Gameful Markets for Collaboration and Learning

Stephan Leutenmayr
University of Munich
Oettingenstr. 67
80538 Munich, Germany
leutenma@cip.ifi.lmu.de

Fabian Kneissl
University of Munich
Oettingenstr. 67
80538 Munich, Germany
kneissl@pms.ifi.lmu.de

Sven Ziemer
Bauhaus Luftfahrt
Lyonel-Feininger-Str. 28
80807 Munich, Germany
sven.ziemer@bauhaus-luftfahrt.net

François Bry
University of Munich
Oettingenstr. 67
80538 Munich, Germany
bry@lmu.de

Abstract

The market metaphor offers a promising form of human computation for gathering and aggregating individual opinions into a collective point of view for groups of people. In this article, we identify different kinds of market forms, their specific goals, and their participants' behavior. Building on this, we first determine appropriate reward designs for a good functioning of the markets as opinion aggregators. Furthermore, we stress potential possibilities for facilitating participants' learning. To illustrate and validate the approach proposed, we investigate two different market applications with different kinds of rewards and goals.

Introduction

For many tasks, information for their successful execution is dispersed among several information sources. A successful execution thus requires to gather and aggregate this information, for example in decision making or linguistic research. Both computer-based and human-based approaches are conceivable for this process. Computer-based approaches often require a large infrastructure and complex algorithms for collecting all the desired information. Human-based approaches are often time-consuming and costly. Recently, approaches have emerged for involving both humans and computers for the solution of complex tasks under the term *human computation* (Quinn and Bederson 2011). There, tasks are distributed for execution over a large group of people. These tasks are often algorithmically difficult and simpler, but well executed, by humans. People then collaborate via the human computation mechanism in order to achieve a goal. In decision making, for example, it may be simpler to engage people in the gathering of information from diverse sources if these sources change from decision to decision and need to respect varying evaluation criteria along the way rather than to develop algorithmic approaches every single time. For aggregating the single contributions of a group of people, an aggregation method is required. Virtual markets have been utilized for several information aggregation tasks, for example in the forecast of presidential elections (Forsythe et al. 1992), the gathering of product preferences (Dahan, Soukhoroukova, and Spann 2010), and the adjustment of research portfolios (Gaspoz 2008). These markets thereby combine human information gathering on their topics and computer-based aggregation of the trades in order to form a market price. Thus, they may also be counted among the human computation-based approaches. Generally, a market organizer is interested in achieving a useful result from the market application. The definition of a useful result thereby depends on the application context of the market. The result in turn is impacted by the trading behavior of the single participants. This behavior can be influenced by providing appropriate rewards. The design of such rewards thus influences the behavior and lastly the quality of the market result. In this article, we investigate the typical market application scenarios of prediction markets, preference markets, and decision markets. We then identify their common contexts of gathering collective forecasts and sincere opinions, respectively, and their respective definitions of a useful result, that is, an accurate forecast and a collective sincere opinion. Finally, we highlight possible rewards helping in generating these outcomes.

Human Computation, Games, and Markets

The term human computation in the modern usage has been defined by Luis von Ahn as a paradigm for taking advantage of human processing power for solving problems that so far cannot be solved by computers alone (von Ahn 2007). Two aspects are regarded necessary for human computation: humans must have a *conscious* role in determining the outcome of the computation, and the organization of the humans' and the computer's work must be under *explicit control* (Law and von Ahn 2011).

One specific type of human computation are human computation games, also called Games With a Purpose (GWAPs). Here, the task that humans have to solve is designed such that it is solved by humans while playing a game (von Ahn and Dabbish 2004). Compared to general human computation in which human workers are often payed for their work, e.g., in Amazon Mechanical Turk (Chen, Menezes, and Bradley 2011), players in GWAPs contribute their processing power for free. Proper incentives and rewards have to be provided for motivating players to participate and to contribute meaningfully.

