

# Incentivizing Evaluation with Peer Prediction and Limited Access to Ground Truth (Extended Abstract)\*

Alice Gao<sup>1†</sup>, James Wright<sup>2</sup> and Kevin Leyton-Brown<sup>3</sup>

<sup>1</sup>University of Waterloo, Waterloo, ON, Canada

<sup>2</sup>University of Alberta, Edmonton, AB, Canada

<sup>3</sup>University of British Columbia, Vancouver, BC, Canada

alice.gao@uwaterloo.ca, james.wright@ualberta.ca, kevinlb@cs.ubc.ca

## Abstract

In many settings, an effective way of evaluating objects of interest is to collect evaluations from dispersed individuals and to aggregate these evaluations together. Some examples are categorizing online content and evaluating student assignments via peer grading. For this problem, one challenge is to motivate participants to conduct such evaluations carefully and to report them honestly, particularly when doing so is costly. Existing approaches, notably peer-prediction mechanisms, can incentivize truth telling in equilibrium. However, they also give rise to equilibria in which agents do not pay the costs required to evaluate accurately, and hence fail to elicit useful information. We show that this problem is unavoidable whenever agents are able to coordinate using low-cost signals about the items being evaluated (e.g., text labels or pictures). We then consider ways of circumventing this problem by comparing agents' reports to ground truth, which is available in practice when there exist trusted evaluators—such as teaching assistants in the peer grading scenario—who can perform a limited number of unbiased (but noisy) evaluations. Of course, when such ground truth is available, a simpler approach is also possible: rewarding each agent based on agreement with ground truth with some probability, and unconditionally rewarding the agent otherwise. Surprisingly, we show that the simpler mechanism achieves stronger incentive guarantees given less access to ground truth than a large set of peer-prediction mechanisms.

## 1 Introduction

In many practical settings, an effective way of evaluating objects of interest is to collect evaluations from a large, dispersed group of agents. These evaluations can then be ag-

gregated together and provided as a service, as in online forums such as Rotten Tomatoes, Yelp, and TripAdvisor, which obtain feedback from millions of users about movies, restaurants, and travel destinations. Alternatively, the aggregated evaluations can be used directly. For example, crowdsourcing platforms are increasingly used to collect semantic labels of images and online content for use in training machine learning algorithms.

However, these agents may not be motivated to invest costly effort to obtain accurate evaluations. Therefore, an important problem in artificial intelligence is to design incentives to motivate large groups of agents to obtain and to reveal accurate information [Prelec, 2004; Miller *et al.*, 2005; Jurca and Faltings, 2009; Faltings *et al.*, 2012; Witkowski and Parkes, 2012; Witkowski *et al.*, 2013; Dasgupta and Ghosh, 2013; Witkowski and Parkes, 2013; Radanovic and Faltings, 2013; Radanovic and Faltings, 2014; Radanovic *et al.*, 2016; Riley, 2014; Zhang and Chen, 2014; Waggoner and Chen, 2014; Kamble *et al.*, 2015; Kong *et al.*, 2016; Shnayder *et al.*, 2016].

We are particularly motivated by the peer grading problem, which we will use as a running example. Students benefit from open-ended assignments such as essays or proofs. However, such assignments are used relatively sparingly, particularly in large classes, because they require considerable time and effort to grade properly. An efficient and scalable alternative is having students grade each other (and, in the process, learn from each other's work). Many peer grading systems have been proposed and evaluated in both the artificial intelligence and education literatures [Kulkarni *et al.*, 2014; Raman and Joachims, 2014; Wright *et al.*, 2015; de Alfaro *et al.*, 2015], albeit with a focus on evaluating the accuracy of grades collected under the assumption of full cooperation by students.

However, no experienced teacher would expect all students to behave non-strategically when asked to invest effort in a time-consuming task. An effective peer grading system must therefore provide motivation for students to formulate evaluations carefully and to report them honestly. Many approaches have been developed to provide such motivation. One notable category is peer-prediction methods [Prelec, 2004; Miller *et al.*, 2005; Jurca and Faltings, 2009;

<sup>†</sup>Contact Author

\*This paper is an extended abstract of an article in Artificial Intelligence [Gao *et al.*, 2019]. This paper was first presented at the second workshop on Algorithmic Game Theory and Data Science at the 17th ACM Conference on Economics and Computation.

