*Computational Intelligence,* Volume 18, Number 4, 2002

# CONTRACTING WITH UNCERTAIN LEVEL OF TRUST

Sviatoslav Braynov

*Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, New York*

Tuomas Sandholm

*Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania*

The paper investigates the impact of trust on market efficiency and bilateral contracts. We prove that a market in which agents are trusted to the degree they deserve to be trusted is as efficient as a market with complete trustworthiness. In other words, complete trustworthiness is not a necessary condition for market efficiency. We prove that distrust could significantly reduce market efficiency, and we show how to solve the problem by using appropriately designed multiagent contracts. The problem of trust is studied in the context of a bilateral negotiation game between a buyer and a seller. It is shown that if the seller's trust equals the buyer's trustworthiness, then the social welfare, the amount of trade, and the agents' utility functions are maximized. The paper also studies the efficiency of advance payment contracts as a tool for improving trustworthiness. It is proved that advance payment contracts maximize the social welfare and the amount of trade. Finally, the paper studies the problem of how to make agents truthfully reveal their level of trustworthiness. An incentive-compatible contract is defined, in which agents do not benefit from lying about their trustworthiness. The analysis and the solutions proposed in this paper could help agent designers avoid many market failures and produce efficient interaction mechanisms.

*Key words*: e-commerce, decision theory, game theory, negotiation, multi-agent systems, distributed AI.

## 1. INTRODUCTION

As electronic commerce develops, its success crucially depends on establishing, maintaining, and managing trust in on-line transactions. Economic agents can fail to perform their tasks or to meet their commitments due to lack of incentives, lack of ability, or due to circumstances beyond their control. In risky environments trust enables cooperation and permits voluntary participation in mutually beneficial transactions which are otherwise costly to enforce or cannot be enforced by third parties.

The concept of trust has been a subject of continuous interest in different research areas, including multi-agent systems (Marsh 1994; Braynov and Sandholm 2002; Castelfranchi and Falcone 1999), sociology (Coleman 1990), risk management (Slovic 1997), economics and game-theory (Dasgupta 1990; Snijders 1996; Dellarocas 2001). The notion of trust is also closely related to the design and implementation of multistage safe exchanges (Sandholm and Ferrandon 2000; Sandholm and Lesser 1995; Matsubara and Yokoo 2000) and deviation-proof plans (Braynov and Sandholm 1999; Braynov 1994).

Trust is usually considered a belief or cognitive stance (Castelfranchi and Falcone 1999; Dimitrakos 2001b) that could eventually be quantified by a subjective probability (Kaplan and Garrik 1981; Jøsang 1999, 2001; Daskalopulu et al. 2002). It is usually assumed that an agent will only engage in a transaction if the level of trust exceeds some personal threshold (the level of acceptable trustworthiness), which depends on the transaction context (Coleman 1990; Kaplan and Garrik 1981).

This paper analyzes the role of trust in contracts which are not completely enforceable. In such contracts agents cannot rely on third parties or control mechanisms for backing up a transaction. In risky contracts each party usually has its own estimate of the trustworthiness of the other party. Such estimates help agents make decisions and reason about other agents'

Address correspondence to Sviatoslav Braynov, Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260; e-mail: sbraynov@cse.buffalo.edu.

© 2002 Blackwell Publishing, 350 Main Street, Malden, MA 02148, USA, and 108 Cowley Road, Oxford, OX4 IJF, UK.

trustworthiness. Obtaining and maintaining such estimates, however, is a matter of strategic planning. Each party has to know whether it is beneficial to invest some resources in improving its current estimate and how accurate the estimate should be in order to guarantee a profitable outcome. It is common practice that in order to save some resources, agents insure themselves by undertrusting, i.e., by placing less trust on the partner. For example, many risk-assessment firms treat the lack of credit history as a lack of trust. This is usually motivated by the fact that the marginal cost of obtaining additional evidence of trust exceeds the marginal benefit of the evidence. Undertrusting is a typical example of a market inefficiency produced by inaccurate trust estimates. In the extreme case, undertrusting could produce such a low estimate of the partner's trustworthiness that an agent might decide not to participate in a transaction, even when the partner is completely trustworthy.

The problem of how trust estimates are actually learned is beyond the scope of this paper. Daskalopulu et al. (2002) use evidence-based subjective reasoning for assessing the performance of agents during the execution of a contract, based on potentially observable evidence. Propagation of trust has been studied by Dimitrakos (2001a). His analysis includes a number of roles and factors that support trust establishment in business transactions. Subjective reasoning in trust-based decision making has been used by Jøsang (1999, 2001). He proposed a formal model of subjective reasoning based on a theory of uncertain probabilities. The model includes discounting and consensus operators for incorporating second-hand evidence in trust-based decision making. In our previous research (Braynov and Sandholm 2002; Braynov 2001) we developed a mechanism for trust learning which does not require trust assessment, and does not rely on collecting and analyzing information about untrustworthy agents. In the mechanism agents truthfully reveal their trustworthiness at the beginning of every interaction.

It should be pointed out, however, that trust assessment remains a serious practical problem for which no satisfactory solution has been found. Trust assessment usually runs into the following problems.

First, trust assessment requires long-term interaction and is usually costly for the learning agent, who has to accept the risk of being abused for learning purposes. Learning costs may include information search costs, costs for obtaining additional guarantees from trusted third parties, etc. If the costs are prohibitively high, then an interaction may fail, regardless of the trustworthiness of the other party.

