying the equilibrium behavior in sequential verification games. A simple equilibrium analysis can reveal qualitative differences about the distribution on tags provided by users.

In considering the ESP game, Jain and Parkes [16] propose a model within which one can seek to understand whether participants will coordinate on "easy" words associated with an image and also propose to perturb the design in order to push the equilibrium towards "hard" words. Easy words model generic words that could apply to large percentage of images that are in fact quite different. Such easy words could include colors. Hard words model specialized words that more accurately describe an image. For example, if we had an image of a red Ferrari, perhaps "red" and "car" would be easy words and "Ferrari" would be a hard word. The ESP game is modeled as a game of incomplete information in which a player makes two decisions. The first decision is to pick an "effort level," which dictates the domain from which a player will sample words for the image. Each image is associated with a universe of words, each of which has an associated probability (or frequency) to capture the likelihood at which a large population of individuals would output a particular word for the image, if they were asked to output only one word. If a user picks "low effort," she samples her dictionary from a small subset of the universe consisting of the most frequent words. If a user picks "medium effort," she samples her dictionary from a larger subset of the universe and finally, if a user picks "high effort," she samples her dictionary from the entire universe. Having privately chosen an effort level and sampled a dictionary of words, each player then picks a sequence with which to report the words.

A lexicographic utility model of *match-early preferences* captures the fact that each player first prefers to match rather than not, and then prefers to match earlier rather than later. It is shown that playing low effort and revealing words in order of decreasing frequency is a Bayesian-Nash equilibrium in the ESP game for any distribution of word frequencies over the universe. This low-effort equilibrium is in agreement with the experimental results of Weber et al. [31], who document a tendency for players to provide labels that are colors, synonyms or generic words.[1] An important question in the area of incentive design for the ESP game is how to design incentives to elicit high effort from players and as a result, richer, more descriptive labels. For example, Jain and Parkes [16] consider an alternate model of *rare-words first preferences*, in which the utility function of a player incorporates the fact that she prefers to match rather than not match, and then match on rare words before more frequent words, and only then break ties in favor of matching sooner rather than later.[2] Under this preference model, it is no longer the case that playing low effort is always a Bayesian-Nash equilibrium. It becomes possible for high effort in conjunction with playing words in order of increasing frequency to be a Bayesian-Nash equilibrium.

In a similar line of work, Jain et al. [15] propose a game-theoretic model for Yahoo! Answers and examine the equilibrium under the current scoring rule and under alternate scoring rules. Anyone can post a question in Yahoo! Answers, and anyone can answer a question. No money is exchanged for information provided, yet Yahoo! Answers elicits useful work. Each user starts out with a fixed number of points and loses points for asking questions while gaining points for answering questions. The system adopts a *best-answer scoring rule*: while every user receives two points for answering a question, the answer that is selected as a "best answer" by the asker receives ten points. In the proposed model, a fixed set of answerers each have a unique piece of information corresponding to a particular question. The strategic problem facing the answerers is to decide when to report their information, and the asker is modeled as satisficing, with a value threshold beyond which she will close the event and assign credit to the best answer (breaking ties at random). In the model, information aggregates as it is posted, with previous answers folded into later ones. The model seeks to capture the response to *factual questions* (e.g., "What are the causes of the economic crisis we are currently facing?") rather than discussion questions [10] (e.g., "What is your favorite movie of all time?"), with the former more likely to induce this kind of information aggregation behavior. Pieces of information may be *complements*, with each successive piece of information is worth more than previous pieces of information, or *substitutes*, with each successive piece of information is worth less than previous pieces of information. Under the best-answer scoring rule, the socially-optimal outcome in which all information is revealed immediately is achieved as the unique equilibrium for substitutes information, while the maximally suboptimal outcome in which all information is reported in the last round is the unique equilibrium for the complements case.

Considering this inefficiency, these authors then consider the effect of alternate scoring rules: an *approval voting rule*, where the asker can choose up to $k > 1$ answers that she likes and reward each of the k answerers equally, and a *proportional-share rule* where the asker distributes the points among the answerers. In

[1] They also find that a softbot trained on the language model learned from the set of ESP game labels can agree with a human player on 81% of the images that have at least one Taboo word *without looking at the image*. This provides fairly strong evidence that users are coordinating on easy words that can be used to describe a large set of images.