Faltings *et al.*, 2012; Witkowski and Parkes, 2012;  
Witkowski *et al.*, 2013; Dasgupta and Ghosh, 2013;  
Witkowski and Parkes, 2013; Radanovic and Faltings, 2013;  
Radanovic and Faltings, 2014; Radanovic *et al.*, 2016;  
Riley, 2014; Zhang and Chen, 2014;  
Waggoner and Chen, 2014; Kamble *et al.*, 2015;  
Kong *et al.*, 2016; Shnayder *et al.*, 2016]. In order to motivate each agent to reveal his private, informative signal, peer-prediction methods offer a reward based on how each agent's reports compare with those of his peers. Such rewards are designed to induce truth telling in equilibrium—that is, they create a situation in which each agent has an interest in investing effort and revealing his private and informative signal truthfully, as long as he believes that all other agents will do the same.

Even if they offer a truthful equilibrium, peer-prediction methods also always induce other uninformative equilibria, the existence of which is inevitable [Jurca and Faltings, 2009; Waggoner and Chen, 2014]. Intuitively, if every other agent's strategy does not depend on their private information, a given agent does not want to deviate from this strategy either: agents can only be rewarded for coordination, not for accuracy. When private information is costly to obtain, uninformative equilibria are typically *less* demanding for agents to play. This raises significant doubt about whether peer-prediction methods can motivate truthful reporting in practice. Experimental evaluations of peer-prediction methods had mixed results. Some studies showed that agents reported truthfully [John *et al.*, 2012; Faltings *et al.*, 2014; Radanovic *et al.*, 2016]; another found that agents colluded on uninformative equilibria [Gao *et al.*, 2014].

Recent progress on peer-prediction mechanisms has focused on making the truthful equilibrium Pareto dominant, i.e., (weakly) more rewarding to every agent than any other equilibrium [Dasgupta and Ghosh, 2013; Witkowski and Parkes, 2013; Kamble *et al.*, 2015; Radanovic and Faltings, 2015; Shnayder *et al.*, 2016]. This can be achieved by rewarding agents based on the distributions of their reports for multiple objects. However, we show in this paper that such arguments rely critically on the assumption that every agent has access to only one private signal per object. This is often untrue in practice; e.g., in peer grading, by taking a quick glance at an essay a student can observe characteristics such as length, formatting and the prevalence of grammatical errors. These characteristics require hardly any effort to observe, can be arbitrarily uninformative about true quality, and are of no interest to the mechanism. Yet their existence provides a means for the agents to coordinate. We build on this intuition to prove that no mechanism can guarantee that an equilibrium in which all agents truthfully report their informative signals is always Pareto dominant in every setting.

Motivated by these negative results, we move on to consider a setting in which the operator of the mechanism has access to trusted evaluators (e.g., teaching assistants) who can reliably provide noisy but informative signals of the object's true quality. This allows for a hybrid mechanism that blends peer-prediction with comparison to trusted reports. With a fixed probability, the mechanism obtains a

trusted report and rewards the agent based on the agreement between the agent's report and the trusted report [Jurca and Faltings, 2005]. Otherwise, the mechanism rewards the agent using a peer-prediction mechanism. Such hybrid mechanisms can yield stronger incentive guarantees than other peer-prediction mechanisms, such as achieving truthful reporting of informative signals in Pareto-dominant equilibrium [Jurca and Faltings, 2005; Dasgupta and Ghosh, 2013]. Intuitively, if an agent seeks to be close to a trusted report, then his best strategy is to reveal his informative signal truthfully.

In fact, the availability of trusted reports is so powerful that it gives us the option of dispensing with peer-prediction altogether. Specifically, we can reward students based on agreement with the trusted report when the latter is available, but simply give students a constant reward otherwise, in an approach we dub the *peer-insensitive mechanism*. Indeed, in [Wright *et al.*, 2015] we introduced such a peer grading system and showed that it worked effectively in practice, based on a study across three years of a large class. This mechanism has even stronger incentive properties than the hybrid mechanism—because it induces a single-agent game, it can give rise to dominant-strategy truthfulness.

Our paper's main focus is on comparing these two approaches in terms of the number of trusted reports that they require. One might expect that the hybrid approach would have the edge, both because it relies on a weaker solution concept and because it leverages a second source of information reported by other agents. Surprisingly, we prove that this intuition is backwards. We identify a simple sufficient condition, which, if satisfied, guarantees that the peer-insensitive mechanism offers the dominant strategy of truthful reporting of informative signals while querying trusted reports with a lower probability than is required for a peer-prediction mechanism to motivate truthful reporting in Pareto-dominant equilibrium. We then show that all applicable peer-prediction mechanisms of which we are aware satisfy this sufficient condition. This means that when we seek to make untruthful equilibria either Pareto-dominated or to eliminate these equilibria entirely, we require fewer trusted reports in the simple, peer-insensitive mechanism than in the hybrid mechanism derived from any of these existing peer-prediction mechanisms.

