

Does making specific investments unobservable boost  
investment incentives?\*

Randolph Sloof

University of Amsterdam

School of Economics and Tinbergen Institute

Roetersstraat 11, 1018 WB Amsterdam

the Netherlands

r.sloof@uva.nl

Hessel Oosterbeek

University of Amsterdam

School of Economics and Tinbergen Institute

Roetersstraat 11, 1018 WB Amsterdam

the Netherlands

h.oosterbeek@uva.nl

Joep Sonnemans

University of Amsterdam

School of Economics and Tinbergen Institute

Roetersstraat 11, 1018 WB Amsterdam

the Netherlands

j.h.sonnemans@uva.nl

Suggested running head: Unobservability and investment incentives

---

\*An earlier version of this paper was presented at conferences at Harvard University (ESA), in Erfurt (ESA) and in Madrid (EEA), and at seminars at NYU, the University of Bonn and the University of Rotterdam. We are grateful to two anonymous referees for their helpful remarks which improved the paper considerably.

**Abstract:** Standard theory predicts that holdup can be alleviated by making specific investments unobservable; private information creates an informational rent that boosts investment incentives. Empirical findings, however, indicate that holdup is attenuated by fairness and reciprocity motivations. Private information may interfere with these, as it becomes impossible to observe whether the investor behaved fair or not. In that way unobservability could crowd out an informal fairness/reciprocity mechanism in place. This paper reports on an experiment to investigate this issue empirically. Our results are in line with standard predictions when there is limited scope for social preferences. But with sufficient scope for these motivational factors, unobservability does not boost specific investments.

# 1 Introduction

In a world with incomplete contracts, relationship-specific investments are vulnerable to appropriation by the trading partner. After investment costs are sunk, the other party may force a renegotiation to obtain a larger share of the ex post surplus. In that way this party can capture some of the returns on investment without sharing in the costs. Anticipating this, the investor will invest less than the efficient level. This is the well-known holdup underinvestment problem (cf. Klein et al. 1978, Williamson 1985).

Acknowledging the risk of potential holdup, the theoretical literature has come up with a number of possible solutions that effectively protect the investor's return on investment. One of these solutions concerns the privacy of the investment decision itself. When only the investor observes the size of the actual investment made – and, as a result, has private information about the size of the actual quasi-surplus so created – she obtains an informational rent in the renegotiation stage. This increases her marginal return on investment and thus generates stronger investment incentives. The above intuition is used by Gul (2001, p. 344)

“...to emphasize the role of allocation of information as a tool in dealing with the hold-up problem. Audits, disclosure rules or privacy rights could be used to optimize the allocation of rents and guarantee the desired level of investment. Controlling the flow of information may prove to be a worthy alternative to controlling bargaining power in designing optimal organizations.”

Existing empirical evidence, however, casts doubt on the effectiveness of investment unobservability as an instrument against holdup. A large number of experimental studies have namely demonstrated that, with symmetric information, the underinvestment problem is less severe than predicted. A plausible explanation for this finding is that subjects are partly guided by social preferences, like fairness and reciprocity. Investment is typically seen as fair or kind behavior, which is therefore rewarded by the non-investor with a larger than predicted return. This in turn makes it worthwhile to invest more than standard theory predicts. With such an informal fairness/reciprocity mechanism in place, making the investment private information

may potentially be harmful. The intuitive idea here is that, because fair/kind behavior can be more easily identified when the investment is observable, the introduction of private information may crowd out the positive effect on investment incentives brought about by social preferences. Private information may therefore not be as effective in enhancing investment incentives. This paper reports an experiment designed to test this hypothesis.

Our experiment concerns a two-stage game between a buyer and a seller. The buyer first decides on an investment that raises her valuation of the good. After that the seller unilaterally determines the trading price. In one condition the seller observes the buyer's investment decision, in another one he does not. The important consequence is that in the latter case the seller does not know the buyer's valuation when setting his price. Trade takes place only when the seller's price does not exceed the buyer's valuation. Otherwise the seller gets nothing and the buyer bears the cost of investment (if applicable). Within these two information conditions, (commonly known) investment costs can take three values: low, intermediate or high.

Standard theory predicts that investment unobservability increases investment levels by the same positive amount, independent of the costs of investment. The informational rent the buyer obtains thus always boosts her investment incentives. In contrast, models based on social preferences predict that with lower investment costs the effect of unobservability is smaller and may even disappear. The intuition is that lower investment costs increase the scope for motivational factors like fairness and reciprocity. And when this scope is large, investment levels are predicted to be independent of whether the investment is observable or not. If, however, this scope is small, the predictions of standard theory pertain.

Our results are in line with the social preferences predictions. When investment costs are high, such that there is limited scope for fairness and reciprocity, mean investments are substantially higher in the unobservable investment case than in the observable investment situation. Moreover, these mean levels are almost identical to the ones predicted by standard theory. With intermediate investment costs, mean investment levels in the two information conditions become more equal. And with low investment costs – so that there is sufficient scope for social preferences – mean investment levels in the two conditions are the same.

The remainder of this paper is organized as follows. In the next section we provide a brief

review of related literature. The third section presents the simple game on which our experiment is based. This section also discusses the standard equilibrium predictions, as well as alternative predictions based on a specific model of social preferences, viz. intention-based reciprocity. Section 4 provides the details of the experimental design. Results are presented and discussed in Section 5. The final section concludes.

## 2 Related literature

This paper focuses on the interaction between private information and social preferences in mitigating holdup. Two strands of literature are particularly relevant here, viz. (i) theoretical models showing and building upon the idea that private information rents may enhance investment incentives and (ii) empirical (experimental) studies documenting the extent of the underinvestment problem in the presence of social preferences. These are briefly discussed in turn.

### 2.1 Theoretical models

The idea that investment unobservability may alleviate underinvestment originates from Tirole (1986). For a very general class of bargaining processes he shows that an investor invests less when the (non-contractible) investment is observable than when it is unobservable (see his Proposition 3). The intuition here is that privacy of the investment decision creates private information about the size of the ex post surplus, and thus enables the investor to capture an informational rent in the renegotiations. This in turn improves her investment incentives. Gul (2001) complements unobservable investment with a specific dynamic bargaining protocol in which only the non-investor makes frequently repeated offers. He shows that in the limit, when the time between successive offers becomes negligible, the investor invests efficiently. Rogerson (1992), Konrad (2001), Lau (2002) and Gonzalez (2004) also develop and build on the intuition that private information rents create investment incentives.