Markets may be counted among human computation approaches. The basic idea is to represent the single alternatives of a given topic as stocks on a market and to have participants buy shares of their favored alternatives and sell shares of the undesired ones. The highest priced alternatives are then considered to be the collectively determined favorites. Thereby, participants gather information from diverse sources and distill it into their personal assessment of the alternatives. They then convey this assessment by trading on the market. The price mechanism of the market aggregates the single contributions of the participants to form an overall assessment. Both aspects of human computation, the humans' conscious role and the explicit control, are apparent. This combination of human information gathering and computer-based aggregation for forming an overall result can be seen as a form of human computation (Bry 2012).

Markets are furthermore related to games and GWAPs because markets can have a game-like character. Different ways for obtaining information and for trading encourage market participants to create individual strategies.

Behavior of Market Participants

As highlighted in the previous section, markets represent a form of human computation for gathering and aggregating information from large groups of people. The general goal of markets for information aggregation is to achieve a useful result in the end. Participants in markets are often assumed to be utility maximizing and to derive this utility from the reward they expect to gain in return for their participation (Manski 2006). Thus, by designing the reward a market offers appropriately, we should be able to influence the participants' utility maximizing efforts to be beneficial for a useful market result. The meaning of *useful* thereby depends on the context of the application situation.

We identified three kinds of market applications from the literature and own studies. The first kind of market is that of *prediction markets*. There, the goal is to aggregate forecasts of an uncertain future event and to obtain an accurate forecast (Graefe 2009). Participants trade shares according to their personal forecast and receive a reward at market end for the accuracy of their forecast. The accuracy is determined by comparing their respective forecast with the actual outcome of the event. As utility maximizers, participants are incentivized to contribute their best forecast in order to maximize their expected reward. In this context, it is best for the participants to contribute their best forecast. This best forecast contributes to the goal of the market, namely an accurate forecast. Beneficial behavior in this market application consists of the steps as depicted in Figure 1. There, a participant develops his personal forecast of the uncertain event, trades in the market in order to represent this forecast and updates his forecast and market holdings in case of new information.

Figure 1: Behavior model for prediction markets

The second market application is that of *preference markets*. Such markets are employed to aggregate the preferences of a group of people on a given topic such as product features and to achieve a collective preference indication (Dahan, Soukhoroukova, and Spann 2007). Ideally, participants buy shares of the preference options they favor and sell shares of the undesired ones. That is, they build their personal preference ranking, represent this ranking in the market by trading and update their preferences on new insights on the preference options. In such markets, however, an external event is missing for creating the reward for the participants. Thus, participants are often rewarded based on a comparison of their preference ranking with the final market outcome. This, in turn, is likely to lead to a change in the behavior of participants. They no longer contribute their own preferences but rather develop a belief on the others' preferences as this will maximize their expected reward. This is known as the *Keynesian Beauty Contest*, described by economist John Maynard Keynes (Keynes 1936). The core of this beauty contest is that participants start to guess the votings of the others and adjust their voting to take advantage of it. In Figure 2, we adapt the model of Dahan et al. to show this behavior. Dahan et al. conclude that preference markets with such rewards aggregate individual forecasts on the group's preference ranking rather than individual preference rankings (Dahan, Soukhoroukova, and Spann 2010).

Figure 2: Behavior model for preference markets

The third kind of markets are *decision markets* (Leutenmayr et al. 2011). The goal in decision markets is to aggregate the assessments of participants on the single options of a pending decision and to produce a collectively selected decision. In such decision markets, participants buy shares of the decision options they approve of and sell shares of the options they do not favor. The introduction of a forecast-based reward would likely lead to a beauty contest as with preference markets. Hence, rewards that would induce participants to engage in a beauty contest or forecasting are deliberately omitted. Rather, the reward consists of the influence on the final decision that participants have. In Figure 3, this model of behavior is shown. A participant develops his own opinion on the given decision options, trades shares in order to represent this opinion in the market and potentially updates his opinion on the arrival of new information. In this case, participants are assumed to derive their utility from the influence they exert on the decision result and thus to maximize the realization of their personal favorite.

