Incentive-Compatible Escrow Mechanisms

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

Sven Seuken
School of Eng. & Applied Sciences
Harvard University
Cambridge, MA, USA
seuken@eecs.harvard.edu

David C. Parkes
School of Eng. & Applied Sciences
Harvard University
Cambridge, MA, USA
parkes@eecs.harvard.edu

Abstract

The most prominent way to establish trust between buyers and sellers on online auction sites are reputation mechanisms. Two drawbacks of this approach are the reliance on the seller being long-lived and the susceptibility to whitewashing. In this paper, we introduce so-called *escrow mechanisms* that avoid these problems by installing a trusted intermediary which forwards the payment to the seller only if the buyer acknowledges that the good arrived in the promised condition. We address the incentive issues that arise and design an escrow mechanism that is incentive compatible, efficient, interim individually rational and ex ante budget-balanced. 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.

Introduction

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 as the costs for running a physical store are higher. This physical distance between buyer and seller, however, also leads to trust problems. Consider the online auction site eBay as an example: its procedure is such that the winning bidder (henceforth: 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, whether or not he received the payment. Since a rational, self-interested buyer can anticipate this, she will not send payment and no trade takes place.[1]

Online marketplaces such as eBay address this trust problem with a reputation mechanism that publishes buyer feedback about a seller's past behavior. From a game-theoretic point of view this is justified by the seminal work on reputation building by Fudenberg and Levine (1989). They assume a long-lived player (seller) facing a sequence of short-lived players (buyers). In each transaction, the buyer decides whether or not to send payment followed by the decision of the seller whether or not to send the good, where future potential buyers observe the seller's actions. Fudenberg and Levine show that reputation effects can lead to sellers and buyers eventually cooperating such that all beneficial transactions take place. These positive results critically rely on three assumptions: first, the seller's actions are publicly observed by all following buyers; second, the seller is long-lived, i.e. he will continue to trade on the marketplace indefinitely; and, third, the seller cannot *whitewash*, i.e. create a new reputation profile once an old one is ran down.

However, these assumptions are rarely met in real-world reputation mechanisms. Buyers' experiences are not public on eBay. Rather, the reputation mechanism relies on buyer feedback. When feedback is published, there are ample reasons for manipulation as, for example, a competitor degrading a seller's reputation profile to push him out of the market, or a buyer who wants to lower prices for future purchases. Furthermore, in settings with bidirectional feedback, retaliation can be a serious problem (e. g., Bolton, Greiner, and Ockenfels, 2011). It is also not true that all sellers are in the market long enough to be incentivized by future returns that are dependent on today's feedback. Moreover, it is very easy to create a new identity and whitewash. It is thus not surprising that eBay looks for alternative ways to establish trust in their market as witnessed by their introduction of the "eBay Buyer Protection" in September 2010. eBay makes a decision about a claim based, in part, on the buyer's and seller's transaction histories. Unfortunately, they introduce a new problem: now buyers sometimes have an incentive to falsely report that they didn't receive the good.

In this paper, we introduce *escrow mechanisms*, in which a trusted intermediary first receives payment from a buyer, and then forwards the payment to the seller only if the buyer acknowledges that the good arrived in the promised condition. In the particular solution we present, buyers sometimes receive a monetary rebate. It is important not to think about this rebate as a reimbursement. Instead, the escrow mechanism is designed to reward buyers for leaving truthful feedback about seller behavior. This in turn incentivizes sellers to cooperate, since otherwise they do not receive their payments, and promotes efficiency while ensuring budget balance. Furthermore, we achieve incentive compatibility without any common knowledge assumptions that we consider problematic in applications. Moreover, escrow mechanisms are "history-free," i.e. they do not rely on the publication of reported feedback. This improves on the state-of-the-art

Copyright © 2011, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

[1] We refer to buyers and sellers as female and male, respectively.

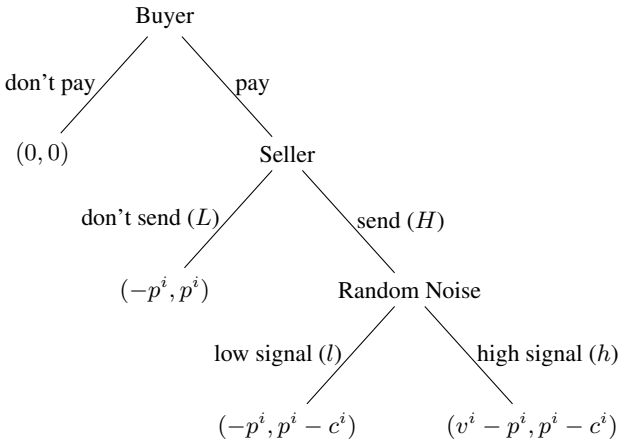

Figure 1: Game tree of a single trade with imperfect monitoring and no escrow mechanism. The first and second number denote the buyer's and the seller's payoff, respectively.

in that it avoids the assumption of long-lived sellers and removes the whitewashing problem. While we illustrate the application of escrow mechanisms to online auction sites like eBay, the general technique applies to a wide range of online markets.

