

# Harvard University, John A. Paulson School of Engineering and Applied Sciences

## CS236r Project Proposal Computational Results For Peer-Prediction and Crowdsourced Judgement Elicitation

Virgile Audi (vaudi@g.harvard.edu)  
Charles Liu (cliu02@g.harvard.edu)

April 13, 2016

### 1 Motivation

Reading the numerous papers for this course left us with the same interrogation: how do the assumptions made to obtain clean and useful results hold when it comes to such implementing mechanisms in real life? This was particularly the case when working on our presentation that focused on Peer-Prediction and Crowd-Sourcing mechanisms. Our project will focus primarily on the “Crowdsourced Judgement Elicitation with Endogenous Proficiency” paper by Dasgupta and Ghosh [1] and “Eliciting Informative Feedback: The Peer-Prediction Method” by Miller, Resnick and Zeckhauser [2]. Although the papers provided very strong results, they required many assumptions to prove their findings. Many of these assumptions seemed to us unrealistic, as well as some results that followed. For our final project we would therefore like to relax some of those assumptions and see if similar results are met. We plan on developing a Python package that can simulate over the number of tasks/participants and estimate expected rewards for participants, using various experimental designs.

We will now present some of the assumptions we plan on relaxing and test the validity of the results under these new assumptions.

### 2 Crowd-Sourcing: how to better model Effort and Proficiency?

In the Crowd-Sourcing paper, two of the central assumptions made are:

- Effort is a binary decision - either 0 effort where you guess the correct result based off a coin flip or you give full effort and guess the correct result based off your proficiency
- Your proficiency for obtaining the correct result is greater than .5

The second assumption is central to proving that giving full effort is an equilibrium (Lemma 4). By relaxing that assumption, this should no longer be optimal for each participant. There are three scenarios to consider here - participant’s own proficiency  $< .5$ , participant’s expected reference rater’s proficiency  $< .5$ , and both.

One of the stronger statements we found the paper to make was Lemma 8:

“Suppose the probability of agent  $i$  using strategy  $(1, X)$  is  $\delta$  and strategy  $(0, r_i)$  is  $1 - \delta$  for each task  $j \in J(i)$ . Suppose  $i$ ’s potential reference raters  $r_j(i)$  use strategies  $(1, X)$  and  $(0, r_{r_j}(i))$  with probabilities  $\epsilon_{r_j}(i)$  and  $1 - \epsilon_{r_j}(i)$  respectively, for each task  $j \in J(i)$ . If  $\epsilon_{r_j}(i) > 0$  for any reference rater with proficiency  $p_{r_j}(i) > \frac{1}{2}$ , then agent  $i$  has a (strict) profitable deviation to  $\delta' = 1$ , i.e., to always using strategy  $(1, X)$ , for all values of  $r_i \in [0, 1]$ ”

If the rater has any chance of giving full effort, then the optimal strategy is to give full effort. Aside from the proficiency being at least .5, this is reliant on the idea that effort is a binary measure. Another way of looking at this is that the proficiency of a participant is some function of effort, where effort is over a distribution. In our implementation, we will specify some prior on effort that would be used as a parameter in our proficiency distribution.

The paper also notes:

“We assume that the maximum proficiency  $p_i \geq \frac{1}{2}$  for all  $i$  this minimum requirement on agent ability can be ensured in online crowdsourcing settings by prescreening workers on a representative set of tasks (Amazon Mechanical Turk, for instance, offers the ability to prescreen workers [13, 4], whereas in peergrading applications such as on Coursera, students are given a set of pre-graded assignments to measure their grading abilities prior to grading their peers, the results of which can be used as a prescreen.)”

We believe this is a very large assumption - in many cases, like the product review setting that was used in the peer prediction paper, the quality of reviewers may not be known. This leads to relaxing different assumptions - on the one hand clearly proficiency, but also the prior on receiving some signal.

### 3 Peer-Prediction

Peer-Prediction relies on many assumptions such as:

- Raters are risk neutral, i.e., that maximizing the expected transfer is equivalent to maximizing expected utility,
- Conditional on the products type, raters signal are independent and identically distributed, i.e. raters have common prior
- Raters have common knowledge of the probability of reporting a signal given a type of a product.

The paper addresses the practical issues that come when comparing these assumptions to reality but in a very swift manner. We would be particularly interested in looking in greater depths at the validity of the results presented under the relaxation of these assumption. We plan on implementing the mechanisms explained in “Peer Prediction without a Common Prior” by Witkowski and Parkes [3] for settings where the agents hold subjective and private beliefs about the state of the world and the likelihood of a positive signal given a particular state.

Relaxing these assumptions could also help answer another issue mentioned in [2] which is which scoring rule to choose?

### 4 Help from the Staff

The relaxation of the common prior assumption is addressed numerously in the literature. We would particularly appreciate the possibility of discussing Professor Chen’s paper “Trick or Treat: Putting Peer Prediction to the Test” [4] with her and especially the design of her experiments (and possibly have access to the data?). If the common prior assumption is discussed abundently, we could not find papers on the relaxation of the risk neutral assumption, any help from the staff on this issue would be appreciated as well.

### 5 References

1. *Crowdsourced Judgement Elicitation with Endogenous Proficiency*, A. Dasgupta & A. Ghosh ([http://www.arpitaghosh.com/papers/elicit\\_arxiv\\_2.pdf](http://www.arpitaghosh.com/papers/elicit_arxiv_2.pdf) )
2. *Eliciting Informative Feedback: The Peer-Prediction Method*, N. Miller, P. Resnick, & R. Zeckhauser (<http://www.hks.harvard.edu/fs/rzeckhau/elicit.pdf> )
3. *Peer Prediction without a Common Prior*, J. Witkowski, D. Parkes ([http://www.eecs.harvard.edu/econcs/pubs/witkowski\\_ec12.pdf](http://www.eecs.harvard.edu/econcs/pubs/witkowski_ec12.pdf) )
4. *Trick or Treat: Putting Peer Prediction to the Test*, X. A. Gao, A. Mao & Y. Chen ([http://0www.eecs.harvard.edu/econcs/pubs/Gao\\_acm13.pdf](http://0www.eecs.harvard.edu/econcs/pubs/Gao_acm13.pdf) )