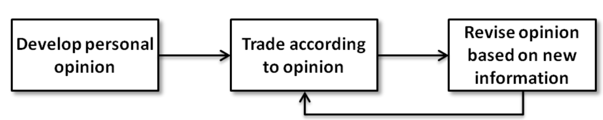

Figure 3: Behavior model for decision markets

From these three market applications, we derive two application contexts. The first context is concerned with achieving a collective forecast of a certain variable as with prediction markets and preference markets. This is represented by the following question that is posed to the participants: “what will the others think of this subject?” In this case, the result is deemed most *useful* if it is as accurate as possible. The second application context deals with aggregating the sincere opinions of group members on a given subject. The question for the group members is thereby: “what do you think of this subject?” There, the metric of accuracy cannot be established as the goal consists in gathering the individual opinions, not forecasts. A *useful* result is therefore characterized by it containing the single sincere opinions of participants.

Reward Design

In both contexts given in the previous section, the achievement of a useful result depends on beneficial behavior of the participants. This beneficial behavior can be encouraged by providing appropriate rewards (see Figure 4).

Figure 4: Factors influencing the result quality

Thus, we want to design the reward for market participants in such a way that they adjust their trading behavior in pursuit of that reward and thus to achieve a useful market result. In the following, we discern two approaches for the reward corresponding to the two contexts identified in the previous section.

Performance-Based Reward

There are market application contexts which can profit from obtaining a forecast on a given variable from a large group of people. Rewarding the forecast of the group’s opinion as in the preferences markets described by Dahan et al. (Dahan, Soukhoroukova, and Spann 2010) may also be perfectly reasonable as a goal for a market application. The reward should then be linked to the accuracy of this forecast. This is a common reward design in prediction markets. The reward may thereby consist of an in-market payment, a leaderboard ranking, and an after-market payoff. The reward is determined by comparing the respective forecast with the actual outcome of the forecast variable. The better the forecast of a participant is, the higher the payment the participant receives or the higher up in the leaderboard he ranks. As participants are assumed to be utility maximizing, they should then aim at maximizing this reward. In this design, they maximize their reward by providing an accurate forecast. This in turn contributes to the useful result the market organizer wants to achieve. Thus, the provided reward should contribute to the achievement of a useful result. This is highlighted in Table 1. A performance-based reward should contribute to the achievement of the forecasting goal. However, such a reward is likely to bias the achievement of an aggregation of individual opinions.

	performance reward	outcome reward
individual opinion goal	−	+
forecast goal	+	+/-

Table 1: Suitability of rewards for market goals

Outcome-Based Reward

In other market application contexts, the goal is to obtain an aggregation of the sincere opinions of the participants. In such contexts, the market organizer does not seek forecasts from participants. Thus, we should not reward any forecasting behavior by participants. In the opinion gathering context, the quality of the market result is subjective and lies in the eye of the respective beholder. There, it is more a matter of satisfaction with the result and the achievement of one’s favored result. Therefore, in such situations, the reward should come from the actual outcome of the market. In such situations, the reward could consist of the influence that participants have on the actual outcome. Such an outcome-based reward would contribute to the aggregation of the single opinions as the participants would be interested in representing their individual opinion. The offering of an outcome-based reward in a situation with a forecasting goal could lead to mixed results as participants would then be drawn between contributing a forecast and realizing their influence on the respective outcome (see Table 1).

For applications of the market metaphor to information aggregation tasks, we therefore suggest to analyze a given application situation with respect to its objectives and the two aforementioned contexts. For aggregation objectives that seek to gather forecasts of a certain variable and for which a reasonable external event may be consulted, a performance-based reward is likely to deliver an accurate forecast. For situations corresponding to the second context, in which the aggregation of individual opinions is sought for, the market should be designed in such a way that participants derive their utility from the actually chosen outcome in order to elicit meaningful contributions.