## 2 Model

A mechanism designer wishes to elicit information about the quality of a set of objects. Each object has a latent quality. There are multiple rational, risk-neutral agents. Each agent evaluates a subset of the objects.

Agents have access to private information about the qualities of the objects of interest, and our goal is to motivate the agents to reveal their private information. In the peer prediction literature, it is standard to assume that each agent receives a single, private signal, which represents the *only* information that agent has about the object of interest. We argue that, in reality, every agent can obtain multiple pieces of information about the object's quality by investing different amounts of efforts. To capture this, we consider a simplified scenario by assuming that, for each object, an agent has

access to two pieces of private information: a *high-quality signal* and a *low-quality signal*.

The high-quality signal refers to a useful piece of information that the mechanism designer wishes to elicit. It is drawn from a distribution conditional on the object’s quality and the identity of the agents evaluating the object. The joint distributions of the high-quality signals are common knowledge among the agents. We assume that obtaining the high-quality signal requires a constant effort  $c^E > 0$ .

The low-quality signal represents information about superficial qualities of the object—it is correlated with the quality of the object, but not sufficiently so. We assume that obtaining the low-quality signal requires no effort, though our results can be straightforwardly extended to the case where the difference between the cost of obtaining the two signals is  $c^E$ . We assume that the mechanism designer prefers to get the high-quality signals rather than the low-quality signals because the high-quality signals are more correlated with the quality of the object than the low-quality signals. However, the low-quality signals are easier to obtain than the high-quality signals because the low-quality signals of different agents are more correlated with each other than the high-quality signals of these agents. As a result, the low-quality signals provide an easier way for agents to coordinate their reports compared to investing costly effort to acquire the high-quality signal.

The low-quality signal is a collection of properties of the object that an agent can observe with negligible effort. For example, by glancing at an essay and skimming several sentences, an agent can observe several superficial attributes of the essay, such as the length of the essay, whether the author provided references or not, the number of spelling and grammatical mistakes, the sentence structure, the vocabulary, and the complexity of the language being used.<sup>1</sup> Similarly, one could base a restaurant review on the decor without ordering food; evaluate a movie by watching its trailer; etc.

Agents may strategize over both whether to incur the cost of effort to observe the high-quality signal and which signal and belief reports to make. The goal of the mechanism designer is to incentivize each agent to incur the cost of effort to obtain the high-quality signal, to report the high-quality signal as his signal report, and to report his posterior belief conditional on the high-quality signal as his belief report. A mechanism has a *truthful equilibrium* when it is a Bayesian Nash equilibrium for every agent to use this strategy.

We focus on *universal* peer-prediction mechanisms.

**Definition 1** (Universal peer-prediction mechanism). *A peer-prediction mechanism is universal if it can be operated without knowledge of the joint distribution of the high-quality signals (i.e., it is “detail free” [Wilson, 1987]) and guarantees the existence of the truthful equilibrium for any number of agents  $n \geq 3$  and any number of tasks.*

We focus on this class of mechanisms for two reasons. First, in practice, it is not always possible for a mechanism

<sup>1</sup>In a large class, it might be unlikely that all the students could agree on which superficial attribute of an essay they should coordinate on. However, by combining multiple superficial attributes together, the students could arrive at similar estimates.

designer to have detailed knowledge of the joint signal distribution, so this allows us to focus on mechanisms that are more likely to be used in practice. Second, it is relatively unrestrictive, as many peer-prediction mechanisms in the literature satisfy universality.

### 3 Impossibility of Pareto-Dominant, Truthful Elicitation

We show that when agents have access to multiple signals about an object, Pareto-dominant truthful elicitation is impossible for any universal mechanism that computes agent rewards solely based on agents’ reports without any access to ground truth. Intuitively, without knowledge of the distributions from which the signals are drawn, the mechanism cannot distinguish the signal that it hopes to elicit from other, irrelevant signals.