A related line of research has developed on truthful feedback for reputation mechanisms. The peer prediction method by Miller, Resnick and Zeckhauser (2005) and its extensions due to Jurca and Faltings (2006; 2009) elicit truthful feedback for opinion forums. In earlier work, we study its application to an online auction setting with a strategic seller (Witkowski 2010). Jurca and Faltings (2007) study a similar problem, but their solution relies on repeated interactions between the buyer and the seller, which makes it inapplicable to e-commerce platforms like eBay or Amazon marketplace where most buyers interact with a particular seller only once. Furthermore, all these mechanisms still rely on publishing reported feedback. Our mechanism is most closely related to that of Dellarocas (2003), who studies online markets and incentivizes truthful seller behavior via listing fees. However, in contrast to our solution, Dellarocas' approach relies on knowing the seller's cost function and the distribution of buyer valuations. Moreover, our mechanism does not distort the sellers' surplus and—in the presence of fraudulent sellers—it remains both efficient and budget-balanced.

The Setting

Consider for concreteness an online auction site like eBay where transactions proceed as follows: a seller in auction i posts a description of the good that he wants to sell and sets a reserve price $m^i \geq 0$. People interested in the good bid for it in a second-price auction, where the bidder with the highest bid wins and has to pay the amount corresponding to the second highest bid or the reserve price, whichever is higher.[2] If the seller's reserve price is higher than the highest bid, the good is not sold and the game ends. We refer to the price of the ith auction as p^i. We assume that bids are natural numbers such as cent values.

Consider Figure 1, which depicts the game tree for a single trade, assuming the good was sold and the winner of the auction has been determined. First, the buyer is asked to transfer the money to the seller and the game ends here with 0 payoff for both parties if the buyer does not pay. If the buyer pays, the seller decides whether to send the good or keep it. That is, a seller can play either high or low effort. The seller's effort with respect to the buyer of auction i is denoted by $e^i \in \{L, H\}$. Exerting low effort corresponds to doing nothing and is free. High effort, however, is costly to the seller. The associated cost is denoted by c^i and is private to the seller. We also overload the notation and use $c^i(L)$ and $c^i(H)$ where more convenient to denote the seller's cost when exerting low or high effort respectively.

Following seller effort, buyer i observes a signal s^i that takes one of two possible values, "low" or "high": $s^i \in S = \{l, h\}$. In the context of online auctions, for example, a high signal refers to "buyer received item in described quality;" a low signal refers to "item significantly not as described, or item not received." Signals depend stochastically on the seller's effort and the probability for signal instantiation $s_m \in \{l, h\}$ given that the seller played $e_k \in \{L, H\}$ is denoted by: $f(s_m|e_k) = Pr(s^i = s_m|e^i = e_k)$. Note that $f(s_m|e_k)$ is assumed to be the same for all of the seller's auctions and is private information to the seller. Low effort leads to a low signal with certainty, i.e. $f(l|L) = 1$, while the probability of a high signal following high effort is positive but the outcome may also be a low signal, that is $1 > f(h|H) > 0$. For example, the seller might send the good as described but it gets lost in the mail.

Buyer i's valuation for signal s_m is denoted by $v^i(s_m)$, where we simplify notation, such that $v^i = v^i(h)$ is her valuation for the good as described. A buyer's valuation for a low signal is 0, i.e. $v^i(l) = 0$, but the value for v^i is private knowledge of the buyer. We assume that both buyers and sellers are risk-neutral. Moreover, we assume they have quasi-linear utility functions, i.e. $u^i(v^i, p^i, s^i = s_m) = v^i(s_m) - p^i$ for the buyers and $u^s(c^i, p^i, e^i = e_k) = p^i - c^i(e_k)$ for the sellers.

To clarify the trust problem, let us apply backward induction to this game. The last action is the effort decision by the seller. If he exerts high effort, i.e. if he sends the good in the described quality, he receives a payoff that is c^i less than if he keeps it which means that he should choose L. Anticipating this, however, the buyer should not pay since receiving a payoff of 0 is better than paying p^i and getting nothing. This is the problem of *moral hazard* (e. g., Dellarocas, 2006).

The Basic Mechanism

The escrow procedure only comes into play if a good is sold and a buyer has been determined. The core concept of an escrow mechanism is that a buyer does not pay the seller directly but via a trusted third party (henceforth the *center*). First, the buyer sends her payment to the center which holds it in escrow. The center then acknowledges receipt of payment to the seller, who then decides whether to exert effort or

[2] Please note that it is straightforward to generalize escrow mechanisms to other auction designs.

not. Once the seller has exerted his effort, the buyer is asked by the center what signal she received. Only if the buyer reports a high signal does the center forward the buyer's payment to the seller.

