Incentive-Compatible Trust Mechanisms (Extended Abstract)

Jens Witkowski
Department of Computer Science
Albert-Ludwigs-Universität
Freiburg, Germany
witkowsk@informatik.uni-freiburg.de

The Problem: Trust in Online Markets

Trading goods online has numerous advantages. One that is particularly compelling is that online merchants can offer their goods at lower prices compared to their offline counterparts. The physical distance between buyers and sellers, however, also leads to problems of trust. Consider the online auction site eBay as an example: its procedure is such that the winning bidder (henceforth the *buyer*) first pays for the good and that the seller is required to send the good only after receipt of this payment. Without any trust-enabling mechanisms in place, the seller is best off keeping the good for himself, even if he received the payment. Since a rational, self-interested buyer can anticipate this, she will not pay for the good in the first place and no trade takes place. This trust problem is usually addressed by a reputation mechanism that publishes buyer feedback about a seller's past behavior [e. g., Dellarocas, 2006]. Reputation mechanisms, however, critically rely on assumptions that are rarely met in real-world marketplaces: first, it is assumed that buyers honestly report their private experiences. Second, it is assumed that the seller cannot *whitewash*, i.e. create a new reputation profile once an old one is ran down, and, third, it is assumed that the seller is long-lived, i.e. that he will continue to trade on the marketplace indefinitely. Consider eBay as an example for a real-world market again. From a game-theoretic point of view, the assumption of honest buyer feedback is problematic for two reasons: first, since reporting feedback is time-consuming and thus costly, a buyer has no incentive to leave feedback at all. Second, when feedback is published, there are ample incentives for manipulation, such as a competitor degrading a seller's reputation profile to push him out of the market. The assumption that sellers cannot whitewash is equally difficult to uphold since it is easy to create a new identity and start anew. Lastly, not all sellers are long-lived which—together with the whitewashing problem—creates a situation where new sellers cannot enter the market.

In my thesis, I design incentive-compatible trust mechanisms that do not rely on any of the aforementioned assumptions. Moreover, I focus on designs that minimize common knowledge assumptions with respect to the players' valuations, costs and beliefs.

Progress to Date

I divide my explanations of the progress to date in two parts: first, I explain how to elicit truthful buyer feedback for trust mechanisms, such as the reputation mechanism employed by eBay. In the second part, I present *escrow mechanisms*, a new class of trust mechanisms, that avoid the assumption of long-lived sellers and remove the whitewashing problem.

Elicitation of Truthful Feedback One way to elicit truthful buyer feedback in product opinion forums, such as Amazon Reviews, is the so-called *peer prediction method* by Miller, Resnick and Zeckhauser (2005) who propose to pay a buyer for her feedback report conditional on the report of another buyer. The comparison of two buyer reports is meaningful as the product's quality is the same for all customers. Consider a digital camera that is bought from Amazon as an example: while different customers may have different experiences due to noise, they all receive the identical model.

This is different for the buyers' experiences at eBay since these primarily depend on the seller's actions, i. e. if the seller sent the good in the prescribed quality. Since a seller can vary this choice from one buyer to the other, it is no longer guaranteed that two buyers' experiences are essentially the same. In our paper *Truthful Feedback for Sanctioning Reputation Mechanisms* (Witkowski 2010), we study the mechanism design space of peer-prediction-based feedback elicitation in the eBay context with strategic sellers. Our results are twofold: first, we prove that it is impossible to elicit truthful buyer feedback in the basic setting with only strategic sellers. Second, we draw on the game-theoretic literature on reputation building and study a model with two types of sellers: a "normal" strategic type and a cooperative commitment type. For this two-type setting, we then show that any positive prior belief for the commitment type allows for a peer-prediction-based payment scheme that elicits truthful buyer feedback.

Incentive-Compatible Escrow Mechanisms In our paper *Incentive-Compatible Escrow Mechanisms* (Witkowski *et al.* 2011), we introduce a new class of trust mechanisms. These escrow mechanisms are "history-free" in that they do not rely on the publication of reported feedback. This improves on the state-of-the-art because it avoids the assumption of long-lived sellers and removes the whitewashing problem. The main idea is that a buyer does not pay the

seller directly but through a trusted third party (the *center*). Once the seller has sent the good, the center asks the buyer for her feedback report and forwards the payment to the seller only if the buyer acknowledges the receipt of the good. The key question is how to proceed with the withheld payments following a negative report. If the center reimbursed every buyer who reports negatively, a rational buyer would give a negative report even if she was satisfied. To avoid this, our escrow mechanism matches two buyers and uses the report of one buyer to determine whether the other buyer receives a payback. We show that this mechanism is incentive compatible, efficient, interim individually rational and ex ante budget balanced. We address collusion by matching buyers from different sellers, so that, in large markets like eBay, the chances for two colluders to be matched with one another are very small. Consequently, the expected utility for collusion is negative even under the assumption of minimal coordination costs. Moreover, and in contrast to previous work on trust and reputation, our approach does not rely on knowing the sellers' cost functions or the distribution of buyer valuations.