Learning in Markets

Among the motivations for people to participate in virtual markets is also the human learning that results from information aggregation and sharing. People learn when they develop strategies for trading in a market, when they gather information to assess the stocks and when they adopt to the collective opinion that emerges from the participation of a group of people. Aggregating and sharing information is a feature that virtual markets share with social media. The use of social media for learning has been studied in the context of *individual learning* in Personal Learning Environments (PLE) (Dabbagh and Kitsantas 2012). The term *informal learning* describes learning that occurs outside formal school settings (Bull et al. 2008). A PLE is thereby a combination of social media that enable learners to create, reorganize and share content. Comparably, virtual markets enable the aggregation of information, its organization and its sharing among a potentially large number of people.

A learning model for PLEs can be formulated as the following steps (Zimmerman 2000): (1) Personal information management, (2) Social interaction and collaboration, and (3) Information aggregation and management. The first step involves the beliefs and goals that people have prior to engaging in a virtual market. The second step refers to the gathering as well as sharing of information between traders. In the final step, participants self-monitor their behavior, review the collected information and consider whether the original beliefs are to be updated due to new information and whether they achieved their goals.

In addition to the three kinds of human learning, the prices that markets generate from the trades of their participants can be interpreted as a form of machine learning. There, the prices represent a one-dimensional reduction of the many opinions that the single participants contribute. A related approach proposes to utilize markets for combining different machine learning methods in order to produce inferences from their different inputs (Storkey 2011).

Two Market Applications

The two kinds of rewards identified in the previous section have been realized in different market-based systems by the authors. Metropolitalia represents the scenario of a beauty contest utilizing the performance-based reward whereas Liquid Decision Making is applicable in situations conforming to the sincere opinion situation and utilizes an outcome-based reward. Below, we introduce the two approaches and highlight the design of their rewards in order to be attractive to participants and to encourage repeated participation.

Metropolitalia

Metropolitalia is a platform for linguistic field research on the Italian language and especially on its regional language varieties (Kneissl and Bry 2012). It is publicly accessible at http://www.metropolitalia.org since August 2012. Two market-based social media gather complementary data and meta-data on phrases in Italian dialects and other language varieties, including the geographical region and social characteristics like age, gender, and level of education of speakers. A screenshot of the platform is shown in Figure 5.

On metropolitalia, users can –amongst others– create so-called assessments consisting of a phrase, its geographical region, and an estimation of how many other users agree to the user's assessment, i.e., choose the same, or a similar, geographical region. This estimation is a prediction of how other users characterize the phrase. The closer the estimation is to the real agreement proportion, the more money the assessment is worth (see Figure 5). As a consequence, success on metropolitalia depends on how one is skilled at forecasting others' conceptions. This is a typical case of the beauty contest, in which participants need to reflect on each others' behavior and adapt their behavior accordingly. While the beauty contest effect is meant by Keynes as a criticism of speculation on financial markets, it contributes to the aim of metropolitalia because the perception of the collective opinion is much more relevant than that of single participants. The performance-based rewards include play-money for the user, a leaderboard ranking of the participants performing best, and an individual dashboard displaying the current valuation of own assessments. These rewards have shown that users participate on metropolitalia and contribute quality data (Bry et al. 2013).

In addition to gaining rewards, users learn on metropolitalia in two ways. On the one hand, users get to know phrases they did not encounter before. Users can furthermore reveal a translation of dialect phrases into standard Italian. This can lead to a reflection on the own usage of words or phrases or the adoption of the phrase into the own vocabulary. On the other hand, the characteristics of phrases are revealed to users. They can learn which phrases or words are distinctive for which geographical region and for speakers of which age, gender, or level of education.

Liquid Decision Making

Liquid Decision Making (LDM) is a market-based approach for decision making in potentially large groups of people. The idea is to represent the single decision options as stocks on a market. Participants buy shares of the decision options they favor and sell shares of the undesired options. The highest ranking stock or stocks are then interpreted as the collectively selected decision. LDM represents a market application conforming to the sincere opinion context as given above. That is, the goal of the market utilized in LDM is to aggregate the individual opinions of the group members on the single decision options. The beneficial behavior of participants would be to assess the decision options and convey their individual opinion by trading accordingly, as highlighted in Figure 3. In such a context, rewarding participants for their forecast accuracy of the end result would lead to the beauty contest of the forecast context and the aggregation of the group forecast. Therefore, we designed the rewards in this market approach in order to avoid such guessing behavior. To this end, the reward for the participants consists of their influence in the finally selected decision option. In this way, participants that have an interest in the final decision option are incentivized to contribute their sincere opinion in order to push their favored decision option. Figure 6 displays the interface of LDM for conveying one's opinion on a decision option. We evaluated LDM in a lab and a case study (Leutenmayr and Bry 2011; Leutenmayr, Ziemer, and Bry 2013) as a prototypical implementation. Amongst others, we found this reward design to be adequate for gathering sincere opinions.