In contrast to presently-used escrow procedures, we do not simply reimburse every buyer who reports a low signal, because that creates a strong incentive for buyers to always report low. Thus, the key question is how to proceed with the withheld payment following a low report. While the escrow procedure only comes into play if an auction clears, the way the withheld payments are handled affects the bidding behavior in the auction preceding the escrow procedure. Thereby, the design of the escrow procedure also affects the efficiency of the overall mechanism.

Consider a straw-man solution to this problem: we could use the reports solely to determine the seller's payment and leave the surplus that it generates with the center. However, this mechanism is not efficient. The intuition is that buyers will take into account $f(l|H)$, i.e. the chance of receiving a low signal. Consequently, they lower their bids which can cause efficient trade to fail. If the center knew $f(l|H)$, then one way to achieve efficiency would be to pay a fixed rebate to every buyer, namely $f(l|H) \cdot p^i$, to offset the buyer's expected losses per transaction. The escrow mechanism that we present in this section is always efficient without relying on the center knowing $f(l|H)$.

The main idea of our mechanism is that whether or not a buyer receives a rebate (equal to her escrow payment) depends on the report of *another* buyer. Essentially, by the buyer reporting a signal, she enters into a lottery for receiving a rebate back from the escrow mechanism.[3] Nothing changes for the seller. It remains the case that the center forwards the payment to the seller if and only if the respective buyer reports a high signal.

Here, we present the mechanism with one seller and two buyers. We use indices 1 and 2 to refer to the corresponding auctions, their buyers, and the reports of those buyers. More generally, we use indices i and $3-i$ (observe that if $i=1$, it holds that $3-i=2$ and vice versa). We need at least two buyers to be able to determine each buyer's rebate payment. Every buyer needs to transfer the escrow payment p^i, i.e. the price of the good, to the center. If she reports a signal then she also has a chance to receive her payment back. We denote buyer i's signal report in auction i by $r^i \in S = \{l, h\}$. Figure 2 depicts how the mechanism groups a seller with two randomly chosen buyers. The critical parts are Steps 8 and 9: if a buyer reports h, the center pays the seller and the other buyer receives nothing back. If a buyer reports l, her payment is not forwarded to the seller, and the second buyer receives back her own original payment. We overload the notation for the price, such that $p^i(r^i)$ denotes the net payment to the seller following report r^i with $p^i(h) = p^i$ and $p^i(l) = 0$. Note that the seller learns about both reports only in Step 7. Similarly, each buyer only knows her own signal observation at the time of reporting.

[3]It is straightforward to use the reports of more than one other buyer to determine a buyer's rebate, so that the expected rebate stays the same but its variance is lower.

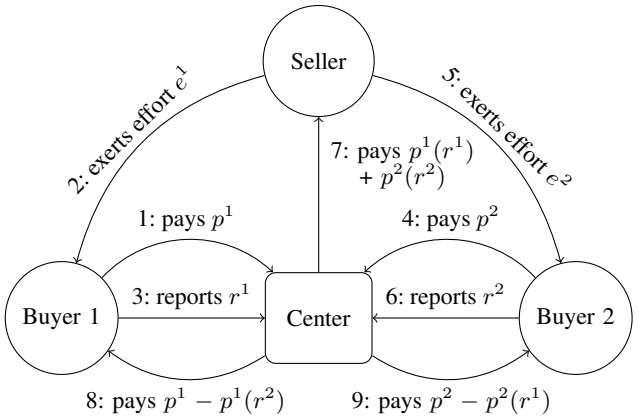

Figure 2: Procedure of the escrow procedure of the basic mechanism once the prices have been determined.

The seller's utility for participating in the mechanism consists of the negative costs for exerting effort plus the report-dependent payments he receives:

$$u^s(e^i, e^{3-i}, r^i, r^{3-i}) = -c^i(e^i) - c^{3-i}(e^{3-i}) + p^i(r^i) + p^{3-i}(r^{3-i}).$$

In practice, buyers may have a small cost for reporting feedback but, for reasons of clarity, we assume that their reporting costs are zero. Note, however, that a buyer only has the chance to receive her good for free when reporting a signal. A buyer's utility when participating both in the auction and the reporting phases depends on her valuation for the observed signal and on her net payment:

$$u^i(s^i, r^i, r^{3-i}) = v^i(s^i) - p^i(r^{3-i}).$$

Theoretical Analysis

In this section, we show that the basic escrow mechanism is incentive compatible, efficient, interim individually rational and ex ante budget-balanced. For the proofs we require the following lemma.