Second, trust typically is learned gradually, but can be destroyed in an instant by misfortune or a mistake. Once trust is lost, it may be costly or it may take a long time to rebuild it. This reflects certain fundamental mechanisms of human psychology known as the asymmetry principle (Slovic 1997). According to it, distrust is not merely the inverse of trust. Trust and distrust are learned in different ways and they function differently. This might be explained by the following psychological reasons (Slovic 1997):

- Humans perceive trust-destroying events as more noticeable than trust-building events.
- Trust-destroying events carry much greater weight than trust-building events in subjective reasoning.
- Sources of trust-destroying events tend to be accepted as more credible than sources of trust-building events.
- Once initiated, distrust tends to reinforce itself. In contrast, trust can be destroyed in an instant.

Third, every trust-learning mechanism can fail with positive probability (Dasgupta 1990). The problem is that any learning agent will have, with positive probability, a run of bad encounters leading to such a low estimate of the partner's trustworthiness that the learning agent no longer wishes to learn.

Fourth, in many cases trust learning may be impossible. The Internet provides vast opportunities to interact with total strangers for whom there is no history of previous transactions. Even if we assume that some information could be obtained from reputation databases or recommender systems, this information is often too general to be applied to the context of a particular transaction. The coverage of current reputation systems is also very limited. There are problems with the aggregation of correlated reputation, fake transactions (Zacharia and Maes 2000), false identity and pseudonyms (Friedman and Resnick 1998), the obtaining of negative feedback (Resnick et al. 2000), interpretability and portability of reputation databases, etc.

Fifth, the process of trust learning seldom produces complete and accurate estimates. Inaccurate beliefs could lead to interaction failures and inefficiencies. For example, an agent with an inaccurate estimate of his partner's trustworthiness might decide not to participate in an interaction, even when the other party is completely trustworthy.

This paper studies the impact of trust estimates and trust beliefs on market efficiency, and especially on multiagent negotiation. We prove that inaccurate trust estimates could have detrimental effects on market efficiency, and show how to solve the problem by using appropriately designed multiagent contracts. We prove that if agents hold accurate trust estimates about one another, then the social welfare, the amount of trade, and the agents' utility levels are maximized. We also show that market efficiency does not require complete trustworthiness. Untrustworthy agents could transact as efficiently as trustworthy agents, provided that they hold accurate estimates of one another. Therefore, what really matters is not the actual level of trustworthiness, but the accuracy of individual estimates. A market in which agents are trusted to the degree they deserve to be trusted is as efficient as a market with complete trustworthiness. We show that inaccurate trust estimates produce socially suboptimal results or market failures.

The paper also discusses advance payment contracts as a solution to the problem of inaccurate trust estimates. We show that advance payment contracts maximize individual utilities, the social welfare, and the amount of trade. In addition, advance payment contracts serve as a screening device. That is, they could be used to separate trustworthy agents from untrustworthy ones.

The paper also analyzes the problem of how to make agents truthfully reveal their level of trustworthiness. It is usually the case that agents have a strong incentive to misrepresent their level of trustworthiness. By declaring more trustworthiness, agents usually enjoy more benefits of cooperation and more opportunities to abuse trust. In the paper we propose an incentive-compatible contract in which agents do not benefit from lying about their trustworthiness.

The paper is organized as follows. Section 2 defines a formal model of a bilateral negotiation. In Section 3 we analyze how trust estimates affect negotiation and the terms of contracts. In Section 4 we show how advance payment contracts could solve the problem of inaccurate trust estimates. Section 5 is devoted to incentive-compatible contracts, i.e., contracts in which agents do not lie about their trustworthiness. Finally, Section 6 concludes by summarizing the results and provides directions for future research.

## 2. THE CONTRACTING PROBLEM

We assume that trust is a bilateral relation that involves an entity manifesting trust called the *trustor* and an entity being trusted called the *trustee*. Further, we assume that

- The trustee knows something about itself or about the world which the trustor does not know, and/or the trustee can perform an action that the trustor cannot monitor.

– The information that the trustee knows and/or the action which it performs have a bearing on the trustor's choice of action.
– The trustor has to choose an action (or make a decision) without knowing the trustee's private information or private action.
– The trustor's choice of action is based on the belief that a particular event $\mathcal{E}$ has already occurred, is occurring, or will occur. The event $\mathcal{E}$ depends on the trustee's private information or private action.
– The event $\mathcal{E}$ is favorable to the trustor. That is, the trustor prefers $\mathcal{E}$ occurring to $\mathcal{E}$ not occurring.

In other words, the trustor depends on the trustee for some favorable event $\mathcal{E}$ which is controlled by the trustee. We assume that the degree of trustworthiness, $\gamma$, could be measured by the probability of $\mathcal{E}$. That is, the trustee behaves favorably with probability $\gamma$. In situations where the value of actual trustworthiness is not directly accessible, it is worthwhile differentiating among perceived trustworthiness, actual trustworthiness, and trust. Perceived trustworthiness is defined as one's estimate of $\gamma$ which is usually subjective and could be different from the actual trustworthiness, i.e., from the actual probability of $\mathcal{E}$.