Participants in LDM learn both on the decision options in the market and on the opinions of the others by observing the market development and the ranking of the decision options. Furthermore, they may also gain insights by communicating with the others using the comments feature of LDM.

Conclusion and Future Work

In this article, we investigated virtual markets as a means for human computation. In human computation, a main chal-

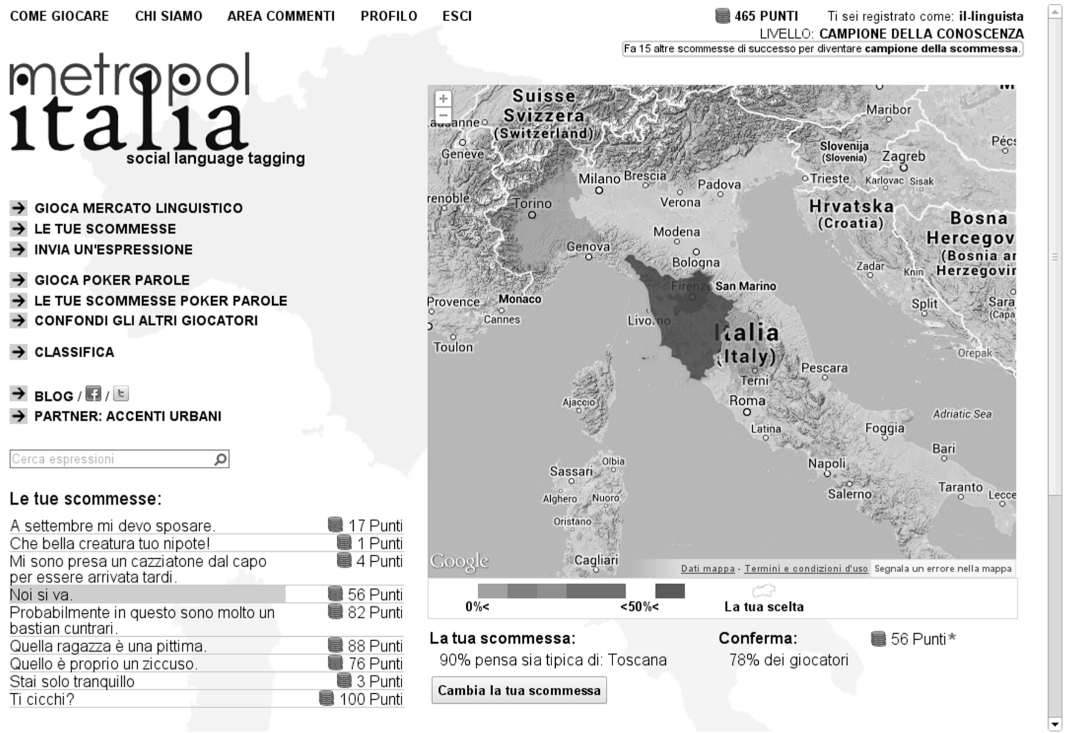

Figure 5: Screenshot of metropolitalia depicting a user's list of assessments. The highlighted phrase "Noi si va" (in English: "Let's go") is characterized as being used in the region Toscana (highlighted on the map) and the user estimates that 90% of all users agree. Currently, the agreement is 78% and thus the assessment is worth 56 points.