Lemma 1. *Given that the seller in auction i plays $e^i = H$, it is a weakly dominant strategy for bidders to bid their valuations.*

Proof. Let v denote a generic bidder's value, and p^i denote the price that the winning bidder has to pay in auction i. The bidder who wins auction i has an expected valuation of $f(h|H) \cdot v + f(l|H) \cdot p^i$, where the first term is the probability of receiving a high signal times the bidder's valuation for the item, and the second term corresponds to the expected rebate, i.e. the probability of receiving a low signal times the price she had to pay. Thus, her expected utility upon winning is $f(h|H) \cdot v + f(l|H) \cdot p^i - p^i = f(h|H) \cdot \left(v^i - p^i\right)$. So, whenever her valuation is at least as high as the price, she obtains positive utility and wants to win the auction. Since this is just the standard second-price auction case, it is a weakly dominant strategy for the bidder to bid her valuation. □

Incentive Compatibility

The information about the state of the game does not change between the seller's first and second effort decision, and so it is strategically equivalent to combine these two actions into a single one. We adopt this viewpoint in the sequel.

Definition 1 (Subgame-Perfect Incentive Compatibility)**.** *An escrow mechanism is* subgame-perfect incentive compatible *if the following strategy profile is a subgame perfect equilibrium*[4]*:*

- ***Seller****: Plays* $e^1 = e^2 = H$.
- ***Buyers****: Bid their valuations* v^1, v^2 *in the auction and report their true signals* $r^1 = s^1, r^2 = s^2$.

To prove that the basic escrow mechanism is incentive compatible, we first prove the following lemma.

Lemma 2. *Given an honest buyer report, setting the reserve price in auction* i *to* $m^i = \lfloor \frac{c^i}{f(h|H)} + 1 \rfloor$ *weakly dominates for the seller any strategy where* $m^i \leq \frac{c^i}{f(h|H)}$.

Proof. Let b_1 and b_2 be the highest and the second highest bid in the auction phase. We simplify notation by setting $C^i = \frac{c^i}{f(h|H)}$ and we distinguish two cases:
1. $b_1 < \lfloor C^i + 1 \rfloor$. For $m^i = \lfloor C^i + 1 \rfloor$, the seller obtains utility 0 as the good is not sold. For any lower value of m^i, he also receives 0 utility: the good is either not sold or it sold for $p^i \leq \lfloor C^i \rfloor$ in which case the seller is either indifferent between H and L or plays L.
2. $b_1 \geq \lfloor C^i + 1 \rfloor$. For $m^i = \lfloor C^i + 1 \rfloor$, the seller's payoff depends on b_2: if $b_2 \geq \lfloor C^i + 1 \rfloor$, the price equals b_2 and his expected utility is $b_2 - C^i$. If $b_2 \leq \lfloor C^i \rfloor$, the price is $p^i = \lfloor C^i + 1 \rfloor$ which results in a seller utility of $p^i - C^i > 0$ after playing H. For $m_i \leq \lfloor C^i \rfloor$, the seller's payoff also depends on b_2: if $b_2 \geq \lfloor C^i \rfloor$, his expected utility is $b_2 - C^i$. If $b_2 \leq \lfloor C^i - 1 \rfloor$, the price is $p^i = \lfloor C^i \rfloor$ which results in seller utility of 0 independent of what he is playing. □

Next, we show that our mechanism is *weakly* incentive compatible. The *weak* notion of incentive compatibility results from the buyers being only indifferent between reporting honestly and lying. Despite this indifference, the fact that we do not use a history of published feedback ensures robustness to certain manipulations that are common in reputation systems (e.g., badmouthing competitors). Later in the paper we present a mechanism that provides strict incentives to report truthful signals.

Theorem 3. *The basic escrow mechanism is (weakly) subgame-perfect incentive compatible.*

Proof. By backward induction. Since buyer i's report r^i does not influence her rebate, she is indifferent between reporting honestly and lying. Thus, in the reporting phase, neither buyer has an incentive to deviate from honest reporting. As the seller's decision does not depend on any other player (given the honest buyers), he can look at his effort decision for each buyer separately. His expected utility for exerting high effort for buyer i is $-c^i + f(h|H) \cdot p^i$. His expected utility for exerting low effort is 0, so that his single best response is to play H if $p^i > \frac{c^i}{f(h|H)}$ for $i = 1, 2$. This condition holds as it follows from Lemma 2 that the seller sets a reserve price $m^i > \frac{c^i}{f(h|H)}$. Thus, the seller has an incentive to play $e^1 = e^2 = H$ and the resulting situation is the standard one-shot second-price auction with reserve price. Incentive compatibility follows from Lemma 1. □

[4]For a definition, see p. 74 in Fudenberg and Tirole (1991)

Efficiency

Let V^i denote the highest valuation of all bidders involved in auction i. Because the winning bidder only receives her value V^i with probability $f(h|H)$, her expected valuation for winning is $f(h|H) \cdot V^i$.

Definition 2 (Efficiency)**.** *A mechanism is efficient if, for every auction* i*, the good is sold to the bidder with the highest value* V^i *whenever*

$$f(h|H) \cdot V^i \geq c^i \tag{1}$$

and remains with the seller otherwise.

We study efficiency for a reserve price $m^i = \frac{c^i}{f(h|H)} + \epsilon$ with $\epsilon \to 0$.