Consider a bilateral negotiation involving a buyer and a seller. The seller produces some commodity and sells it to the buyer. We assume that the seller always delivers the commodity i.e., he is completely trustworthy. The buyer's trustworthiness, however, may vary. In other words, the buyer pays with some probability $\gamma$, $\gamma \in [0, 1]$. We assume that the seller delivers first and after that the buyer pays. In this case, the seller depends on the buyer for the event $\mathcal{E} = \{\text{the buyer pays}\}$, and the seller has to choose whether to enter into a transaction without being able to control $\mathcal{E}$.

We consider the general case where both the buyer and the seller are uncertain about the actual trustworthiness $\gamma$. Let $\alpha$ and $\beta$ denote, respectively, the seller's and the buyer's estimates of $\gamma$. Every agent calculates his estimate based on his frame of discernment which is often different from the frame of discernment of the other agent. As a result, agents could come up with different estimates of $\gamma$. It may seem unnatural at first that the buyer has to estimate his own actual trustworthiness $\gamma$. In many cases, however, the buyer might not have absolute control over his payment or might not have first-hand evidence of a money transfer that amounts to payment.

Depending on the context, the actual trustworthiness $\gamma$ may have an objective, subjective, or a mixed nature. On the other hand, the individual estimates $\alpha$ and $\beta$ are usually subjective. One way to interpret $\gamma$ is to consider trustworthiness as willingness to perform one's share in a transaction. In this case $\gamma$ is an indicator of the buyer's willingness to pay. It could be subjective if we look at trustworthiness as dependable intention (Dimitrakos 2001a) or objective if the buyer pays with some probability that maximizes his profits for a given resource endowment.

A second way to look at $\gamma$ is to see it as the buyer's ability or capacity to pay (the buyer, for example, may be willing to pay, but may not have available funds). In this case $\gamma$ could be objective if the buyer's payment depends on circumstances beyond his control; for example, the payment processing time. In another interpretation $\gamma$ is the probability that the contract between the buyer and the seller can be enforced by an enforcement agency. That is, $\gamma$ could be the probability of detecting particular violation, imposing particular type of sanctions, and using particular type of procedures for adjudicating disputes. It is usually the case that the optimal amount of law enforcement depends on the amount and type of resources and punishments used to enforce a piece of legislation. As a result many law enforcement policies are probabilistic (Becker and Landes 1974).

In this section we assume that the buyer truthfully declares his estimate $\beta$. Later on we will drop this assumption and will show that the buyer can manipulate the seller by declaring false levels of trustworthiness. We assume that $\alpha$ and $\beta$ are common knowledge. That is, everybody knows $\alpha$ and $\beta$, everybody knows that everybody knows $\alpha$ and $\beta$, etc.

The seller's estimate, $\alpha$, could differ from the buyer's estimate, $\beta$. Although the buyer declares the value of $\beta$, there is no reason for the seller to believe him. It is usually the case that the seller has a strong incentive to misrepresent its level of trustworthiness. At first sight it may seem unusual that $\alpha$ and $\beta$ differ even though they are common knowledge. It is a well-known fact, however, that disagreement on mutual beliefs can occur among rational agents, if the agents have different priors (Aumann 1976), such as is often the case in e-commerce settings.

Before we analyze the negotiation between the buyer and the seller, we introduce some standard notation. Let $q$ be the quantity of the commodity which is produced and sold. The seller's cost function is denoted by $C(q)$. Let $V(q)$ be the buyer's value function (how much value the buyer places on quantity $q$). The contract price is denoted by $P$. Making the usual assumption of quasi-linearity, the expected utility of the buyer is

$$U_B(q) = V(q) - \beta P. \tag{1}$$

The expected utility of the buyer indicates that with probability $1 - \beta$ the buyer can obtain the quantity $q$ without paying for it. The expected utility of the seller is

$$U_S(q) = \alpha P - C(q). \tag{2}$$

The expected utility of the seller reflects his subjective judgment that with probability $1 - \alpha$ he will not be paid. For notational simplicity we refer to expected utility as utility whenever this is not a source of confusion.

We assume that the value function, $V(q)$, and the cost function, $C(q)$, are not negative:

$$V(q) \geq 0, C(q) \geq 0.$$

We assume that agents negotiate the transaction terms using the most common bargaining solution: the *Nash bargaining solution* (Nash 1950). It was introduced by John Nash in 1950, and plays a central role in the theory of bargaining. The Nash bargaining solution is of fundamental importance because it is the only bargaining agreement that satisfies the following desirable properties (Nash 1950; Osborne and Rubinstein 1990). First, it always yields an efficient outcome, i.e., an outcome that exhausts all opportunities for making both agents better off. Second, the Nash solution is socially fair in the sense that neither agent is preferred in any way to the other. Third, the solution does not depend on irrelevant alternatives. That is, the outcome of a bargain cannot depend on alternative bargains that are not actually chosen by players. Fourth, the solution is independent of the ways agents measure their utilities. Applied to our setting, the Nash bargaining solution is

$$\arg \max_P (V(q) - \beta P)(\alpha P - C(q)). \tag{3}$$

Once the contract price has been decided, the quantity $q$ is chosen so as to maximize the utility functions of both agents. Later we will prove that for a given contract price there always exists a quantity $q$ that maximizes the agents' utility functions.