A potential downside of asymmetric information at the renegotiation stage is that efficient trade is no longer guaranteed. There thus may exist a tradeoff between efficient trade deci-

sions and ‘high-powered’ investment incentives. This tradeoff may be relevant in a variety of contexts. Riordan (1990), for instance, argues that vertical integration leads to a change in information structure; the downstream firm becomes better informed about upstream costs. This weakens the upstream firm’s incentives to invest in cost reduction. The choice between vertical integration and market contracting is then between distorted investment incentives and distorted production decisions. Schmidt (1996) identifies a similar tradeoff between public and private ownership. Under nationalization the government has precise information about a firm’s costs and profits, but under privatization it has not. The costs of privatization are then a less efficient production level, while the benefits amount to better incentives for managers to save on production costs. Finally, Cremer (1995) argues that the choice of monitoring technology can be seen as a commitment device. Unobservable investments then reflect a situation in which the investor keeps the non-investor at ‘arm’s length’. This enables the non-investor to commit to a single unconditional trade price. Under observable investment such a commitment is non-credible and thus cannot be used to provide investment incentives. Without commitment the parties can always take the efficient trade decision though.

## 2.2 Empirical studies

Data limitations hamper a clean assessment of the extent of the underinvestment problem using field data. This holds in particular because the root of the holdup problem lies in the non-verifiability of the investment, i.e. the investment is difficult to objectively measure and observe. Most empirical studies therefore rely on laboratory experiments. By now a vast number of holdup experiments have been conducted. The typical finding is that, in complete information settings, subjects on average invest significantly more than predicted.<sup>1</sup> The explanation usually put forward in these studies is positive reciprocity. Investment can be considered kind because it increases the joint surplus that can be divided, and is therefore rewarded with a larger than predicted return. This in turn stimulates investment.<sup>2</sup> Recent models of social preferences

---

<sup>1</sup>See e.g. Berg et al. (1995), Ellingsen and Johannesson (2004a, 2004b), Fehr and List (2004), Gantner et al. (2001), Hackett (1993, 1994), Königstein (2000), Oosterbeek et al. (2003), Sloof et al. (2004), Sloof (2006) and Sonnemans et al. (2001).

<sup>2</sup>The empirical relevance of this informal mechanism is confirmed by the field data study of Leuven et al. (2005). They conduct a survey among a representative sample of the Dutch labor force. Among other things, the

that take fairness and/or reciprocity motivations into account, like the ones of e.g. Charness and Rabin (2002) and Dufwenberg and Kirchsteiger (2004), indeed rationalize the experimental findings.

Some of the earlier holdup experiments allude to the main idea motivating this paper that private information may crowd out reciprocal behavior. Hackett (1994) considers a setting in which higher investments increase the probability that the publicly observable quasi-surplus is large (see Oosterbeek et al. (2006) for a similar setting). Because informational rents are absent by construction, standard theory predicts that investment levels are independent of whether they are observable or not. However, Hackett finds that investments are somewhat higher when they are observable. In the experiment of Ellingsen and Johannesson (2005) an investor makes an ultimatum offer about how to divide the (observable) surplus created by her investment. The actual costs of investment are either privately or publicly observed. The authors first formally show that when non-investors are motivated by fairness considerations, they may under private information rationally fear being treated unfairly and therefore turn down profitable (and fair) offers. This in turn may induce high-cost investors to refrain from investing. Nevertheless, in line with standard predictions Ellingsen and Johannesson find that investment rates do not differ significantly between information conditions.

Although especially the findings of Hackett (1994) are suggestive, they are in itself inconclusive about whether informational rents are an effective instrument in mitigating holdup. The reason is that in all these papers the ex post surplus is always public information, such that informational rents are absent at the bargaining stage. Standard theory therefore predicts that investment levels are independent of whether the actual investment costs borne are observable or not. The characterizing feature of the present experiment is that in our unobservable investment condition both *the investment itself* and *its actual return* are private information. Hence the actual surplus up for renegotiation is known to the investor only, and she is predicted to obtain an informational rent in the bargaining stage. According to standard theory, this data set contains information about participation in work-related training, who paid for this training, and the worker's reciprocal attitude. The firm-sponsored training rate of workers with a high sensitivity to reciprocity appears to be around 48%, while it only equals 33% for workers with a low sensitivity to reciprocity. The difference of 15 percentage points is highly significant. Employers thus seem to actively rely on reciprocity as informal remedy against holdup.

improves her investment incentives as compared to the situation in which the investment (and its return) is publicly observed. The next section explains this in more detail.

### 3 Theory

#### 3.1 Basic setup of the model

Consider a bilateral relationship between a female buyer and a male seller. Both parties are assumed to be risk neutral. The order of play is as follows:

1. The buyer decides whether to make a specific investment ( $I = 1$ ) or not ( $I = 0$ ). Investment costs equal  $C$  and are immediately borne by the buyer. Without investment her valuation of the seller's good equals  $V$ , with investment this becomes  $V + W$ .
2. The seller makes a price demand  $P \in [0, V + W]$  for which he is willing to sell the good. In case his price is weakly below the buyer's actual valuation, trade takes place at the demanded price. Otherwise, trade does not take place.

The second stage captures in reduced form a situation in which the buyer has (almost) no bargaining power at all. She is forced to accept the seller's price demand as long as this demand is weakly below her valuation. If the demanded price is higher, her only option is to reject. Clearly this is not particularly realistic. In practice parties can at least indicate whether they accept or reject the terms of trade. We focus on the reduced form specification though, because it makes the equilibrium analysis under social preferences much simpler and the interpretation of the results more clear cut. In the concluding section we will return to this issue.

The seller's valuation is unaffected by the investment and normalized to zero. We assume that  $0 < C < W$ . This implies that making the investment is efficient. We also assume that  $V > 0$ , such that trade is always efficient. Maximum net overall surplus equals  $V + W - C$ .