Proposition 4. *The basic escrow mechanism is efficient.*

Proof. "⇒" If the good is sold, we know that the price is at least as high as the reserve price, i.e. $p^i \geq m^i = \frac{c^i}{f(h|H)}$. According to Lemma 1, potential buyers bid their valuations, so that $v^i \geq p^i$ and, taken together, we obtain $v^i \geq \frac{c^i}{f(h|H)}$ which satisfies the property in Definition 2.
"⇐" Whenever $f(h|H) \cdot V^i \geq c^i$, we know that $V^i = v^i$ and thus $v^i \geq \frac{c^i}{f(h|H)}$. Thus, the highest bidder's bid is larger or equal to the seller's reserve price $m^i = \frac{c^i}{f(h|H)}$, and the good is sold. □

Individual Rationality

A common problem of online reputation mechanisms, including the one employed by eBay, is that a buyer's *realized* utility can still be negative, for example when the good never arrives and there is no way to get a full refund. While this also applies to our mechanism, we nevertheless establish a formal result on participation.

Definition 3 (Interim IR)**.** *A mechanism is* interim individually rational *if the following holds true in equilibrium:*

1. *Knowing his costs, every seller has a non-negative expected utility when participating in the market:* $U^s(e^i, e^{3-i}, r^i, r^{3-i}|c^i) \geq 0$.
2. *Knowing her valuation for the good, each buyer has a non-negative expected utility when participating in the market:* $U^i(s^i, r^i, r^{3-i}|v^i) \geq 0$.

Proposition 5. *The basic escrow mechanism is interim IR.*

Proof. According to Lemma 2, the seller sets his reserve prices m^i strictly larger than $\frac{c^i}{f(h|H)}$ and thus he is guaranteed non-negative utility. Next, without loss of generality, consider the first buyer's expected utility after the price has

been set but before she learns her signal: $U^1(r^1, r^2|v^1) = E[v^1] - E[p^1(r^2)] = f(h|H)(v^1 - p^1)$. This is non-negative. □

Budget Balance

As can be seen by inspecting Figure 2, the mechanism's budget for auctions 1 and 2 is $B = p^1(r^2) - p^1(r^1) + p^2(r^1) - p^2(r^2)$.

Definition 4 (Budget Balance)**.** *Assuming the seller and the buyers play the equilibrium strategies defined in Definition 1, an escrow mechanism is* ex ante budget-balanced *if the expected budget equals zero, i.e.* $E[B] = 0$*. An escrow mechanism is* ex post budget-balanced *if for every equilibrium outcome the realized budget equals zero, i.e.* $B = 0$*.*

Proposition 6. *The basic escrow mechanism is ex ante budget-balanced.*

Proof. In equilibrium, the expected budget is:

$$\begin{aligned} E[B] &= E\left[p^1(r^2) - p^1(r^1) + p^2(r^1) - p^2(r^2)\right] \\ &= E\left[p^1(r^2)\right] - E\left[p^1(r^1)\right] + E\left[p^2(r^1)\right] - E\left[p^2(r^2)\right] \\ &= f(h|H)\cdot p^1 - f(h|H)\cdot p^1 + f(h|H)\cdot p^2 - f(h|H)\cdot p^2 \\ &= f(h|H)\left(p^1 - p^1 + p^2 - p^2\right) = 0 \end{aligned}$$

The third line follows from the second since the conditional signal probability $f(h|H)$ is the same for both buyers and in equilibrium, the buyers report their true signals. □

Observe that if $p^1 = p^2$, the realized budget is always zero. However, if $p^1 \neq p^2$ and it happens that the buyers receive different signals due to random noise, then the *realized* budget is not balanced. Thus, the basic escrow mechanism is not ex post budget-balanced. Note that in large markets like eBay, ex ante budget balance is already a satisfying concept since there, the center's long-term risk is minimal. Furthermore, by matching auctions with similar prices, the center can reduce the variance of its budget.

Strict Buyer Incentives

A drawback of the mechanism discussed so far is that a buyer is only indifferent between honest reporting and lying. In this section we show that if seller abilities vary and if the center has distributional information about these variations, we can design an incentive compatible escrow mechanism with strict buyer incentives by adapting the *peer prediction* method (Miller, Resnick, and Zeckhauser 2005). The general idea of the peer prediction method is to compare the reported signals of the two buyers and construct a payment rule that depends on the *comparison* of these two reports. Based on prior probability knowledge about the seller's abilities, the center can design payments such that buyers maximize their expected utility when reporting honestly.

Extended Setting with Type Uncertainty

For the peer prediction method to be applicable to settings with moral hazard, it is crucial that there is seller type uncertainty in equilibrium, such that a buyer observing a signal learns something about another buyer's signal from the same seller. We study a model where type uncertainty stems from one seller being more accurate in describing the good than another seller. Imagine, for example, a used laptop being sold on eBay. The same laptop would be described differently by two different sellers. While one seller would state that it is "running perfectly", another would describe the laptop's condition more accurately. This difference in behavior influences the signal distribution of these two sellers, in that the latter has a higher probability to satisfy his customers and thus a higher probability of inducing high signals. It is important to note that neither seller intends to cheat, but that they are of different nature.