In this paper we confine our attention to the case when the buyer is distrusted, i.e., $\alpha < \beta$. According to the asymmetry principle (Slovic 1997) this is the more common situation.

## 3. UNDERTRUSTING

In this section we prove that undertrusting ($0 < \alpha < \beta$) leads to an inefficient allocation of resources. Specifically, the quantity exchanged and the utilities of both agents will be smaller than those obtained in the case when the seller places the correct amount of trust on the buyer ($\alpha = \beta$).

*Proposition 1.* The general solution of the Nash bargaining problem (3) in our setting is given by

$$P = \frac{\alpha V(q_1) + \beta C(q_1)}{2\alpha\beta}$$

$$U_S(q_1) = \frac{\alpha V(q_1) - \beta C(q_1)}{2\beta} \tag{4}$$

$$U_B(q_1) = \frac{\alpha V(q_1) - \beta C(q_1)}{2\alpha} \tag{5}$$

Proof. Consider the maximization problem (3). The first-order condition with respect to $P$ is

$$\alpha V(q) - 2\alpha\beta P + \beta C(q) = 0$$

Therefore,

$$P = \frac{\alpha V(q_1) + \beta C(q_1)}{2\alpha\beta}$$

Substituting $P$ in (1) and (2) we obtain (4) and (5). The agents will choose the quantity $q_1$ so as to maximize $\alpha V(q) - \beta C(q)$. □

By means of Proposition 1 we can prove that the bargaining problem (3) is well defined. That is, once the contract price (as a function of quantity, $q$) has been decided, the quantity $q$ can be determined unambiguously.

*Proposition 2.* For a given contract price $P$ the seller and the buyer prefer the same quantity.

Proof. From (4) and (5) it follows that both the seller and the buyer will choose the quantity that maximizes $\alpha V(q) - \beta C(q)$. □

The following proposition follows immediately from Proposition 1.

*Proposition 3.* If $\alpha = \beta$, then the quantity exchanged, $q_1$, maximizes $V(q) - C(q)$ and

$$U_B(q_1) = U_S(q_1) = \frac{V(q_1) - C(q_1)}{2}$$

$V(q) - C(q)$ in Proposition 3 is the social welfare, i.e., the net benefits to the agents. Therefore, when trust matches trustworthiness ($\alpha = \beta$), social welfare is maximized. Proposition 3 also implies that an underestimation of the buyer's trustworthiness produces socially suboptimal results. Maximizing social welfare, however, does not imply that the maximal output will be produced and exchanged. The following proposition provides the result we need.

*Proposition 4.* If the value function $V(q)$ satisfies the standard diminishing returns requirements:

$$\frac{\partial V(q)}{\partial q} \geq 0, \frac{\partial^2 V(q)}{\partial q^2} \leq 0,$$

and if the cost function is increasing and convex:

$$\frac{\partial C(q)}{\partial q} \geq 0, \frac{\partial^2 C(q)}{\partial q^2} \geq 0,$$

then the quantity $q_1$ that maximizes the social welfare $V(q) - C(q)$ is the maximal possible output.

Proof. Let $q_1$ maximize $V(q) - C(q)$. According to Proposition 2 the buyer and the seller choose quantity $q_2$ that maximizes $\alpha V(q) - \beta C(q)$. Therefore, we have to show that $q_2 \leq q_1$. Since $q_2$ is an extremum of the function $\alpha V(q) - \beta C(q)$, we have

$$\alpha V'(q_2) - \beta C'(q_2) = 0,$$

where $V'(q)$ and $C'(q)$ are the first derivatives of the functions $V(q)$ and $C(q)$. From $0 < \alpha < \beta$ it follows that

$$\alpha V'(q_2) - \beta C'(q_2) < \alpha V'(q_2) - \alpha C'(q_2).$$

Therefore,

$$0 < V'(q_2) - C'(q_2).$$

For $q_1$ we have:

$$V'(q_1) - C'(q_1) = 0.$$

That is,

$$V'(q_2) - C'(q_2) > V'(q_1) - C'(q_1).$$

The function $V'(q) - C'(q)$ is decreasing ($V''(q) \leq 0$ and $C''(q) \geq 0$). Therefore, $q_2 \leq q_1$. □

According to Propositions 3 and 4, when $\alpha = \beta$ the social outcome is efficient in terms of quantity and welfare. That is, when trust matches trustworthiness, the agents enjoy the maximal social benefit and the maximal amount of trade. Maximizing social welfare, however, does not imply maximizing individual utilities. It may be the case that there is a conflict between individual and social interest. The following proposition asserts that in the case where $\alpha = \beta$, the individual interests coincide with the social interest.

*Proposition 5.* When trust matches trustworthiness ($\alpha = \beta$), the seller and the buyer both maximize their individual utility functions.