Two different information conditions are considered. First, in the *observable investment* case the buyer's investment decision is publicly observable. Here the seller knows the buyer's valuation when he chooses his price demand  $P$ . Second, in the *unobservable investment* case the seller does not observe the buyer's investment choice. Then the seller does not know what the

Table 1: Reduced strategic form under unobservable investment

	$P = V$	$P = V + W$
$I = 0$	0, $V$	0, 0
$I = 1$	$W - C$ , $V$	$-C$ , $V + W$

buyer's actual valuation is when he makes his price demand. This situation is formally equivalent to one in which the buyer and the seller simultaneously decide on  $I$  and  $P$ , respectively. In both information conditions the setup of the game and the values of  $V$ ,  $W$  and  $C$  are common knowledge. In particular, also in the unobservable investment case the seller knows the costs of investment  $C$  (although he does not observe whether the investment is actually made or not).

### 3.2 Standard equilibrium predictions

Consider first the observable investment case. Solving the game through backward induction, the seller chooses  $P_1^* = V + W$  after investment and  $P_0^* = V$  after no investment. Here  $P_I^*$  denotes the seller's equilibrium price after observed investment decision  $I \in \{0, 1\}$ . Anticipating this pricing strategy, the buyer will not invest in order to save on the investment costs. Hence the unique subgame perfect equilibrium predicts holdup to be complete:  $q_{obs}^* \equiv \Pr(I = 1) = 0$ . There is no trade inefficiency, because the buyer and the seller always trade. Predicted net social surplus equals  $V$ . The efficiency loss owing to holdup is  $W - C$ .

In the unobservable investment case the seller cannot condition his price on the buyer's investment decision. Although he may demand any price in  $[0, V + W]$ , in equilibrium he will choose between  $P = V$  and  $P = V + W$  only. The reduced strategic form therefore corresponds to the  $2 \times 2$  simultaneous-move game depicted in Table 1. This game has a unique mixed-strategy equilibrium:  $q_{un}^* = \frac{V}{V+W}$  and  $p^* \equiv \Pr(P = V) = \frac{C}{W}$ .

Our interest lies in the effect of investment unobservability on the propensity to invest. The above analysis yields the following prediction:

**Standard theory**  $q_{un} - q_{obs}$  is positive and independent of  $C$ ; private information always boosts investment incentives.<sup>3</sup>

---

<sup>3</sup>Risk aversion does not affect this prediction. To see this, note that the buyer always chooses  $q_{obs}^* = 0$ ,

Our two treatments reflect the tradeoff between efficient trade and high-powered investment incentives discussed in Subsection 2.1. To illustrate, compared to the observable investment case, privacy of the investment decision induces an efficiency gain of  $q_{un}^* \cdot (W - C)$ . At the same time it also introduces inefficient separations with probability  $(1 - q_{un}^*) \cdot (1 - p^*)$ . Inefficient separations occur when the seller demands a high price while the buyer did not invest. In that case the potential surplus of trade  $V$  is wasted. In our simple setup the expected gain owing to more investment and the expected loss due to inefficient separations actually cancel out; expected net social surplus under unobservable investment also equals  $V$ .

### 3.3 Predictions based on social preferences

Based on existing experimental evidence it was suggested in the Introduction that private information may potentially crowd out the investment incentive effects due to social preferences. For our simple game this hypothesis can be made formal using the theory of intention-based reciprocity as developed by Rabin (1993) and further refined by Dufwenberg and Kirchsteiger (2004). For ease of exposition we assume that the buyer is selfish and motivated by money maximization only.<sup>4</sup> The seller may be motivated by reciprocity though, implying that he may be willing to sacrifice in order to reward the buyer's good intentions and/or to punish her bad intentions. In particular, following Dufwenberg and Kirchsteiger the seller's utility equals:

$$u_S = \pi_S + Y_S \cdot \kappa \cdot \lambda.$$

Here  $\pi_S$  denotes the seller's monetary payoffs whereas term  $Y_S \cdot \kappa \cdot \lambda$  gives his reciprocity payoffs.

---

independent of her own or the seller's risk attitude. In the unobservable investment case the equilibrium is necessarily in mixed strategies. The buyer's mixing probability  $q_{un}^*$  should then make the seller indifferent, so it only depends on the risk attitude of the latter. Now the more risk averse the seller is, the larger  $q_{un}^*$  will be. The intuition for this is that in equilibrium the seller necessarily mixes between the safe option  $P = V$  and the risky option  $P = V + W$ . The latter price is risky, because a profitable sale is lost if the buyer did not invest. The more risk averse the seller is, the more concerned he is about this potential loss and the more he likes to play safe by choosing  $P = V$ . To counterbalance this and to make the seller indifferent again such that he is willing to mix,  $q_{un}^*$  should increase to reduce the probability of losing a profitable sale when  $P = V + W$ . Overall, if anything risk aversion thus only strengthens the incentive effect of private information.

<sup>4</sup>Dufwenberg and Kirchsteiger (2000) make the same simplifying assumption in their analysis of employer-worker relationships. In the earlier working-paper version Sloof et al. (2005) we provide the complete equilibrium analysis for the more general case in which also the buyer may be reciprocal. This leads to the same qualitative predictions.

Parameter  $Y_S \geq 0$  reflects the reciprocal attitude of the seller. The higher  $Y_S$ , the more sensitive to reciprocity he is. Factor  $\kappa$  measures the seller's kindness towards the buyer. This factor is positive if the seller is kind to the buyer and negative if he is unkind to her. Here kindness is measured with reference to the range of monetary payoffs the seller thinks he could give the buyer in principle. Factor  $\lambda$  gives the seller's belief about how kind the buyer is to him. It is positive when the seller believes that the buyer is kind to him, and negative when he thinks she is unkind. Dufwenberg and Kirchsteiger (2004) provide exact definitions of how  $\kappa$  and  $\lambda$  are calculated. The key ingredient of the model is that a reciprocal seller has an incentive to match the sign of her own kindness  $\kappa$  with the sign of the perceived kindness  $\lambda$  of the buyer.

Because the reciprocity payoffs depend on the players' beliefs, psychological game theory is needed to derive equilibrium predictions. Within this framework Dufwenberg and Kirchsteiger define and prove the existence of a *sequential reciprocity equilibrium* (SRE). This concept requires each player to maximize his utility given correct beliefs, and also invokes a subgame perfection requirement. The formal equilibrium analysis is somewhat involved and is therefore relegated to Appendix A. Table 2 summarizes the main insights from the analysis.