Proposed Plan for Research

Until my graduation in May 2013, I will continue to work on both the elicitation of truthful feedback and the design of incentive-compatible escrow mechanisms.

Elicitation of Truthful Feedback A major drawback of the peer prediction method are its strong common knowledge assumptions. Consider again a digital camera bought from Amazon: the method's first assumption is that every buyer has the same prior belief about the camera's true quality. For example, the buyers may believe that the camera is of high quality with a probability of 70%. Second, once a buyer receives the camera, she experiences a noisy signal of its true quality and it is assumed that the probability for a particular signal given a particular quality is the same for all buyers. For example, the buyers may believe that they have a positive experience with a probability of 90% and 30% if the camera is of high and low quality, respectively.

These assumptions on beliefs entail two points: first, that every buyer has identical beliefs and, second, that the mechanism knows them. In future work, we will design a mechanism that does not rely on either of these points. The basic idea is to take advantage of the fact that right after ordering a good, the buyer cannot have experienced it yet which allows the mechanism to ask for two probabilistic belief reports: one before the buyer receives the good and another one after she has experienced it. By eliciting these two beliefs, the mechanism can infer the buyer's experience as it is reflected in the belief change: if the second belief report is higher than the first, the experience must have been good and vice versa. Please note that while the second belief is more accurate, it is not sufficient to elicit only the second belief report. The reason is that the inference of the buyer's experience is required to condition the other buyers' payments. Also note that a property of this mechanism is that buyers have to report probabilities. We will, however, "hide" them and employ a user interface with a point scale from 0 to 10. These points directly correspond to probabilities but instead of asking for probability reports, it allows buyers to interact with the system in a way they are familiar with from other online rating sites.

Incentive-Compatible Escrow Mechanisms The general escrow mechanism technique is applicable to a wide class of settings. A market that is particularly in need for a trust mechanism is the crowdsourcing platform Amazon Mechanical Turk. On this platform, people are paid small rewards to do human computation tasks, such as annotating images. A particularity of this market is that the verification of a task, i.e. to learn whether the task was duly completed, is costly. At the same time, the fact that tasks are digital allows for an external verification, i.e. the person who posted the task (the *requester*) can ask others to verify it. It is therefore not uncommon that the verification of task is itself made a task. In fact, for every original task, a requester usually creates two verification tasks and asks other workers to vote if the original task was properly executed. Unfortunately, this scheme does not properly incentivize effort, and fraudulent behavior is a major problem. In future work, we will therefore develop an escrow mechanism for this market that provides proper incentives, and increases efficiency by reducing the number of necessary verification tasks. In addition to escrowing the payments, we incentivize verification workers by comparing their votes in a way that is similar to the peer prediction method. Once we have designed the mechanism and proven its theoretical properties, we will run an experiment with different designs on Amazon Turk itself. We are particularly interested to study the optimal trade-off between the mechanism's theoretical properties and the cognitive costs incurred by the respective level of complexity.

Acknowledgements

I want to thank my coauthors for inspiring collaboration. My special thanks go to David Parkes for hosting me at Harvard University, where most of this research agenda developed, and my advisor Bernhard Nebel for his support and positive attitude towards my plans and ideas. I am also grateful for financial support through PhD fellowships from the Landesgraduiertenförderung Baden-Württemberg and the German Academic Exchange Service.

References

Chrysanthos Dellarocas. Reputation Mechanisms. In Terry Hendershott, editor, *Handbook on Information Systems and Economics*. Elsevier Publishing, 2006.

Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting Informative Feedback: The Peer-Prediction Method. *Management Science*, 51(9):1359–1373, 2005.

Jens Witkowski, Sven Seuken, and David Parkes. Incentive-Compatible Escrow Mechanisms. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI'11)*, 2011.

Jens Witkowski. Eliciting Honest Reputation Feedback in a Markov Setting. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, 2009.

Jens Witkowski. Truthful Feedback for Sanctioning Reputation Mechanisms. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI'10)*, 2010.