Formally, each seller is of one of two types, namely of the "good" or the "bad" type: $\theta \in \Theta = \{\theta_G, \theta_B\}$. All agents share a common prior belief $Pr(\theta_G)$ that the seller is of the good type θ_G with $1 > Pr(\theta_G) > 0$ and $Pr(\theta_G) + Pr(\theta_B) = 1$. The signal of a buyer depends on the seller's type. The probability for signal s_m given that the seller played e_k and is of type θ_t is assumed to be common knowledge and denoted by: $f_t(s_m|e_k) = Pr(s^i = s_m|\theta = \theta_t, e^i = e_k)$. The probability of a high signal following high effort is larger for a seller of the good type than it is for a seller of the bad type: $f_G(h|H) > f_B(h|H)$. Low effort leads, independent of the seller's type, to a low signal: $f_t(l|L) = 1$ for all $\theta_t \in \Theta$.

Because of the type uncertainty and because buyers now have beliefs over sellers' types, we adopt perfect Bayesian equilibrium for the analysis of the peer prediction escrow mechanism.

Definition 5 (Perfect-Bayesian Incentive Compatibility)**.** *An escrow mechanism is* perfect-Bayesian incentive compatible *if the following strategy profile is a perfect Bayesian equilibrium*[5]*:*

- ***Seller****: Plays* $e^1 = e^2 = H$*.*
- ***Buyers****: Bid their valuations* v^1, v^2 *in the auction and report their true signals* $r^1 = s^1, r^2 = s^2$*.*

The Peer Prediction Escrow Mechanism

As before, we overload the notation of p^i and let $p^i(r^i, r^{3-i})$ denote buyer i's net payment to the seller which now depends on both r^i and r^{3-i}. The seller's utility is defined as before, but the utility function of a buyer changes slightly:

$$u^i(s^i, r^i, r^{3-i}) = v^i(s^i) - p^i(r^i, r^{3-i}).$$

The escrow procedure still requires the buyer to transfer p^i to the center right after the auction clears. The probability of buyer $3-i$'s signal given buyer i's signal is denoted by $g^i(s_j|s_m) = Pr(s^{3-i} = s_j|s^i = s_m)$. For peer prediction to work, one buyer's signal must tell you something about the probability of the other buyer's signal. Formally, this concept is called *stochastic relevance* (Miller, Resnick, and Zeckhauser 2005). With only two signals h and l, this corresponds to the requirement that $g^i(h|h) \neq g^i(h|l)$ which follows from our setting if the seller plays $e^1 = e^2 = H$ because $f_G(h|H) > f_B(h|H)$. We refer to the paper by Miller,

[5] For a definition, see p. 325 in Fudenberg and Tirole (1991)

Resnick and Zeckhauser (2005) for the Bayesian transformations required to compute $g(s_j|s_m)$. Note that the center needs to perform additional Bayesian updates in case a buyer buys multiple items from the same seller over time.

Our goal is to design a payment rule such that in equilibrium, buyer i's expected utility when reporting honestly is strictly higher than when lying. Thus, we analyze the situation where the seller and buyer $3-i$ play the equilibrium strategies defined in Definition 1 and construct net payments $p^i(r^i, r^{3-i})$ such that reporting honestly is buyer i's unique best response in the reporting phase. In equilibrium the seller plays $e^1 = e^2 = H$ and thus the $g^i(s_j|s_m)$ are symmetric with regard to the buyers and we can drop its superscript. The expected utility of buyer i following signal s_m is then:

$$U^i(r^i|s^i = s_m) = v^i(s_m) - \sum_{s_j \in S} g(s_j|s_m) \cdot p^i(r^i, s_j).$$

We now describe the construction of a linear program whose solution is a payment rule $p^i(r^i, r^{3-i})$ with the desired properties, using the formulation due to Jurca and Faltings (2006). The incentive compatibility constraints require that the expected payment following an honest report is lower than that of a dishonest report. That is, for all $s_m,\ s_d \in S,\ s_m \neq s_d$:

$$\sum_{s_j \in S} g(s_j|\, s_m) \cdot p^i(s_m, s_j) < \sum_{s_j \in S} g(s_j|\, s_m) \cdot p^i(s_d, s_j)$$

To preserve budget balance and interim IR, a buyer must pay the same in expectation as in the basic setting. Note that the prior probability for a high signal report by the other buyer was $f(h|H)$ and is now $Pr(s^{3-i} = h)$:

$$\sum_{s_m \in S} Pr(s_m)\Big(\sum_{s_j \in S} g(s_j|\, s_m) \cdot p^i(s_m, s_j)\Big) = Pr(s^{3-i} = h) \cdot p^i$$