[2] We can think of the Taboo words in the ESP game as implementing a step function version of *rare-words first preferences*, in that all words that are not Taboo are rewarded equally if matched upon, yet no points are awarded if both players enter the same Taboo word.

the proportional-share rule, each answerer receives credit for the marginal value she contributes to the system, which depends on the time period in which she participates and how many answerers that have gone before her. Both rules are able to achieve the most efficient outcome in equilibrium for the complements case under certain restrictions on the valuation function of the asker and retain the most efficient outcome for the substitutes case, while differing in terms of whether they also introduce an inefficient equilibrium for the substitutes case.

3. RESEARCH DIRECTIONS

We see from existing work that it is possible to formalize some aspects of human computation systems within game-theoretic frameworks, and understand possible effects of changes in design on the equilibrium behavior within the confines of a model. In this section, we survey some future research directions in expanding the role of game-theoretic analysis in the design and understanding of human computation systems.

A Theory of Design for Human Computation. Mechanism design theory formalizes the idea of designing a game to promote good outcomes in a game-theoretic equilibrium [13]. The problem of mechanism design is to elicit private information (e.g., about preferences and capabilities) in order to make a decision about how to pick an alternative (e.g., an allocation of tasks or resources) to promote a set of desiderata. Canonical examples of mechanisms are voting protocols or auction protocols. Mechanism design poses an ambitious question: *what is the protocol design, across all possible protocols, that will optimally meet a set of desiderata?* By appealing to the "revelation principle," this problem is solved through a focus on direct revelation mechanisms in which design is limited without loss of generality to simple, one-shot games in which agents make a direct claim about their private information.

But the design challenges in GWAP and peer production are quite different from those of mechanism design theory. Rather than eliciting private information from participants, the challenge faced in the design of such systems is to elicit private *actions* (or work) from agents. Were the center able to make agents take particular actions, then human computation would reduce to a (still interesting) question of design to promote efficient, coordinated work. It seems unlikely that a revelation principle will exist for human computation systems because the "details" of the design clearly matter. The problem is more closely related to problems of *agency* and moral-hazard from microeconomic theory than mechanism design [4, 18]. For example, in human computation the effort that a user exerts in performing work is typically unobservable. The problem is related also to the problem of *environment design*, in which a designer seeks to perturb user environments in order to induce useful behaviors (and perhaps this requires understanding a user's underlying preferences) [33, 34].

It is interesting, then, to seek a theory for the design of game forms in which the equilibrium will elicit socially-efficient actions from users. The goal is likely not one of *de novo* design, but rather design from some existing space of possible game forms.

Open Question 1. *Can we agree on a design objective and parameterize the design space for particular problems of human computation, in order to enable a game-theoretic analysis that identifies the design that induces optimal behavior in equilibrium?*

Part of the challenge, of course, will be to formalize the design space. One useful step in this direction is provided by a recent taxonomy for GWAP into three categories: output-agreement games, inversion-problem games, and input-agreement games [27]. In *output-agreement games*, such as the ESP Game, players are given the same input and an individual round ends when both players have produced the same output. In *inversion-problem games*, such as Peekaboom, Phetch, and Verbosity, one player (the describer) is given a secret object and must reveal information about it to a second player (the guesser). The round ends when the guesser accurately guesses the secret object. In *input-agreement games*, such as TagATune, players are given inputs that are either the same or different. Players describe their inputs to each other and must decide whether they have the same input or not.[3] One concrete next step would be to couple a game-theoretic model with each of these three abstract game paradigms.

Altruism, Fun and Other-Regarding Preferences. Existing game-theoretic models for human computation have assumed that all agents participating in the system have selfish motivations. But there is evidence that users are behaving altruistically in peer-production systems such as Wikipedia, Yahoo! Answers, and YouTube [23]. In fact, game-theoretic analysis is not constrained to selfish utility functions, and it will be useful to expand our models of what motivates users. For example, altruism can naturally be incorporated within the selfish actor model by introducing other-regarding preferences [5] or inequality-averse utility functions [24]. To make a meaningful advance in this direction does present a challenge, however, in terms of identifying the right model of users in this larger space. A related challenge is to develop an understanding of what kinds of cognitive activities are enjoyable for users, and to take this into account during design. For example, can the lessons from existing GWAP design such as TagATune be generalized to understand the kinds of tasks that are inherently enjoyable?