**Theorem 1.** *For any universal mechanism, if obtaining the high-quality signals requires an additional cost of effort of  $c > 0$  compared to obtaining any low-quality signal, then there exists a multi-signal environment in which the truthful equilibrium is not Pareto dominant.*

Any universal mechanism does not have access to the joint distributions of the signal. Therefore, with multiple signals, there is no way for a universal mechanism to ensure that the truthful equilibrium yields the highest utility for the agents. The truthful equilibrium is Pareto dominant only if the high-quality signal happens to be drawn from a distribution yielding sufficiently higher reward than every other signal to compensate for the cost of effort.

One way for the mechanism designer to ensure that agents are reporting the high-quality signal is to stochastically compare agents’ reports to ground truth. In the next section, we introduce a class of mechanisms that takes this approach.

### 4 Combining Elicitation with Limited Access to Ground Truth

Elicitation mechanisms are designed for situations where it is infeasible for the mechanism designer to evaluate each object herself. However, in practice, it is virtually always possible, albeit costly, to obtain unbiased evaluations of a subset of the objects, which we will call *trusted reports*. In the peer grading setting, the instructor and teaching assistants can mark some of the assignments. Similarly, review sites could hire an expert to evaluate restaurants or hotels that its users have reviewed; and so on.

In this paper, we study a class of mechanisms—*spot checking mechanisms*—that take advantage of this limited access to ground truth to circumvent the result from Section 3. The mechanism performs a *spot check* on each object with some probability. When a spot check is performed, the mechanism obtains a trusted report  $s_j^t$ , which is an unbiased estimate of the object’s quality, and rewards each agent by comparing the agent’s signal and belief reports with the trusted report.

Specifically, we compare two approaches to using limited access to ground truth for elicitation. The first approach is to augment an existing peer-prediction mechanism with spot-checking; we call this kind of mechanism a *spot-checking*

*peer-prediction mechanism*. The second approach is to rely exclusively on ground truth access to incentivize truthful reporting, using what we call a *peer-insensitive mechanism*. When a spot check is performed, the peer-insensitive mechanism rewards each agent by comparing their signal report to the trusted report; when a spot check is not performed, each agent receives a constant payment.

## 5 When Does Peer Prediction Help?

We compare the peer-insensitive mechanism with the class of universal spot-checking peer-prediction mechanisms. In Lemma 1, we derive an expression for the minimum spot-check probability  $p_{ds}$  at which the truthful strategy is a dominant strategy for the peer-insensitive mechanism. When the spot-check probability is  $p_{ds}$ , any agent is indifferent between investing effort by reporting truthfully and playing their best strategy conditional on investing no effort when the object is spot checked.

**Lemma 1.** *The minimum spot-check probability  $p_{ds}$  at which the truthful strategy is dominant for the peer-insensitive mechanism satisfies the following equation.*

$$p_{ds} \mathbb{E}[y(s^h, s^t)] - c^E = p_{ds} \mathbb{E}[y(g^l(s^l), s^t)]. \quad (1)$$

Next, we derive a lower bound for  $p_{\text{Pareto}}$ , the minimum spot-check probability at which the truthful equilibrium is Pareto dominant in a universal, spot-checking peer-prediction mechanism. Theorem 2 states that, if a simple sufficient condition is satisfied, then compared to all universal spot-checking peer-prediction mechanisms, the peer-insensitive mechanism can achieve stronger incentive properties (dominant-strategy truthfulness versus Pareto dominance of truthful equilibrium) while requiring a smaller spot-check probability.

**Theorem 2** (Sufficient condition for Pareto comparison). *For any spot-checking peer-prediction mechanism, if the low-effort equilibrium exists and Pareto dominates the truthful equilibrium when the cost of effort  $c^E = 0$  and the probability of spot-checking  $p = 0$ , then  $p_{\text{Pareto}} \geq p_{ds}$  for all  $c^E \geq 0$ .*

Finally, we show that, under very natural conditions, every universal peer-prediction mechanism of which we are aware in the literature satisfies the conditions of Theorem 2; hence, in this setting, the peer-insensitive spot-checking mechanism requires less ground truth access than any spot-checking peer-prediction mechanism.