Table 2 and the corresponding intuition can be understood as follows. First consider the observable investment case. If the seller observes that the buyer did not invest, he will (in equilibrium) interpret this as unkind behavior. Investment, on the other hand, is seen as kind. The buyer can thus always convince the seller of her kindness, simply by choosing to invest. Investment is then rewarded by the seller by claiming less than the available surplus. In particular, the seller demands a price equal to  $P_1^* = V + \frac{2}{Y_S}$ . Note that this price gives the buyer a larger return on investment the more reciprocal the seller is. If he is sufficiently reciprocal, i.e. if  $Y_S$  is high enough, the buyer gets sufficient incentives to invest. This happens when the return on investment  $W - \frac{2}{Y_S}$  exceeds the investment costs  $C$ . *Positive reciprocity* can thus effectively operate as an investment incentive instrument only for  $C$  sufficiently small (or, equivalently, for  $Y_S$  sufficiently high). Otherwise the buyer does not invest at all.

Next turn to the unobservable investment case. Because the seller in that case does not observe the buyer's actual kindness, he has to form beliefs about her *intended* kindness. Although in equilibrium these beliefs are necessarily correct, they can be self-fulfilling. Suppose,

for instance, that the seller starts from the presumption that the buyer is kind, i.e. that she is inclined to invest. Given no factual observation to the contrary, there is no reason to update this prior belief. A reciprocal seller will thus prefer to reward the buyer with an expected return on investment. The more reciprocal the seller is, the higher this reward. For a sufficiently reciprocal seller the buyer is then indeed induced to invest, corroborating the seller’s initial belief that the buyer is kind. Like under observable investment, therefore, a positive reciprocity equilibrium with  $q_{un}^* = 1$  exists for sufficiently high values of  $Y_S$ . As Table 2 reveals it actually holds that  $q_{un}^* = 1$  is possible for a larger set of  $Y_S$ -values than  $q_{obs}^* = 1$  is. The scope for positive reciprocity is thus larger under unobservable investment.<sup>5</sup>

However, the seller may also start from an a priori belief that the buyer is unkind. Given that investment is unobservable, the buyer can do nothing to change this belief. The seller thus finds no reason to reward her with an expected return on investment. And without such a return the buyer lacks strong incentives to invest, again confirming the seller’s initial belief. Because there is no way in which the buyer can “prove” to the seller that she is actually kind, the parties may always end up in such a *negative reciprocity* equilibrium. That is, for any value of  $Y_S$  there exists a SRE in which the buyer invests with low probability  $0 < q_{un}^* \leq \frac{V}{V+W}$ .<sup>6</sup> As explained above, this is not the case under observable investment, because there kind behavior can always easily be identified. The scope for negative reciprocity is thus also larger when the investment is private information.

The main conclusion that follows from Table 2 is that when reciprocity considerations are weak, the buyer invests more under unobservable investment than under observable investment (just like standard theory predicts). This situation becomes more likely the higher are the costs of investment  $C$  (while keeping  $W$  fixed). However, when the seller is sufficiently sensitive to intention-based reciprocity, private information does not boost investment incentives. This case

---

<sup>5</sup>As Theorem 1 in Appendix A makes clear, for  $Y_S > \frac{2}{W-C} \cdot \frac{W}{V+W}$  also a (third) SRE exists in which  $\frac{1}{2} < q_{un}^* < 1$ . This equilibrium is also based on positive reciprocity and therefore also predicts a higher investment rate than under standard theory (note that  $V < W$  implies that  $\frac{V}{V+W} < \frac{1}{2}$ ). Because the existence of this SRE does not affect the main qualitative predictions nor the intuition, we have left it out from Table 2.

<sup>6</sup>Note that  $q_{un}^* = 0$  cannot occur. This holds because strategy  $q_{un} = 0$  is never interpreted by the seller as kind and he thus chooses  $P = V$  in response. But when  $P = V$  for sure, the selfish buyer prefers to invest, i.e. to choose  $q_{un} = 1$ . More generally, in the negative reciprocity equilibrium  $q_{un}^*$  must make the seller indifferent between  $P = V$  and  $P = V + W$ , taking both his monetary payoffs and his reciprocity payoffs into account.

Table 2: Predicted investment probabilities

		Observable	Unobservable
Standard theory	All $C$	$q_{obs}^* = 0$	$q_{un}^* = \frac{V}{V+W}$
Social preferences:			
Weak	$C > V + W - \frac{2}{Y_S}$	$q_{obs}^* = 0$	$0 < q_{un}^* \leq \frac{V}{V+W}$
Medium	$W - \frac{2}{Y_S} < C < V + W - \frac{2}{Y_S}$	$q_{obs}^* = 0$	$0 < q_{un}^* < \frac{V}{V+W}$ or $q_{un}^* = 1$
Strong	$C < W - \frac{2}{Y_S}$	$q_{obs}^* = 1$	$0 < q_{un}^* < \frac{V}{V+W}$ or $q_{un}^* = 1$

*Remark:*  $q$  denotes the probability of investment. In the ‘Medium’ and ‘Strong’ case multiple equilibria exist side by side when the investment is unobservable. The analysis under social preferences assumes that  $V < W$ .

is likely to apply when  $C$  is relatively low.<sup>7</sup> We thus obtain the following qualitative prediction based on social preferences:

**Social preferences**  $q_{un} \leq q_{obs}$  when  $C$  is low and  $q_{un} > q_{obs}$  when  $C$  is high; private information enhances investment incentives only when  $C$  is high.

These predictions also imply that, relative to the standard predictions, the increase in the investment rate owing to social preferences is larger when the investment is observable than when it is unobservable. The intuition that the impact of reciprocity on investment incentives is much more substantial in the observable investment case is the main driving force behind our crowding out hypothesis.

Clearly, in practice subjects are heterogeneous. Some care strongly about the intended kindness of others whereas others are completely selfish. Hence even at a low cost level a fraction of subjects is likely to behave selfish. Likewise, even at a high cost level some subjects may reveal a concern for reciprocity. Yet we expect that when we aggregate over all subjects, the above qualitative predictions will pertain. It is also important to point out that the same qualitative

<sup>7</sup>Comparing the equilibrium in which  $q_{un}^* < \frac{V}{V+W}$  with  $q_{obs}^* = 1$  in the ‘Strong’ case suggests that private information may even *weaken* investment incentives. This only strengthens our qualitative prediction that when  $C$  is low, private information cannot be used as an instrument to encourage investments.

predictions are obtained when we assume instead that subjects care about the final distribution of payoffs, rather than about how this distribution came about (see the earlier working-paper version Sloof et al. (2005)). In particular, assuming quasi-maximin preferences introduced by Charness and Rabin (2002) or inequality-averse subjects like in Fehr and Schmidt (1999) also yields the prediction that when social preferences are effectively important (i.e. when  $C$  is low), private information has less or even no impact on investment behavior. Our experiment thus should not be taken as a test of the intention-based reciprocity model per se.