Proof. Let $q_1$ maximize $V(q) - C(q)$ and let $q_2$ maximize $\alpha V(q) - \beta C(q)$. We have to show that $U_B(q_2) < U_B(q_1)$ and $U_S(q_2) < U_S(q_1)$.

$$\begin{aligned} U_B(q_2) &= \frac{\alpha V(q_2) - \beta C(q_2)}{2\alpha} < \frac{\alpha V(q_2) - \alpha C(q_2)}{2\alpha} \\ &= \frac{V(q_2) - C(q_2)}{2} \leq \frac{V(q_1) - C(q_1)}{2} = U_B(q_1) \end{aligned}$$

Analogously,

$$U_S(q_2) = \frac{\alpha V(q_2) - \beta C(q_2)}{2\beta} < \frac{\beta V(q_2) - \beta C(q_2)}{2\beta}$$
$$= \frac{V(q_2) - C(q_2)}{2} \leq \frac{V(q_1) - C(q_1)}{2} = U_S(q_1). \qquad \square$$

The proof of Proposition 5 shows that an underestimation of the buyer's trustworthiness reduces the utility function of each agent, making both agents worse off. According to Propositions 3–5, the case where the seller trusts the buyer to the extent that the buyer deserves to be trusted is optimal for society. By an optimal outcome we mean one that maximizes the social welfare, the quantity produced, and the agents' utility functions. To obtain efficiency it is not necessary that the buyer be trustworthy. The only relevant factor is the seller's accuracy in estimating the buyer's trustworthiness. Any underestimation of the buyer's trustworthiness tends to harm each agent and society as a whole.

The next example illustrates Propositions 3–5. Suppose that the buyer's value function is

$$V_B(q) = 16\ln(q+1) \tag{6}$$

and the seller's cost function is

$$C_S(q) = q^2 - 2q + 2. \tag{7}$$

If $\alpha = \beta = 1 = \gamma$, then $q = 3$, $P = 13.59$, and $U_S(3) = U_B(3) = 8.59$. That is, if the buyer is completely trustworthy and the seller believes him, each agent's utility will be 8.59 and the quantity exchanged will be 3. Suppose now that the buyer is still completely trustworthy, but the seller believes him with probability 0.375. In other words, $\alpha = 0.375$ and $\beta = \gamma = 1$. Then $q = 2$, $P = 11.46$, $U_S(2) = 2.30$, and $U_B(2) = 6.12$. Therefore, lack of trust reduces the quantity exchanged and the utility of each agent.

## 4. IMPROVING TRUSTWORTHINESS BY ADVANCE PAYMENTS

In this section we discuss the problem of how a distrusted buyer can convince the seller of his trustworthiness. One possible way for a distrusted buyer to signal his trustworthiness is to make an advance payment to the seller, i.e., to pay some amount $P_0$ before the seller delivers the commodity. We assume that in the absence of payment the seller is relieved from his obligations. That is, if the buyer does not make an advance payment after he has promised to do so, the seller is relieved from his obligation to deliver.

*Proposition 6.* If $0 < \alpha < \beta$ and the agents choose an advance payment contract, then according to the Nash bargaining solution, the amount of advance payment is

$$P_0 = \frac{\alpha V(q_1) + \beta C(q_1)}{\alpha + \beta}.$$

The quantity exchanged, $q_1$, maximizes the function $V(q) - C(q)$, the amount of contract payment is zero ($P = 0$), and

$$U_B(q_1) = \frac{\beta}{\alpha + \beta}(V(q_1) - C(q_1)), \tag{8}$$

$$U_S(q_1) = \frac{\alpha}{\alpha + \beta}(V(q_1) - C(q_1)). \tag{9}$$

The quantity exchanged with an advance payment contract equals the quantity exchanged with an uncertain payment contract.

Proof. The utility functions of the buyer and the seller are

$$U_B(q) = V(q) - P_0 - \beta P, \tag{10}$$

$$U_S(q) = \alpha P + P_0 - C(q). \tag{11}$$

Therefore, the Nash bargaining solution is

$$\arg\max_P (V(q) - P_0 - \beta P)(\alpha P + P_0 - C(q)).$$

This is maximized by

$$P = \frac{\alpha V(q) + \beta C(q) - (\alpha + \beta)P_0}{2\alpha\beta}.$$

Substituting $P$ in (10) and (11) we obtain

$$U_B(q) = \frac{\alpha V(q) - \beta C(q) + (\beta - \alpha)P_0}{2\alpha}, \tag{12}$$

$$U_S(q) = \frac{\alpha V(q) - \beta C(q) + (\beta - \alpha)P_0}{2\beta}. \tag{13}$$

The seller and the buyer will choose a quantity, $q_1$, that maximizes $\alpha V(q) - \beta C(q) + (\beta - \alpha)P_0$. Since $P_0$ is constant, $q_1$ also maximizes $\alpha V(q) - \beta C(q)$. With an uncertain payment contract, the seller produces a quantity that maximizes the same function. Therefore, advance payment and uncertain payment contracts lead to the same quantity and the same social welfare.

How do the agents choose the advance payment $P_0$? $P_0$ has to maximize $\alpha V(q_1) - \beta C(q_1) + (\beta - \alpha)P_0$ subject to the constraint that $P(q_1) \geq 0$. Thus, we obtain

$$P_0 = \frac{\alpha V(q_1) + \beta C(q_1)}{\alpha + \beta}.$$

Substituting $P_0$ in (12) and (13) we obtain (8) and (9). It is straightforward to check that for this value of $P_0$ we have $P = 0$. That is, the buyer pays the entire contract price in advance. □

According to Proposition 6 if the buyer is trustworthy, he should pay the entire price in advance. This corresponds to our intuition that paying before or after the delivery does not make a difference for a trustworthy buyer. Furthermore, Proposition 6 says that if a distrusted buyer makes an advance payment, then the social welfare, $V(q) - C(q)$, is maximized. It is not obvious, however, that the advance payment contract is preferred by both agents to the standard uncertain payment contract. According to the following proposition, if the buyer is distrusted, the advance payment contract gives each agent higher utility than the uncertain payment contract.