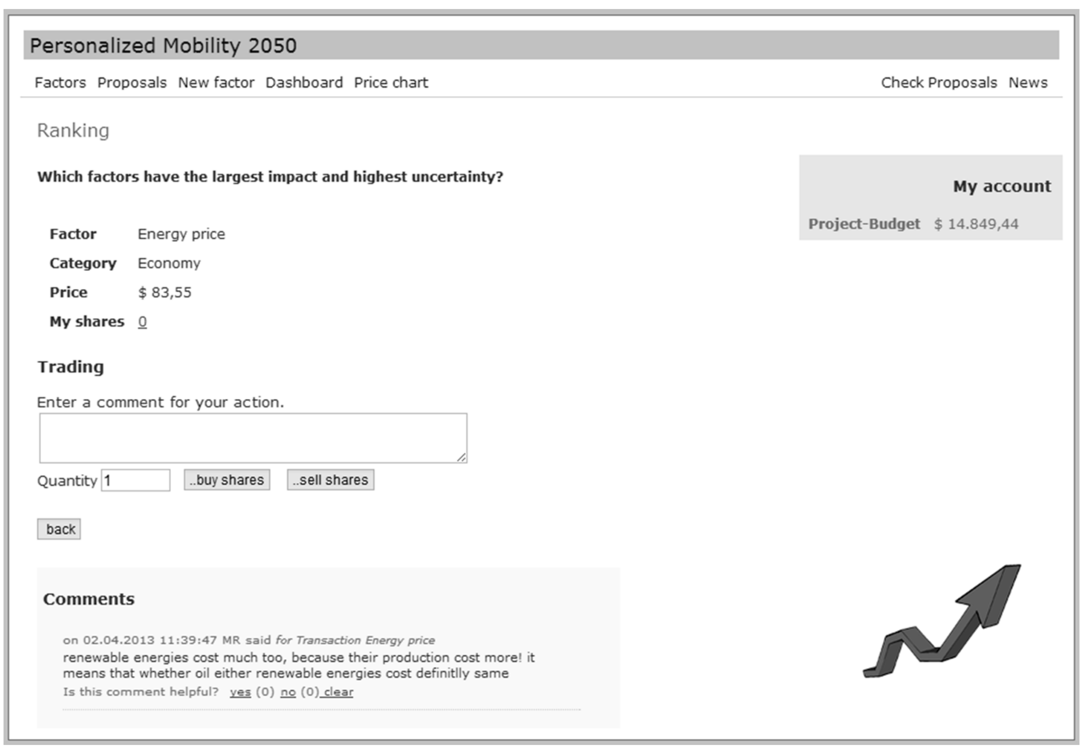

Figure 6: Screenshot showing the trading interface of LDM. The participant has a remaining budget of $ 14849.44 and may specify a quantity for trading in the currently selected decision option "energy price". Additionally, he may leave a comment.

lenge is thereby to encourage beneficial behavior in participants that contributes to the achievement of the human computation goal. Rewards play a central role in encouraging beneficial behavior in virtual markets. We identified the contexts of a forecast situation and a sincere opinion situation for designing such rewards. Forecast situations would require to reward the accurate forecast of a group's opinion, whereas sincere opinion situations would require to reward the contribution of sincere opinions. Accuracy can be based on the comparison of the variable to be forecast and the forecast of the participant. Sincere opinion contribution is achievable by the direct influence that participants have on the produced collective opinion.

We furthermore envisage to explore extra means for encouraging beneficial behavior in markets. Further studies on the effects of the two types of rewards on participants' perception, contribution and possible biases can help to identify opportunities and drawbacks of both types of rewards. For metropolitalia, the potential existence of a bias induced by the performance-based reward needs to be investigated carefully. In LDM, additional rewards are conceivable that do not rely on market performance but rather reward engagement, for example for repeated participation or for helpful comments. Both the effectiveness of such measures and the interplay between these different kinds of rewards would then be interesting to investigate in further studies.

Acknowledgments

This research has been funded in part by the German Foundation of Research (DFG) within the project Play4Science number 578416.

References

Bry, F.; Kneissl, F.; Krefeld, T.; Lücke, S.; and Wieser, C. 2013. Crowdsourcing for a Geographical and Social Mapping of Italian Dialects. In *Proceedings of the 2nd International Workshop on Social Media for Crowdsourcing and Human Computation at ACM Web Science (SoHuman)*.