## 4 Experimental design

The experiment is based on a  $2 \times 3$  design. For both the observable and the unobservable investment case we considered three levels of investment costs:  $C \in \{20, 40, 60\}$ . The two other parameters always equalled  $V = 50$  and  $W = 80$ . We ran six sessions in total, all of them in April 2002. Three sessions considered the observable investment case, the other three the unobservable investment case. All subjects within a session were confronted with all three values of  $C$ . Overall 120 subjects participated, with 20 participants per session. The subject pool consisted of the undergraduate student population of the University of Amsterdam. Sixty percent were students in economics, 64 percent of the participants were male.<sup>8</sup> Average earnings were 27.65 euros in about one and a half hours. Earnings varied considerably though, with the minimum actual earnings equal to 7.60 euros and a maximum of 51.50 euros.

The sessions in which the investment was observable necessarily displayed a sequential game structure. We also used a sequential decision structure in the unobservable investment case; subjects knew that buyers decided on their investment before sellers chose their price demand. We did so to make both information conditions fully comparable. To exclude dominated strategies in the observable investment case, the seller could never ask for more than the actual pie.

---

<sup>8</sup>Recruitment was done in two ways. First of all, the experimental research group CREED at the University of Amsterdam maintains a database of prospective participants, which mainly consists of people who have participated in some other (unrelated) CREED experiment before. We sent an e-mail announcement to a randomly selected sample from this database. Apart from that, we also recruited subjects through poster announcements in one of the main university buildings, which houses the two departments of economics & business and psychology (among some other much smaller ones). This explains why economics students are overrepresented in our sample. In Subsection 5.4 we verify whether this biases our results.

Figure 1 depicts the structure of the experimental games.

[ Insert Figure 1 about here ]

Each session contained 36 rounds. We employed a block structure to control for learning and order effects. In particular, we divided the 36 rounds into six blocks of six rounds. Within each block the costs of investment were kept fixed. In two out of three sessions per information condition we used the ‘upward’ order (20, 40, 60, 20, 40, 60) of investment costs. In the remaining session we employed the opposite ‘downward’ order of (60, 40, 20, 60, 40, 20). By comparing (within a session) different blocks that consider the same value of  $C$  we can test for learning effects. By comparing the two different orders we can control for order effects. The start of every new block and the change in the value of the investment costs  $C$  were both verbally announced and shown on the computer screen. Both buyers and sellers were thus explicitly informed about the exact costs of investment  $C$  that applied in each round.

From each block of six rounds we selected – before the experiment started – one round that was actually paid. After the final round, subjects learned which six rounds were selected and they obtained the number of points they had earned in these rounds, on top of their initial endowment of 75 points. (The conversion rate was one euro for 10 points.) Subjects were explicitly informed about this procedure at the start of the experiment. The rationale for paying only one round per block is that it strengthens the one-shot nature of each interaction.<sup>9</sup>

Subject roles’ varied over the rounds. Within each block each subject had the role of buyer three times, and the role of seller also three times.<sup>10</sup> The experiment used a stranger design. Subjects were anonymously paired and their matching varied over the rounds. Within each block subjects could meet each other only once. Subjects were explicitly informed about this. Moreover, within a session we divided the subjects into two groups of ten subjects. Matching of pairs only took place within these matching groups.

---

<sup>9</sup>This holds because subjects know that they cannot compensate gains or losses within the same block. In that way our payment procedure also makes it more likely that fairness/reciprocity motivations are restricted to each interaction in isolation.

<sup>10</sup>We used role switching for two reasons. First, it enhances subjects’ awareness of the other player’s decision problem. Alternating roles provide subjects with an opportunity to see things from the other player’s viewpoint and thus to understand the game better. Second, it also doubles the number of investors in the experiment.

The experiment was computerized. Subjects started with on-screen instructions. Before the experiment started all subjects had to answer a number of control questions correctly. They also received a summary of the instructions on paper.<sup>11</sup> At the end of the experiment subjects filled out a short questionnaire and the earned experimental points were exchanged for money.

## 5 Results

In presenting our results we pool the data from sessions that are completely similar in the order of treatments they consider, because no significant differences are found between these sessions. We also pool the results from sessions that differ only in the order of the  $C$ -values.<sup>12</sup> Although some order effects can be detected, these are only minor. Further aggregations are not possible, because it appears that behavior evolves over time. Most findings are therefore reported separately for the first and second half of the experiment.

### 5.1 Investment levels

Within each block of six rounds subjects have the role of buyer three times. For each subject we calculate for each block his or her mean investment level, which equals either 0,  $\frac{1}{3}$ ,  $\frac{2}{3}$  or 1. Statistical tests can then be based on a comparison of these *individual mean* investment levels. Per treatment we have 60 individual investors. In addition we perform our tests on the group level data. As discussed in Section 4 we divided the 20 subjects within a session into two groups that were independently matched. By doing so we created six independent group observations per treatment and we can compare the *group mean* investment levels across treatments. In the sequel we base our inferences on the results of both types of tests. If not stated otherwise, a significance level of 5% is employed.

The first result compares mean investment levels across information conditions.

**Result 1.** (a) *With high or intermediate investment costs, mean investment levels are higher*

---

<sup>11</sup>A direct translation of this summary sheet can be downloaded from: [www1.fee.uva.nl/scholar/mdw/sloof/InstructionsUnobservable.pdf](http://www1.fee.uva.nl/scholar/mdw/sloof/InstructionsUnobservable.pdf).

<sup>12</sup>The test results on session effects, order effects and learning effects are reported in a web-appendix available at: [www1.fee.uva.nl/scholar/mdw/sloof/WebAppendixUnobservable.pdf](http://www1.fee.uva.nl/scholar/mdw/sloof/WebAppendixUnobservable.pdf).