*Proposition 7.* If $q_1$ and $q_2$ are quantities exchanged in advance and uncertain payment contracts, respectively, and if $0 < \alpha < \beta$ then

$$U_B(q_1) = \frac{\beta}{\alpha + \beta}(V(q_1) - C(q_1)) \geq U_B(q_2) = \frac{\alpha V(q_2) - \beta C(q_2)}{2\alpha}$$

and

$$U_S(q_1) = \frac{\alpha}{\alpha+\beta}(V(q_1) - C(q_1)) \geq U_S(q_1) = \frac{\alpha V(q_2) - \beta C(q_2)}{2\beta}.$$

Proof. Since $0 < \alpha < \beta$, $V(q) \geq 0$, and $C(q) \geq 0$, for every $q$ we have

$$\alpha\beta V(q) - \alpha\beta C(q) \geq \alpha^2 V(q) - \beta^2 C(q).$$

This leads to

$$\frac{\beta V(q) - \beta C(q)}{\alpha+\beta} \geq \frac{\alpha V(q) - \beta C(q)}{2\alpha}. \tag{14}$$

Let $q_1$ maximize $\frac{\beta}{\alpha+\beta}(V(q_1) - C(q_1))$ and $q_2$ maximize $\frac{\alpha V(q) - \beta C(q)}{2\alpha}$. Then

$$U_B(q_1) = \frac{\beta}{\alpha+\beta}(V(q_1) - C(q_1)) \geq \frac{\beta}{\alpha+\beta}(V(q_2) - C(q_2)).$$

From (14) it follows that

$$\frac{\beta}{\alpha+\beta}(V(q_2) - C(q_2)) \geq \frac{\alpha V(q_2) - \beta C(q_2)}{2\alpha}.$$

That is,

$$U_S(q_1) \geq U_S(q_2).$$

Consider now $U_S(q_1)$ and $U_S(q_2)$.

$$\begin{aligned} U_S(q_1) &= \frac{\alpha}{\alpha+\beta}(V(q_1) - C(q_1)) \geq \frac{\alpha}{\alpha+\beta}(V(q_2) - C(q_2)) \\ &\geq \frac{\alpha V(q_2) - \beta C(q_2)}{2\beta} = U_S(q_2). \end{aligned}$$ □

From Propositions 3–5 it follows that when trust matches trustworthiness, the uncertain payment contract is optimal. Analogously, from Propositions 6 and 7, it follows that when trustworthiness is underestimated, the advance payment contract is optimal. Proposition 6 suggests that advance payment contracts could serve as a screening device. That is, they could help separate trustworthy agents from untrustworthy agents. If the buyer is trustworthy, he should not object to paying the whole price in advance. Therefore, a buyer who declines an advance payment contract is not as trustworthy as he claims to be.

Compare both types of contracts. In both, social welfare is maximized and the maximal output is produced. The only difference between the two contract types is the distribution of welfare. In the case of an uncertain payment contract, it is divided equally between the agents. In the case of an advance payment contract, it is divided as a proportion

$$\frac{U_S}{U_B} = \frac{\alpha}{\beta}. \tag{15}$$

According to the following proposition, uncertain payment and advance payment contracts can be regarded as different types of the Nash bargaining solution (Nash 1950).

*Proposition 8.* The advance payment contract is equivalent to the uncertain payment contract concluded with a nonsymmetric Nash bargaining function

$$(V(q) - P)^{\beta}(P - C(q))^{\alpha}. \tag{16}$$

Proof. Analogous to the proof of Proposition 1. □

## 5. AN INCENTIVE-COMPATIBLE NEGOTIATION MECHANISM UNDER UNCERTAIN LEVEL OF TRUST

In this section we drop the assumption that the buyer honestly declares his estimate $\beta$. Let the declared trustworthiness be $\hat{\beta}$ which may differ from the actual trust estimate, $\beta$. We analyze the possibility of strategic manipulation on the part of the buyer. Let $\alpha$ still be the seller's estimate of $\gamma$. As usual, we assume that $\alpha$ and $\hat{\beta}$ are common knowledge. Propositions 3–5 might suggest that it is profitable for the buyer to declare less trustworthiness than he deserves. For example, the buyer could agree with the seller and accept the seller's estimate of his trustworthiness. That is, the buyer could declare $\hat{\beta} = \alpha$. However, according to the following proposition it is not beneficial for the buyer to declare less trustworthiness than he deserves.

*Proposition 9.* If $0 < \alpha < \beta$, then it is not beneficial for the buyer to declare $\hat{\beta}$, $\hat{\beta} < \beta$.