**Corollary 1.** *For spot-checking peer-prediction mechanisms based on [Faltings et al., 2012; Witkowski et al., 2013; Dasgupta and Ghosh, 2013; Waggoner and Chen, 2014; Kamble et al., 2015; Radanovic and Faltings, 2015] and [Shnayder et al., 2016], the minimum spot-check probability  $p_{\text{Pareto}}$  for the Pareto dominance of the truthful equilibrium is greater than or equal to the minimum spot-check probability  $p_{ds}$  at which the truthful strategy is a dominant strategy for the peer-insensitive mechanism.*

**Corollary 2.** *For spot-checking peer-prediction mechanisms based on [Witkowski and Parkes, 2012; Witkowski and Parkes, 2013; Radanovic and Faltings, 2013;*

*Radanovic and Faltings, 2014] and [Riley, 2014], if the peer-prediction mechanism uses a symmetric proper scoring rule, then the minimum spot-check probability  $p_{\text{Pareto}}$  for the Pareto dominance of the truthful equilibrium is greater than or equal to the minimum spot-check probability  $p_{ds}$  at which the truthful strategy is a dominant strategy for the peer-insensitive mechanism.*

This surprising finding is intuitive in retrospect. Peer-prediction mechanisms can only motivate agents to behave in a certain way as a group. An agent has a strong incentive to be truthful if all other agents are truthful; conversely, when all other agents coordinate on investing no effort, the agent again has a strong incentive to coordinate with the group. Peer-prediction mechanisms thus need to provide a strong enough incentive for agents to deviate from the most attractive uninformative equilibrium in the worst case, whereas the peer-insensitive mechanism only needs to motivate effort and honesty in what is effectively a single-agent setting.

## 6 Conclusions and Future Work

We consider the problem of using limited access to noisy but unbiased ground truth to incentivize agents to invest costly effort in evaluating and truthfully reporting the quality of some object of interest. Without such spot-checking, peer-prediction mechanisms already guarantee the existence of a truthful equilibrium that induces both effort and honesty from the agents. However, this truthful equilibrium may be less attractive to the agents than other uninformative equilibria.

Existing mechanisms ensure that the truthful equilibrium is the most attractive (i.e., Pareto dominant) equilibrium to the agents. However, these mechanisms rely crucially on the unrealistic assumption that agents' only means of correlating are via the signals that the mechanism aims to elicit. We show that when agents have access to more than one signal, no universal peer-prediction mechanism has a Pareto-dominant truthful equilibrium in all settings.

In contrast, we present a simpler, peer-insensitive mechanism that provides incentives for effort and honesty only by checking the agents' reports against ground truth. While one might expect that peer-prediction would require less frequent access to ground truth to achieve stronger incentive properties than the peer-insensitive mechanism, we proved the opposite for all universal spot-checking peer-prediction mechanisms.

Many exciting future directions remain to be explored. For example, we assumed that the principal does not care about the total amount of the artificial currency rewarded to the agents. One possible direction would consider a setting in which the principal seeks to minimize both spot checks and the agents' rewards. Also, in our analysis, we assumed that the spot-check probability does not depend on the agents' reports. Conditioning the spot-check probability on the agents' reports might allow the mechanism to more efficiently detect and punish uninformative equilibria. We are particularly excited about designing more sophisticated spot check mechanisms where the spot-check probability is a function of the set of reports for a particular submission. (We show how this can work in the case of binary signals in very recent work [Zarkoob et al., 2019].) In addition, we are interested in ex-

ploring the scenario in which some agents are altruistic and always invest the effort to obtain the high-quality signal.