Table 3: Mean investment levels by treatment and tests for equality

	first: rounds 1-18			second: rounds 19-36		
	unobs.	obser.	$p$ -values	unobs.	obser.	$p$ -values
$C = 20$	.711	.772	.1304	.644	.572	.4335
	[.385]	[0]	.4192	[.385]	[0]	.2207
$C = 40$	.561	.461	.1957	.456	.189	.0000
	[.385]	[0]	.1481	[.385]	[0]	.0156
$C = 60$	.333	.122	.0001	.383	.078	.0000
	[.385]	[0]	.0215	[.385]	[0]	.0031

*Remark:* Standard equilibrium predictions (based on self-interest) within square brackets.  $p$ -values correspond to a Mann-Whitney ranksum test comparing the unobservable investment case with the observable investment case. For each level of  $C$  the upper (lower)  $p$ -value is based on individual (group) level data with  $m = n = 60$  ( $m = n = 6$ ).

*under unobservable investment than under observable investment. (b) With low investment costs, mean investment levels are independent of the information condition.*

Evidence supporting Result 1 is provided in Table 3. This table reports the mean investment levels by treatment and gives the test statistics for equality of these levels across treatments (ranksum tests). When  $C = 20$  the investment rate is independent of whether the investment itself is observable or not. In case  $C = 40$  we observe a significant difference only when subjects are confronted with this costs level during the second half of the experiment.<sup>13</sup> For  $C = 60$  the difference is significant for both halves, and largest in absolute and relative magnitude.

Our second result compares mean investment levels across different costs of investment.

**Result 2.** *(a) In both information conditions, mean investment levels are decreasing in the costs of investment. (b) When the costs of investment are high, mean investment levels are very close to the standard predictions.*

Result 2 follows from comparing the mean investment levels in the different rows of Table 3.

<sup>13</sup>When we pool the data from the first and second halves, the difference between the unobservable and observable case is also significant;  $p = .0023$  at the individual level and  $p = .0247$  at the group level.

Table 4:  $p$ -values of comparative statics tests by information condition

	first: rounds 1-18		second: rounds 19-36	
	unobs.	obser.	unobs.	obser.
$C = 20$ vs. $C = 40$	.0054	.0000	.0003	.0000
	.0273	.0277	.0350	.0273
$C = 20$ vs. $C = 60$	.0000	.0000	.0001	.0000
	.0277	.0273	.0277	.0277
$C = 40$ vs. $C = 60$	.0006	.0000	.1298	.0013
	.0277	.0277	.4593	.0345

*Remark:* The reported  $p$ -values correspond to a Wilcoxon signed-rank test. For each comparison the upper (lower)  $p$ -value is based on 60 (6) matched pairs of individual (group) mean investment levels.

Under unobservable investment, mean investment levels fall from around 68% to around 36% when  $C$  increases from 20 to 60. With observable investment, mean investment levels fall from around 67% to around 10%. For low costs of investment the mean investment levels are well above the predicted levels of  $38\frac{1}{2}\%$  and 0% respectively. But for  $C = 60$  mean investment levels are fairly close to these standard predictions. This is especially true during the second half of the experiment. Table 4 reports the relevant  $p$ -values. Because comparisons are on a within-subjects basis, we make use of the Wilcoxon signed-rank test for matched pairs. In the observable investment case we observe 5 (out of 6) significant differences. In the unobservable investment case all six comparisons yield significant differences. Hence the negative relationship between investment levels and investment costs appears to be robust.

Results 1(a) and 2(b) are in line with standard equilibrium predictions, Results 1(b) and 2(a) are not. The self-interest model namely predicts that for both information conditions the propensity to invest is independent of  $C$ . Social preferences provide an explanation. As discussed in Section 3 the scope for (intention-based) reciprocity decreases with  $C$ . Therefore, when  $C$  increases, buyers should be less willing to invest. This is exactly what we observe. Result 2(b) demonstrates that with  $C$  large enough, the impact of reciprocity is likely to be weak and the predictions of standard theory and social preferences theories will coincide. Overall we conclude

that unobservability of the specific investment made does boost investment incentives. But, it only does so when alternative (social preferences) motivations do not provide strong enough incentives to invest.

## 5.2 Pricing behavior

Although our main interest lies in buyers' investment decisions, to understand these we have to analyze sellers' price demands. In the observable investment case the seller can condition his price on the investment level observed. Here we thus consider the contingencies of no-investment and investment separately. Figures 2 and 3 depict the frequency distributions of price demands by treatment. In these figures separate demand decisions rather than the (individual or group) mean demands are the units of observation. These demands are bunched into intervals of 10 experimental points; demands that are not divisible by 10 are rounded upwards to the nearest multiple of 10. We also group the data from the first 18 and the last 18 rounds, because the shapes of the distributions are very similar over time.

[ Insert Figures 2 and 3 about here ]

First consider the observable investment case. When no investment is made almost always  $P = 50$  is chosen. For all values of  $C$  the frequency of exactly this demand is over 90%. These demands are fully in line with standard predictions, but are much higher than those typically observed in dictator games (cf. Camerer 2003). The latter points at the importance of intention-based negative reciprocity.<sup>14</sup> If subjects would only care about the final distribution of payoffs, we would predict no differences between the situation in which the small pie is exogenously fixed (like in a dictator game) and the one where it is endogenously chosen (as in our game). Our finding that there is a difference is in line with previous experimental results that intentions do matter, see e.g. Falk et al. (2000, 2003).

When the buyer invests the price demands are more dispersed. For all cost levels there is a large peak at  $P = 130$ , with a minimum mass of 39% when  $C = 20$ . For the higher cost levels

---

<sup>14</sup>The reciprocity model of Subsection 3.3 predicts that  $P_0^* = V$ , see Proposition 1 in Appendix A.1.