Proof. If the buyer declares the true value $\beta$, his utility will be

$$U_B(q_2) = \frac{\alpha V(q_2) - \beta C(q_2)}{2\alpha}.$$

If the buyer declares $\hat{\beta}$, $\hat{\beta} < \beta$, his utility is

$$\begin{aligned} U_B(q_1) &= V(q_1) - \beta \frac{\alpha V(q_1) + \hat{\beta} C(q_1)}{2\alpha\hat{\beta}} \\ &= \frac{(2\alpha\hat{\beta} - \alpha\beta)V(q_1) - \hat{\beta}\beta C(q_1)}{2\alpha\hat{\beta}}. \end{aligned}$$

Since $\hat{\beta} < \beta$, we have

$$2\alpha\hat{\beta} - \alpha\beta < \alpha\hat{\beta}.$$

Therefore,

$$\begin{aligned} \frac{(2\alpha\hat{\beta} - \alpha\beta)V(q_1) - \hat{\beta}\beta C(q_1)}{2\alpha\hat{\beta}} &< \frac{\alpha\hat{\beta} V(q_1) - \hat{\beta}\beta C(q_1)}{2\alpha\hat{\beta}} \\ &= \frac{\alpha V(q_1) - \beta C(q_1)}{2\alpha} \leq \frac{\alpha V(q_2) - \beta C(q_2)}{2\alpha}. \end{aligned}$$

Thus, $U_B(q_2) > U_B(q_1)$. □

The buyer can still gain by declaring higher values of trustworthiness ($\hat{\beta} > \beta$). The possibility of such strategic manipulation depends on the particular trust level that the seller places on him ($\alpha$), as well as on the value function and the cost function. For example, suppose that the value and the cost are defined by (6) and (7), respectively, the true value of $\beta$ is 0.5 and $\alpha = 0.375$. The utility that the buyer obtains for different values of $\hat{\beta}$ is shown in Table 1.

The data in Table 1 show that for this particular case, the buyer has a strong incentive to declare higher trustworthiness. Therefore, the uncertain payment contract is not incentive-compatible when the level of trust is uncertain. From Proposition 7 it follows that the advance

TABLE 1. Utility and Quantity as a Function of Trustworthiness

| Declared value $\hat{\beta}$ | Quantity exchanged $Q$ | Buyer's Utility $U_B(q)$ |
|---|---|---|
| 0.5 | 2.64 | 7.87 |
| 0.6 | 2.45 | 9.48 |
| 0.7 | 2.30 | 10.48 |
| 0.8 | 2.18 | 11.13 |
| 0.9 | 2.08 | 11.55 |
| 1 | 2 | 11.84 |

payment contract is also not incentive-compatible. Since in an advance payment contract the buyer receives utility proportional to his declared trustworthiness, he will always declare the maximum trustworthiness ($\hat{\beta} = 1$). Therefore, the symmetric Nash bargaining solution and the nonsymmetric one given by (16) cannot guarantee that the buyer will truthfully reveal the level of his trustworthiness. Let us consider the following nonsymmetric Nash function:

$$F = (V(q) - \hat{\beta} P)^{\alpha} (\alpha P - C(q))^{\hat{\beta}}.$$

By an *F-contract* we mean a contract in which the price, $P$, maximizes the function $F$ and the quantity produced, $q$, maximizes the agents' utility functions. Every F-contract can be either an uncertain payment contract or an advance payment contract. The next proposition establishes desirable properties of the advance payment F-contract.

*Proposition 10.* If the agents choose an advance payment F-contract, then according to the Nash bargaining solution the amount of advance payment is

$$P_0 = \frac{V(q_1) + C(q_1)}{2}.$$

The quantity exchanged, $q_1$, maximizes $V(q) - C(q)$, the amount of contract payment is zero ($P = 0$) and

$$U_B(q_1) = U_S(q_1) = \frac{V(q_1) - C(q_1)}{2}.$$

Proof. Analogous to the proof of Proposition 6. □

Proposition 10 says that every advance payment F-contract maximizes social welfare. In addition, every advance payment F-contract is egalitarian, i.e., it equalizes individual utilities. Another advantage of an advance payment F-contract is that the buyer cannot benefit by declaring a false level of trustworthiness.

*Proposition 11.* In an advance payment F-contract, the buyer cannot benefit by revealing a false level of trustworthiness.

Proof. Since $U_B(q_1)$ does not depend on $\beta$, the buyer has no incentive to declare an underestimated or overestimated value of $\beta$. □

An advance payment F-contract can be used whenever the agents hold inaccurate estimates of trustworthiness. In particular, an advance payment F-contract can be used when the buyer is distrusted. Proposition 11 guarantees that such a contract eliminates the inefficiencies caused by information asymmetry.

## 6. CONCLUSIONS

In the paper we analyzed the impact of trust on market efficiency, and especially on multiagent negotiation. We proved that if agents hold accurate trust estimates about one another, then the social welfare, the amount of trade, and the agents' utility levels are maximized. We also showed that market efficiency does not require complete trustworthiness. Untrustworthy agents could transact as efficiently as trustworthy agents, provided that they hold accurate estimates of one another. Therefore, what really matters is not the actual level of trustworthiness, but the accuracy of individual estimates. A market in which agents are trusted to the degree they deserve to be trusted is as efficient as a market with complete trustworthiness.

We also proved that distrust could have a detrimental effect on market efficiency. Namely, it reduces the amount of trade, individual profits, and the social welfare. We demonstrated that an attempt by an agent to insure himself by placing a reduced amount of trust on the contract partner may threaten the success of cooperation.

In the paper we analyzed the advance payment contract as a solution to the problem of distrust. We proved that every advance payment is not only individually rational, but it also maximizes the social welfare and the amount of trade. In addition, advance payment contracts could serve as a screening device. That is, they could be used to separate trustworthy agents from those that are untrustworthy.