Open Question 2. *How can altruism be best incorporated into game-theoretic models of human computation? Can we understand the role of altruistic preferences within a theory of design and change designs in accordance with the prevalence of other-regarding preferences? How sensitive are the socially-efficient equilibrium to the number of altruistic users? Can formal models be developed to predict the amount of fun that users will receive from cognitive activities?*

Another known issue in the design of human cooperation systems is that of "crowding out" of altruistic behaviors. For example, one famous example is that when one country moved from voluntary blood donations to small monetary incentives for donating blood, the donation rates went down instead of up [7]. This kind of interaction between motivations seems challenging to model; when does a positive reinforcement in one domain become a negative reinforcement in another domain? There is also plenty of evidence from behavioral economics about the influence of contextual factors on user decision-making [20].

The Macro-Level Design Problem. Existing work in game-theoretic modeling for human computation has focused on modeling a single-interaction between a fixed set of users; e.g., one ESP game, or a particular question in Yahoo! Answers. But in practice, these are typically complex dynamic systems, in which users participate multiple times (often with different sets of players) and in which previous experiences can influence future behavior. One aspect of game-theoretic modeling of the larger ecosystem will involve adopting the framework of *population games*. The theory of population games is suitable for systems of large numbers of

[3] The problem of tagging images was unsuccessfully implemented as an output-agreement game, yet ingeniously solved as an input-agreement game [19]. This shows the importance of being able to quantify fun in human computation systems and the value of being able to predict the amount of fun that users will receive.

agents, where agents interact anonymously and each agent's payoffs depends on the distribution over opponents' choices. Individual agents in the population are "small"; any single agent has little effect on any other agents' payoffs [25]. The goal of such modeling would be to examine the effects of such macro-level design issues as how to match players so that desirable behaviors (from a system-wide perspective) propagate through the population. This agenda can also borrow from the extensive literature of diffusion over social networks in the computer science and economics literature [14, 22]. Another interesting aspect of the macro-level problem is the problem of coordination, where the productive resources in a system are best coordinated to solve a global problem.

Open Question 3. *By adopting a macro-level, population-wide view of human computation systems, can we understand how to design information flows, matching and coordination algorithms to promote both the propagation of beneficial behaviors through a user society and a more efficient (e.g. coordinated) use and reuse of human resources?*

For example, in question and answer forums, perhaps a macro-level model should capture the observed behavior of some answerers in strategically choosing which questions to answer based on the amount of money or points promised to the winner [32, 23] or the number of other users that have already participated in the thread [32]. Are there some fraction of answerers who carefully select the question they answer to improve their best-answer percentage (which is displayed prominently in their profile)? How can we get these answerers to reveal relevant information, whenever they have it, rather than hand-picking the question they answer to attain "best-answer" status?

4. CONCLUSIONS

Looking ahead, there are a number of exciting directions for future work examining the role of game theory in formalizing the design of human computation systems, where incentives are a first-order consideration. This work should be informed by empirical evidence as much as possible. Indeed, the abundance of data presents one of the most compelling reasons to continue to pursue such an agenda because broader advances could also be made in folding models of human behavior back into game theory.

Acknowledgements

We thank Yiling Chen and Haoqi Zhang for helpful conversations and feedback. The first author is supported in part by an AT&T Labs Fellowship and an NSF Graduate Research Fellowship.

5. REFERENCES

[1] E. Adar and B. Huberman. Free riding on Gnutella. *First Monday*, 2000.

[2] C. Aperjis, M. J. Freedman, and R. Johari. A comparison of bilateral and multilateral exchanges for peer-assisted content distribution. In *Workshop on Network Control and Optimization*, 2008.

[3] Y. Benkler. Coase's Penguin, or, Linux and the nature of the firm. *The Yale Law Journal*, 112, 2002.

[4] P. Bolton and M. Dewatripont. *Contract Theory*. MIT Press, 2005.

[5] C. F. Camerer and E. Fehr. Measuring social norms and preferences using experimental games: A guide for social scientists. *Foundations of Human Sociality*, 2004.

[6] B. Cohen. Incentives build robustness in BitTorrent. In *Workshop on Economics of Peer-to-Peer Incentives*, 2003.