Table 5: Mean demands in observable case and tests for equality

	predictions		first: rounds 1-18		second: rounds 19-36	
	$I = 0$	$I = 1$	$I = 0$	$I = 1$	$I = 0$	$I = 1$
$C = 20$	50	130	48.66	101.44	49.12	114.65
$C = 40$	50	130	49.11	107.53	48.90	107.94
$C = 60$	50	130	48.40	94	49.85	110
$C = 20$ vs. $C = 40$			.2701 (30)	.4059 (51)	.1765 (50)	.1115 (24)
			1.000 (6)	.1730 (6)	.6002 (6)	.8927 (5)
$C = 20$ vs. $C = 60$			.2288 (33)	.3923 (18)	.2376 (50)	.1322 (12)
			.7532 (6)	.2249 (5)	.2809 (6)	<i>n.a.</i>
$C = 40$ vs. $C = 60$			.1300 (50)	1.000 (14)	.0770 (60)	.0861 (7)
			.2489 (6)	.0431 (5)	.0277 (6)	<i>n.a.</i>

*Remark:* The columns labeled ‘predictions’ report the standard equilibrium predictions.  $p$ -values correspond to a Wilcoxon signed-rank test. For each comparison the upper (lower)  $p$ -value is based on individual (group) level data. Within parentheses appear the number of observations (individual or group means) on which the test is based; these vary because (groups of) sellers may never be confronted with a particular investment choice ( $I = 0$  or  $I = 1$ ) in a treatment. *n.a.* indicates that no sensible test statistic is available, because there are too few observations.

the mass equals around 53%. In all three cases there is also a second smaller peak. This peak is at  $P = 60/70$  when  $C = 60$ , at  $P = 80/90$  when  $C = 40$  and at  $P = 100$  when  $C = 20$ . Here the frequencies are around 25% overall. Note that these second peaks roughly occur at demands  $130 - C - \epsilon$ , allowing the buyer to make a small return of  $\epsilon \leq 10$  on investment. Positive reciprocity provides an explanation for this (outcome-oriented social preferences that assume that subjects care about a ‘fair’ distribution of payoffs do as well).

In the unobservable investment case subjects typically choose between  $P = 50$  and  $P = 130$ , see Figure 3. The frequency with which the low demand is chosen increases with the costs of investment: 29% when  $C = 20$ , 49% for  $C = 40$  and 68% in case  $C = 60$ . These percentages are well in line with the ones of 25%, 50% and 75% predicted by standard theory. Note, however, that these predicted percentages belong to equilibria with an investment rate equal to  $38\frac{1}{2}\%$ , independent of the costs of investment. But from Result 2 we already know that the actual investment rate decreases with  $C$ . If sellers reasonably guess that buyers are less likely to invest when  $C$  is high, or figure this out after a couple of rounds, they have an incentive to play safe by choosing  $P = 50$  more often. This could also explain the observed pattern in Figure 3. More generally, because sellers know the costs of investment  $C$  and thus might be able to infer whether buyers are likely to make investments, there may be less than ‘true’ unobservability in practice.<sup>15</sup>

Apart from the above considerations based on self-interest, Figure 3 also provides clear indications for alternative motivations. Demands between 50 and 130 can be considered fair / reciprocal. We find that the number of these demands is modest, but decreases with  $C$  as predicted: 22% when  $C = 20$ , 10% when  $C = 40$  and 5% when  $C = 60$ .<sup>16</sup>

The upper parts of Table 5 and 6 present the mean demands in the various treatments, together with the (expected) price predicted by standard theory. The lower parts present the  $p$ -values of signed-rank tests that compare the different costs situations. For the observable

---

<sup>15</sup>Clearly, when sellers would not know the costs of investment  $C$  such inferences would not be possible. But then also the equilibrium predictions of Section 3 would change. In particular, if the value of  $C$  is private information to the buyer as well, sellers can not make their mixing probability  $p^* = \Pr(P = V)$  dependent on the *actual* costs of investment  $C$ .

<sup>16</sup>Reciprocal/fair behavior of the seller can take the form of a demand between 50 and 130 or, alternatively, a mixing strategy between 50 and 130 with a lower probability of the 130-demand than standard theory predicts.

Table 6: Mean demands in unobservable case and tests for equality

	predictions	first: rounds 1-18	second: rounds 19-36
$C = 20$	110	92.92	102.79
$C = 40$	90	80.00	85.41
$C = 60$	70	67.58	68.43
$C = 20$ vs. $C = 40$		.0019	.0000
		.1159	.0277
$C = 20$ vs. $C = 60$		.0000	.0000
		.0277	.0277
$C = 40$ vs. $C = 60$		.0057	.0000
		.1159	.0277

*Remark:* The column labeled ‘predictions’ reports the standard equilibrium predictions.  $p$ -values belong to a signed-rank test. For each comparison the upper (lower)  $p$ -value is based on 60 (6) matched pairs of individual (group) mean demands.

investment case the mean demand is largely independent of  $C$ . Although after investment the high demand of  $P = 130$  is chosen with a higher probability when  $C$  is high, the second peak occurs at  $130 - C$  which is lower in case  $C$  is high (cf. Figure 2). Our data suggests that these two effects cancel out. In the unobservable investment case mean demands are significantly decreasing in  $C$ . Moreover, actual average demands are somewhat below the expected demand of  $P = 130 - C$  predicted by standard theory.

The findings of this subsection are summarized in Result 3.

**Result 3.** (a) *When  $I = 0$  is observed, sellers almost always demand  $P = 50$ . In case  $I = 1$  is observed, sellers demand either  $P = 130$  or  $P = 130 - C - \epsilon$  (with  $\epsilon \leq 10$ ). The mean demand does not vary with  $C$ .* (b) *When the investment decision is unobserved, sellers typically demand either  $P = 50$  or  $P = 130$ . For higher cost levels the distribution shifts towards  $P = 50$ . Mean demands are decreasing in  $C$ .*

### 5.3 Efficiency

We next take a look at efficiency. In the unobservable investment condition there are two types of inefficiencies. First, the buyer may decide not to invest, leading to lower gains from trade. Second, the seller may demand too much, inducing no trade at all. The latter cannot occur in the observable investment case, because there by design the seller can never demand more than the actual pie. Standard theory predicts a lower investment inefficiency and a higher trade inefficiency under unobservable investment than under observable investment. Overall, however, these inefficiencies are predicted to cancel out. Compared to the standard predictions, reciprocity motivations lead to lower (investment) inefficiencies under observable investment. In the unobservable investment case the effect of reciprocity is ambiguous, because there multiple equilibria exist. When we focus on the positive reciprocity equilibrium in which the buyer invests for sure (cf. Table 2), overall efficiency improves to the same extent as under observable investment. The negative reciprocity equilibrium, however, lowers efficiency as compared to the selfish benchmark. No clear cut implications of social preferences can therefore be derived. If anything, we expect that (overall) inefficiencies are somewhat larger under unobservable investment than under observable investment. Our final result relates to this.