Finally, we demand that the best situation for a buyer is that she has to pay nothing while the worst is that she receives nothing back. That is, for every $r^i, r^{3-i} \in S$, it holds that:

$$0 \leq p^i(r^i, r^{3-i}) \leq p^i$$

Given these constraints, we maximize the rebate given to buyers who had bad experiences, i.e. our objective function maximizes

$$p^i - \sum_{s_j \in S} g(s_j|\, l)\ \cdot\ p^i(l, s_j).$$

The payments $p^i(r^i, r^{3-i})$ for buyer i can then be found as the solution to the following linear program:

$$\begin{aligned}
\text{max.}\ \ & p^i - \sum_{s_j \in S} g(s_j|\, l)\ \cdot\ p^i(l, s_j) \\
\text{s.t.}\ \ & \sum_{s_j \in S} g(s_j|s_m)\Big(p^i(s_d, s_j) - p^i(s_m, s_j)\Big) \geq \epsilon \\
& \qquad \text{for all } s_m,\ s_d \in S,\ s_m \neq s_d \\
& \sum_{s_m \in S} Pr(s_m)\Big(\sum_{s_j \in S} g(s_j|s_m) \cdot p^i(s_m, s_j)\Big) = Pr(h) \cdot p^i \\
& 0 \leq p^i(s_m, s_j) \leq p^i \ \ \text{for all } s_m, s_j \in S
\end{aligned}$$

Theorem 7. *The peer prediction escrow mechanism is (strictly) perfect-Bayesian incentive compatible, efficient, interim individually rational and ex ante budget-balanced.*

Proof Sketch. When buyer i receives signal s^i, the expected net payment $E\big[p^i(r^i, r^{3-i})|s^i\big]$ is strictly smaller when she honestly reports $r^i = s^i$ than when she lies. After making an honest report, a buyer's expected rebate is equal to her price times the probability of not receiving the good. Thus, she will bid her valuation in the auction, and together with reporting honestly in the reporting phase, this constitutes incentive compatibility of the mechanism. Efficiency, interim individual rationality and ex ante budget balance follows directly from incentive compatibility and the proofs for these properties proceed analogously to the proofs for Propositions 4 to 6.

Collusion and Cross-Seller Matching

One drawback of the escrow mechanisms discussed so far is their susceptibility to collusion. There are two possible types of collusion: first, two buyers colluding with each other and, second, the seller colluding with one of the buyers. Throughout, it is helpful to split the net payment $p^i(r^i, r^{3-1})$ into two parts: the *escrow part* and the *peer prediction part*. The escrow part corresponds to the rebate paid out in the basic escrow mechanism while the peer prediction part is the payment needed to guarantee strict incentives in the reporting phase.

In the buyer-buyer collusion attack, two buyers agree to buy from the same seller and both report a low signal, such that both receive a rebate. This attack can involve multiple collusion costs C, which may include account creation costs, bidding costs, and the costs for paying for an item obtained in the auction. Note that we can set the peer prediction part of the net payment arbitrarily small. To make honest reporting as attractive as possible, we set the peer prediction part of the net payment such that the maximally attainable peer prediction payment equals $C - \varepsilon$ for some small $\varepsilon > 0$. Then, a collusion attack on the peer prediction payments is not beneficial and, consequently, we only need to worry about collusion attacks on the escrow part of the payment.

To obtain robustness to buyer-buyer collusion in regard to the escrow part, we introduce "cross-seller matching." Up to this point, every buyer was "matched" with another buyer from the same seller. With cross-seller matching, the escrow mechanism is unchanged, except that every buyer is matched with a buyer from a *different* seller, randomly chosen from all auctions clearing within a particular time period, such as a day. Thus, a buyer could now be matched with a buyer from a seller of a different type to determine the escrow part of the net payment. The peer prediction part of the net payment, however, is still determined by the report from another buyer from the *same* seller. Assuming that every buyer only participates in one auction per time period, it is easy to see that introducing cross-seller matching preserves the peer prediction escrow mechanism's properties. The intuition is that a priori, i.e. before receiving a signal from the seller, the buyer's belief regarding her own seller's type is equal to the prior belief of any other seller's type.

Using cross-seller matching, we can make the peer prediction escrow mechanism robust against buyer-buyer collusion. For each time period, we let N denote the total number of clearing auctions, and we assume that each buyer only participates once. We obtain the following theorem:

Theorem 8. *The peer prediction escrow mechanism with cross-seller matching is robust against buyer-buyer collusion attacks if* $C > \frac{1}{N-2} \cdot p^i$.

Proof Sketch. The intuition for collusion-robustness is that in large enough markets, i.e. where N is large, the probability of being matched with a colluder is very small, and thus, the collusion costs C will be larger than the expected gains from colluding. We leave the details of the proof to a longer version of the paper.