[7] B. S. Frey and R. Jegen. Motivation crowding theory: A survey of empirical evidence. *Journal of Economic Surveys*, 15:589–611, 2001.

[8] E. J. Friedman, J. Y. Halpern, and I. A. Kash. Efficiency and nash equilibria in a scrip system for p2p networks. In *Proc. 7th ACM Conf. on Electronic Commerce*, 2006.

[9] S. Hacker and L. von Ahn. Matchin: Eliciting user preferences with an online game. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 2009.

[10] F. M. Harper, D. Moy, and J. A. Konstan. Facts or friends? Distinguishing informational and conversational questions in social Q&A sites. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 2009.

[11] C.-J. Ho, T.-H. Chang, and J. Y.-J. Hsu. Photoslap: A multi-player online game for semantic annotation. In *Twenty-Second Conf. on Artificial Intelligence*, 2007.

[12] C.-J. Ho and K.-T. Chen. On formal models for social verification. In *Proc. Human Computation Workshop*, 2009. To appear.

[13] M. O. Jackson. Mechanism theory. In U. Derigs, editor, *The Encyclopedia of Life Support Systems*. EOLSS Publishers, 2003.

[14] M. O. Jackson and L. Yariv. Diffusion of behavior and equilibrium properties in network games. *American Economic Review*, 97(2):92–98, 2007.

[15] S. Jain, Y. Chen, and D. C. Parkes. Designing incentives for online question and answer forums. In *Proc. 10th ACM Conf. on Electronic Commerce*, 2009. To appear.

[16] S. Jain and D. C. Parkes. A game theoretic analysis of games with a purpose. In *Proc. 4th Intl. Workshop on Internet and Network Economics*, 2008.

[17] D. M. Kreps and R. Wilson. Sequential equilibria. *Econometrica*, 50:863–894, 1982.

[18] J.-J. Laffont and D. Martimort. *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press, 2001.

[19] E. Law and L. von Ahn. Input-agreement: A new mechanism for data collection using human computation games. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 2009.

[20] R. LeBoeuf and E. Shafir. Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making*, 16(77-92), 2003.

[21] D. Levin, K. LaCurts, N. Spring, and B. Bhattacharjee. BitTorrent as an auction: Analyzing and improving BitTorrent's incentives. In *SIGCOMM*, 2008.

[22] S. Morris. Contagion. *Review of Economic Studies*, 67(1):57–78, January 2000.

[23] K. K. Nam, M. S. Ackerman, and L. A. Adamic. Questions in, Knowledge iN?: A study of Naver's question answering community. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 2009.

[24] D. Ray, B. King-Casas, P. R. Montague, and P. Dayan. Bayesian model of behaviour in economic games. In *Proc. 22nd Conf. on Neural Information Processing Systems*, 2008.

[25] W. H. Sandholm. *Population Games and Evolutionary Dynamics*. MIT Press, 2009. To be published.

[26] L. von Ahn and L. Dabbish. Labeling images with a computer games. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 2004.

[27] L. von Ahn and L. Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.

[28] L. von Ahn, S. Ginosar, M. Kedia, and M. Blum. Improving accessibility of the web with a computer game. In *ACM Conf. on Human Factors in Computing Systems, CHI Notes*, 2006.

[29] L. von Ahn, M. Kedia, and M. Blum. Verbosity: a game for collecting common-sense facts. In *ACM Conf. on Human Factors in Computing Systems, CHI Notes*, 2006.

[30] L. von Ahn, R. Liu, and M. Blum. Peekaboom. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, 2006.

[31] I. Weber, S. Robertson, and M. Vojnovic. Rethinking the ESP game. Technical report, Microsoft Research, 2008.

[32] J. Yang, L. A. Adamic, and M. S. Ackerman. Crowdsourcing and knowledge sharing: Strategic user behavior on Taskcn. In *Proc. 9th ACM Conf. on Electronic Commerce*, 2008.

[33] H. Zhang and D. C. Parkes. Value-based policy teaching with active indirect elicitation. In *Twenty-Third Conf. on Artificial Intelligence*, 2008.

[34] H. Zhang, D. C. Parkes, and Y. Chen. Policy teaching through reward function learning. In *Proc. 10th ACM Conf. on Electronic Commerce*, 2009. To appear.