Bry, F. 2012. Human Computation and Economics. Research Report PMS-FB-2012-3, Institute for Informatics, University of Munich.

Bull, G.; Thompson, A.; Searson, M.; Garofalo, J.; Park, J.; Young, C.; and Lee, J. 2008. Connecting Informal and Formal Learning Experiences in the Age of Participatory Media. *Contemporary Issues in Technology and Teacher Education* 8(2):100–107.

Chen, J. J.; Menezes, N. J.; and Bradley, A. D. 2011. Opportunities for Crowdsourcing Research on Amazon Mechanical Turk. Technical report.

Dabbagh, N., and Kitsantas, A. 2012. Personal Learning Environments, Social Media, and Self-regulated Learning: A Natural Formula for Connecting Formal and Informal Learning. *The Internet and Higher Education* 15(1):3–8.

Dahan, E.; Soukhoroukova, A.; and Spann, M. 2007. Preference Markets: Organizing Securities Markets for Opinion Surveys with Infinite Scalability. *SSRN eLibrary*.

Dahan, E.; Soukhoroukova, A.; and Spann, M. 2010. New Product Development 2.0: Preference Markets – How Scalable Securities Markets Identify Winning Product Concepts and Attributes. *Journal of Product Innovation Management* 27(7):937–954.

Forsythe, R.; Nelson, F.; Neumann, G. R.; and Wright, J. 1992. Anatomy of an Experimental Political Stock Market. *The American Economic Review* 82(5):1142–1161.

Gaspoz, C. 2008. Prediction Markets as an Innovative Way to Manage R&D Portfolios. In *Proceedings of the CAiSE-DC'08 15th Doctoral Consortium*.

Graefe, A. 2009. *Prediction Markets versus Alternative Methods - Empirical Tests of Accuracy and Acceptability*. Ph.D. Dissertation, Institute of Information Systems and Management, Karlsruhe Institute of Technology.

Keynes, J. M. 1936. *The General Theory of Employment, Interest, and Money*. Macmillan Cambridge University Press.

Kneissl, F., and Bry, F. 2012. Metropolitalia: A Crowdsourcing Platform for Linguistic Field Research. In *Proceedings of the IADIS International Conference WWW/Internet*.

Law, E., and von Ahn, L. 2011. Human Computation. In Brachman, R. J.; Cohen, W. W.; and Dietterich, T., eds., *Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool Publishers. 1–121.

Leutenmayr, S., and Bry, F. 2011. Liquid Decision Making: An Exploratory Study. In *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services (iiWAS)*, 391–394.

Leutenmayr, S.; Bry, F.; Schiebler, T.; and Brodbeck, F. 2011. Work in Progress: Do They Really Mean It? Assessing Decision Market Outcomes. In *Proceedings of 4. Workshop Digitale Soziale Netze im Rahmen der 41. Jahrestagung der Gesellschaft für Informatik*.

Leutenmayr, S.; Ziemer, S.; and Bry, F. 2013. Decision Markets for Continuously Reflected Collective Decisions. In *Proceedings of the Third International Conference on Social Eco-Informatics (SOTICS)*. to appear.

Manski, C. F. 2006. Interpreting the Predictions of Prediction Markets. *Economics Letters* 91(3):425–429.

Quinn, A. J., and Bederson, B. B. 2011. Human Computation: A Survey and Taxonomy of a Growing Field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 1403–1412.

Storkey, A. J. 2011. Machine Learning Markets. *Journal of Machine Learning Research* 15:716–724.

von Ahn, L., and Dabbish, L. 2004. Labeling Images with a Computer Game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 319–326.

von Ahn, L. 2007. *Human Computation*. Ph.D. Dissertation, Carnegie Mellon University.

Zimmerman, B. J. 2000. Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective. In Boekaerts, M.; Pintrich, P. R.; and Zeidner, M., eds., *Handbook of Self-Regulation*. San Diego: Academic Press. 13–39.