## References

- [Dasgupta and Ghosh, 2013] Anirban Dasgupta and Arpita Ghosh. Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd International Conference on the World Wide Web*, pages 319–330, 2013.
- [de Alfaro *et al.*, 2015] Luca de Alfaro, Vassilis Polychronopoulos, and Michael Shavlovsky. Incentives for truthful peer grading. *UC Santa Cruz Technical Report*, 2015.
- [Faltings *et al.*, 2012] Boi Faltings, Jimmy J Li, and Radu Jurca. Eliciting truthful measurements from a community of sensors. In *3rd International Conference on the Internet of Things (IOT)*, pages 47–54. IEEE, 2012.
- [Faltings *et al.*, 2014] Boi Faltings, Radu Jurca, Pearl Pu, and Bao Duy Tran. Incentives to counter bias in human computation. In *Second AAAI Conference on Human Computation and Crowdsourcing*, 2014.
- [Gao *et al.*, 2014] Xi Alice Gao, Andrew Mao, Yiling Chen, and Ryan Prescott Adams. Trick or treat: putting peer prediction to the test. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, pages 507–524. ACM, 2014.
- [Gao *et al.*, 2019] Xi Alice Gao, James R Wright, and Kevin Leyton-Brown. Incentivizing evaluation with peer prediction and limited access to ground truth. *Artificial Intelligence*, 275:618–638, 2019.
- [John *et al.*, 2012] Leslie K John, George Loewenstein, and Drazen Prelec. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5):524–532, 2012.
- [Jurca and Faltings, 2005] Radu Jurca and Boi Faltings. Enforcing truthful strategies in incentive compatible reputation mechanisms. *Internet and Network Economics*, pages 268–277, 2005.
- [Jurca and Faltings, 2009] Radu Jurca and Boi Faltings. Mechanisms for making crowds truthful. *Journal of Artificial Intelligence Research*, 34(1):209–253, 2009.
- [Kamble *et al.*, 2015] Vijay Kamble, Nihar Shah, David Marn, Abhay Parekh, and Kannan Ramachandran. Truth serums for massively crowdsourced evaluation tasks. *arXiv preprint*, arXiv:1507.07045, 2015.
- [Kong *et al.*, 2016] Yuqing Kong, Katrina Ligett, and Grant Schoenebeck. Putting peer prediction under the micro(economic)scope and making truth-telling focal. *arXiv preprint*, arXiv:1603.07319, 2016.
- [Kulkarni *et al.*, 2014] Chinmay E Kulkarni, Richard Socher, Michael S Bernstein, and Scott R Klemmer. Scaling short-answer grading by combining peer assessment with algorithmic scoring. In *Proceedings of the First ACM Conference on Learning @Scale*, pages 99–108, 2014.
- [Miller *et al.*, 2005] Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.
- [Prelec, 2004] Dražen Prelec. A Bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.
- [Radanovic and Faltings, 2013] Goran Radanovic and Boi Faltings. A robust Bayesian truth serum for non-binary signals. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence, AAAI 2013*, pages 833–839, 2013.
- [Radanovic and Faltings, 2014] Goran Radanovic and Boi Faltings. Incentives for truthful information elicitation of continuous signals. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.
- [Radanovic and Faltings, 2015] Goran Radanovic and Boi Faltings. Incentives for subjective evaluations with private beliefs. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [Radanovic *et al.*, 2016] Goran Radanovic, Boi Faltings, and Radu Jurca. Incentives for effort in crowdsourcing using the peer truth serum. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(4):1–28, 2016.
- [Raman and Joachims, 2014] Karthik Raman and Thorsten Joachims. Methods for ordinal peer grading. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1037–1046. ACM, 2014.
- [Riley, 2014] Blake Riley. Minimum truth serums with optional predictions. In *Proceedings of the 4th Workshop on Social Computing and User Generated Content (SC14)*, 2014.
- [Shnayder *et al.*, 2016] Victor Shnayder, Arpit Agarwal, Rafael Frongillo, and David C Parkes. Informed truthfulness in multi-task peer prediction. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 179–196. ACM, 2016.
- [Waggoner and Chen, 2014] Bo Waggoner and Yiling Chen. Output agreement mechanisms and common knowledge. In *Proceedings of the 2nd AAAI Conference on Human Computation and Crowdsourcing*, 2014.
- [Wilson, 1987] R. Wilson. Game-theoretic approaches to trading processes. In *Advances in Economic Theory: Fifth World Congress*, pages 33–77, 1987.
- [Witkowski and Parkes, 2012] Jens Witkowski and David C Parkes. A robust Bayesian truth serum for small populations. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI’12)*. Association for the Advancement of Artificial Intelligence, 2012.
- [Witkowski and Parkes, 2013] Jens Witkowski and David C Parkes. Learning the prior in minimal peer prediction. In *Proceedings of the 3rd Workshop on Social Computing and User Generated Content at the ACM Conference on Electronic Commerce*, pages 39:1–39:12, 2013.
- [Witkowski *et al.*, 2013] Jens Witkowski, Yoram Bachrach, Peter Key, and David Christopher Parkes. Dwelling on the negative: Incentivizing effort in peer prediction. In *Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing*, 2013.
- [Wright *et al.*, 2015] James R Wright, Chris Thornton, and Kevin Leyton-Brown. Mechanical TA: Partially automated high-stakes peer grading. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, pages 96–101, 2015.
- [Zarkoob *et al.*, 2019] Hedayat Zarkoob, Hu Fu, and Kevin Leyton-Brown. Report-sensitive spot-checking in peer grading systems. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 2306–2308. International Foundation for Autonomous Agents and Multiagent Systems, 2019.
- [Zhang and Chen, 2014] Peter Zhang and Yiling Chen. Elicitability and knowledge-free elicitation with peer prediction. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multiagent Systems*, pages 245–252. International Foundation for Autonomous Agents and Multiagent Systems, 2014.