In the paper we also studied the problem of how to make an untrustworthy agent truthfully report his actual level of trustworthiness. This is of special importance to risky transactions in which agents could lie about or misrepresent their trustworthiness. We proposed the advance payment F-contract as an efficient and incentive-compatible solution to the problem of lying. We showed that in an advance payment F-contract agents cannot benefit by declaring false trustworthiness.

The analysis and the solutions proposed in this paper could help agent designers avoid many market failures and produce efficient interaction mechanisms. The applicability of the solutions proposed in the paper is not limited to e-commerce applications only. The results could be applied to any problem involving negotiation and contracting with a untrustworthy agent, such as task and resource allocation, multiagent planning, etc.

There are many open questions for further research. Given the results of this paper, the problem of learning accurate trust estimates is especially important for market efficiency. Another problem of practical importance is how to design an interaction mechanism for efficient trust management. We hope that our work will facilitate further research on this topic as well.

## REFERENCES

AUMANN, R. 1976. Agreeing to disagree. The Annals of Statistics **4**:1236–1239.

BECKER, G., and W. LANDES. 1974. Essays in the Economics of Crime and Punishment. Columbia University Press for the National Bureau of Economic Research, New York.

BRAINOV[*sic*], S., and T. SANDHOLM. 1999. Power, dependence and stability in multiagent plans. *In* Proceedings of the National Conference on Artificial Intelligence (AAAI'99), Orlando, pp. 11–16.

BRAINOV[*sic*], S. 1994. Deviation-proof plans in open multiagent environments. *In* Proceedings of the European Conference on Artificial Intelligence (ECAI'94), Amsterdam, pp. 274–278.

BRAINOV[*sic*], S. 2001. An incentive compatible trading mechanism for trust revelation. *In* Proceedings of the IJCAI'01 Workshop on Economic Agents, Models and Mechanisms, pp. 62–70.

BRAYNOV, S., and T. SANDHOLM. 2002. Trust revelation in multiagent interaction. *In* Proceedings of CHI'02 Workshop on the Philosophy and Design of Socially Adept Technologies, Minneapolis, pp. 57–60.

CASTELFRANCHI, C., and R. FALCONE. 1999. The dynamics of trust: From beliefs to action. *In* Proceedings of the Second Workshop on Deception, Fraud and Trust in Agent Societies, pp. 41–54.

COLEMAN, J. 1990. Foundations of Social Theory. Harvard University Press, Cambridge, MA.

DASGUPTA, P. 1990. Trust as a commodity. *In* Making and Breaking Cooperative Relations. *Edited by* D. Gambeta. Basil Blackwell, Oxford.

DASKALOPULU, A., T. DIMITRAKOS, and T. MAIBAUM. 2002. Evidence-based electronic contract performance monitoring. INFORMS Journal of Group Decision and Negotiation (in press).

DELLAROCAS, C. 2001. Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. *In* Proceedings of the ACM Conference on Electronic Commerce, Tampa, Florida.

DIMITRAKOS, T. 2001a. System models, e-risk and e-trust. *In* 1st IFIP Conference on E-Commerce, E-Business, E-Government. Kluwer Academic Publishers, Zurich, Switzerland.

DIMITRAKOS, T. 2001b. Towards a formal model of trust in e-commerce. *In* Proceedings of the the AI-2001 Workshop on Novel E-Commerce Applications of Agents, Ottawa, Canada.

FRIEDMAN, E., and P. RESNICK. 1998. The social cost of cheap pseudonyms. Working Paper.

JØSANG, A. 1999. Trust-based decision making for electronic transactions. *In* Proceedings of the Fourth Nordic Workshop on Secure Computer Systems, Sweden, pp. 496–502.

JØSANG, A. 2001. A logic for uncertain probabilities. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, pp. 279–311.

KAPLAN, S., and B. GARRIK. 1981. On the quantitative definition of risk. Risk Analysis **28**:11–27.

MARSH, S. 1994. Formalizing trust as a computational concept. Ph.D. thesis, University of Stirling, UK.

MATSUBARA, S., and M. YOKOO. 2000. Defection-free exchange mechanism for information goods. *In* Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS), Boston, pp. 183–190.

NASH, J. 1950. The bargaining problem. Econometrica **28**:155–162.

OSBORNE, M., and A. RUBINSTEIN. 1990. Bargaining and markets. Academic Press, New York.

RESNICK, P., R. ZECKHAUSER, F. FRIEDMAN, and K. KUWABARA. 2000. Reputation systems. Communications of the ACM **43**:45–48.

SANDHOLM, T., and V. FERRANDON. 2000. Safe exchange planner. *In* Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS), Boston, 255–262.

SANDHOLM, T., and V. LESSER. 1995. Equilibrium analysis of the possibilities of unenforced exchange in multiagent systems. *In* Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Montreal, Canada, pp. 694–701.

SLOVIC, P. 1997. Risk perception and trust. *In* Trust: Fundamentals of Risk Analysis and Risk Management. *Edited by* V. Molak. Lewis Publishers, Boca Raton, Florida.

SNIJDERS, C. 1996. Trust and commitments. Ph.D. thesis, University of Utrecht, The Netherlands.

ZACHARIA, G., and P. MAES. 2000. Trust management through reputation mechanisms. Applied Artificial Intelligence **14**:881–907.