Next, we consider the second collusion attack, i.e. a seller colluding with one of his buyers. In this attack, the seller exerts low effort which results in a low signal with certainty, but the buyer reports a high signal. Thus, the payment from the buyer is forwarded to the seller, so that the colluders have neither lost nor gained money. Considering that there is also a strictly positive chance of receiving the escrow payback, this collusion attack is powerful: the attackers never lose money and sometimes win money. Even with cross-seller matching, the attackers gain $Pr(s^i = l) \cdot p^i$ in expectation. However, this attack can also involve many different collusion costs C. These can include the costs for creating a buyer and a seller account, the costs for listing the item as a seller, the costs for bidding on the item as a buyer, and finally, whatever sales fee the market platform charges. Taking these collusion costs into account, we obtain the following theorem:

Theorem 9. *The peer prediction escrow mechanism with cross-seller matching is robust against buyer-seller collusion attacks if* $C > Pr(s^i = l) \cdot p^i$.

Proof. In the buyer-seller collusion attack, buyer i's payment is forwarded to the seller with certainty, no value is gained or lost, and no money is gained or lost. Thus, we only need to consider the rebate part. Because of cross-seller matching, neither the buyer nor the seller can influence the rebate, and the probability of receiving the rebate is equal to $Pr(s^i = l)$, i.e. the prior probability of a low signal. Thus, the expected rebate is $Pr(s^i = l) \cdot p^i$. Consequently, if $C > Pr(s^i = l) \cdot p^i$, the expected utility from the buyer-seller collusion attacks is negative. □

In real-world marketplaces, the biggest part of the collusion cost for the buyer-seller collusion attack would most likely be some kind of sales fee. Note that eBay, for example, charges a *final value fee* of 9% (capped at $100) for all items sold via regular auctions. Thus, in this case, the collusion costs are proportional to the selling price, namely $p^i \cdot$ fee. For the collusion attack to be unattractive, we thus need that $p^i \cdot \text{fee} > Pr(s^i = l) \cdot p^i$ which is true whenever $\text{fee} > Pr(s^i = l)$. In practice, we would expect that the probability of not receiving the good in the promised condition is relatively low, most certainly smaller than 9%, i.e., $Pr(s^i = l) < 0.09$. Thus, for markets with large sales fees, cross-seller matching is effective to provide robustness against buyer-seller collusion attacks.

Conclusion

In this paper, we have introduced *escrow mechanisms* to address moral hazard in online markets. The main idea is to install a trusted intermediary that forwards the payment from the buyer to the seller only if the buyer reports that she has received the good in the promised condition. A distinct property of escrow mechanisms is that they do not publish feedback reports and are thus history-free. In contrast to reputation mechanisms, escrow mechanisms do not rely on long-lived sellers and do not have a whitewashing problem. Furthermore, we have shown how they can be used to properly incentivize sellers and buyers without any knowledge of the buyers' valuations or the sellers' cost functions. It was interesting to see that the rebate we pay back to the buyers is necessary for the mechanisms to be fully efficient. Because we do not publish any feedback, there is no reputation profile that buyers can manipulate. This enabled us to achieve strict incentives in the reporting phase with any positive budget for the peer prediction part of the payment, while maintaining ex ante budget balance. Combining this with cross-seller matching, we were able to make the mechanism robust against collusion attacks.

Acknowledgements

We thank Yiling Chen, Michael Manapat and Haoqi Zhang for very helpful discussion on this work.

References

Bolton, G.; Greiner, B.; and Ockenfels, A. 2011. Engineering Trust – Reciprocity in the Production of Reputation Information. Working Paper 42, University of Cologne.

Dellarocas, C. 2003. Efficiency through Feedback-contingent Fees and Rewards in Auction Marketplaces with Adverse Selection and Moral Hazard. In *Proceedings of the 4th ACM Conference on Electronic Commerce (EC'03)*.

Dellarocas, C. 2006. Reputation Mechanisms. In Hendershott, T., ed., *Handbook on Information Systems and Economics*. Elsevier Publishing.

Fudenberg, D., and Levine, D. K. 1989. Reputation and equilibrium selection in games with a patient player. *Econometrica* 57:759–778.

Fudenberg, D., and Tirole, J. 1991. *Game Theory*. MIT Press.

Jurca, R., and Faltings, B. 2006. Minimum Payments that Reward Honest Reputation Feedback. In *Proceedings of the 7th ACM Conference on Electronic Commerce (EC'06)*.

Jurca, R., and Faltings, B. 2007. Obtaining Reliable Feedback for Sanctioning Reputation Mechanisms. *Journal of Artificial Intelligence Research (JAIR)* 29:391–419.

Jurca, R., and Faltings, B. 2009. Mechanisms for Making Crowds Truthful. *Journal of Artificial Intelligence Research (JAIR)* 34:209–253.

Miller, N.; Resnick, P.; and Zeckhauser, R. 2005. Eliciting Informative Feedback: The Peer-Prediction Method. *Management Science* 51(9):1359–1373.

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