**Result 4.** *Investment inefficiency is weakly larger under observable investment, while trade inefficiency is always larger under unobservable investment. When subjects have gained experience overall inefficiencies are not significantly different from each other.*

Result 4 follows from comparing the various inefficiencies in the different columns of Table 7. In the observable investment case trade inefficiency is zero by design. Investment and overall inefficiencies thus coincide and are therefore reported in a single column labelled ‘inv/overall’. Comparing investment inefficiencies, we observe that these are typically larger under observable investment than under unobservable investment. To illustrate, when  $C = 60$  in the second half of the experiment (next to last row), investment inefficiencies equal 12.33 under unobservable investment and 18.44 under observable investment. Similarly, average trade inefficiencies under unobservable investment are always strictly positive (see the fourth column in Table 7), and thus always higher than the zero trade inefficiency under observable investment. However, overall

Table 7: Inefficiencies by treatment and tests for equality

	rounds	unobservable			observable	$p$ -values	
		invest	trade	overall	inv/overall	invest	overall
$C = 20$	1-18	17.33	9.17	26.5	13.67	.4657	.0534
	19-36	21.33	11.39	32.72	25.67	.2207	.4688
		[36.92]	[23.08]	[60]	[60]		
$C = 40$	1-18	17.56	10	27.56	21.56	.1481	.2615
	19-36	21.78	13.61	35.39	32.44	.0192	.4201
		[24.62]	[15.38]	[40]	[40]		
$C = 60$	1-18	13.33	9.44	22.77	17.56	.0215	.0054
	19-36	12.33	7.77	20.11	18.44	.0031	.6242
		[12.31]	[7.69]	[20]	[20]		

*Remark:* Predicted inefficiencies based on standard theory appear in square brackets. Under observable investment trade inefficiency is zero by design and investment and overall inefficiency coincide.  $p$ -values refer to Mann-whitney ranksum tests performed on group level data (with  $m = n = 6$ ).

efficiency losses between the two information conditions look very similar (5th and 6th column in Table 7), especially in the second half of the experiment.

The above observations are supported by statistical (ranksum) tests. Because efficiency losses can only be calculated for a buyer-seller pair, tests cannot be based on individual means. We therefore only consider tests performed at the aggregate matching group level. The relevant  $p$ -values are reported in Table 7. The second to last column reports the test statistics of comparing investment inefficiencies. The results reiterate our earlier conclusions about mean investment levels. The last column concerns the comparison of overall inefficiencies. Although these are typically smaller when the investment decision is observable, the differences are not statistically significant, with the exception of  $C = 60$  in the first 18 rounds. Once subjects have gained experience overall inefficiencies do not vary with the information condition. This concurs with the predictions of standard theory.

## 5.4 Robustness

In the previous subsections we reported the results for the first and second half of the experiment separately. We did this because comparing the two halves some learning effects can be detected (cf. the web-appendix). Learning can have two causes. First, some subjects may understand the subtleties of the game only after a few rounds. Second, subjects may adapt their beliefs about the population characteristics. For example, a buyer who expects sellers to act reciprocal, may be disappointed after some rounds and change his/her investment decisions accordingly.

In the unobservable investment case almost no learning effects can be detected. Although for  $C = 20$  and  $C = 40$  investment levels are lower in the second half of the experiment (cf. Table 3), differences are insignificant for all cost levels. The increase in sellers' demands observed in Table 6 is significant only for  $C = 20$ . However, also here mean demands stay below the standard prediction of  $130 - C$ . This indicates that social preferences remain to play a role.

Learning is more prominent under observable investment. For low and intermediate cost levels investment rates decrease significantly over time (when  $C = 60$  the difference is insignificant at the 5% level). Changes in sellers' demands provide a partial explanation for this. Note that sellers practically always ask the whole pie of 50 points if no investment is made (cf. Table 4). However, in case the buyer invests, sellers demand a larger part of the pie in the second half of the experiment. Differences are significant only when  $C = 20$ . These changes in demands are such that in the first part of the experiment buyers make a modest profit on a low cost investment. In the second half this turns into a small loss. As a consequence, selfish buyers prefer not to invest in the second part of the experiment, in line with the significant decrease in investment levels we observe. Still, the investment rate equals 57% when  $C = 20$  in the second half of the experiment. Social preferences suggest an explanation. On average buyers lose a few points (5.53) when they invest, but the increase in sellers' earnings is substantial (65.53 points). So, already a relatively weak concern for the payoffs of the other (or efficiency in general) rationalizes investment.

A potential consequence of learning is that the effect of private information on investment incentives become more important over time. We indeed observe that the difference in mean

investment levels between the two information conditions (i.e.  $q_{un} - q_{obs}$ ) is always higher in the second half of the experiment. In order to formally test whether these differences are significant, we compare the change in the unobservable investment rate over time (second half versus first half) with the change in the observable investment rate over time. Ranksum tests reveal a significant difference only for  $C = 40$  if tests are performed at the individual level ( $p = .037$ ). Group level tests yield insignificant differences for all cost levels though. Overall we therefore conclude that learning effects do not affect our main findings concerning investment behavior.

Besides past experience, subjects may also have been influenced by their education. As noted in Section 4, economics students are overrepresented in our sample (60%). This could potentially bias our results, because as Frank et al. (1993) have observed economics students tend to be less cooperative than non-economics students are. If economists are indeed less guided by social preferences, the scope for such motivational factors would actually be (even) larger in practice than our results suggest. To explore whether such a bias exists, we compare the behavior of economics students with non-economists. Economics students indeed tend to invest less. With just a single exception this is true for all cost levels, both halves of the experiment and for both the observable and unobservable investment case. None of the differences are, however, significant at the 5%-level (cf. the web-appendix). Moreover, all the comparative statics results with regard to investment levels (cf. Results 1 and 2) continue to hold for both types of subjects.

Demands are also very similar between economics students and other subjects. The single significant difference occurs in the first half of the observable investment case when  $C = 20$  and the buyer did not invest. Sellers that study economics then demand 50 on average, non-economists slightly less (46.88). Taken together these results indicate that our results are not biased due to the fact that we have a large fraction of economics students in our subject pool.

## 6 Conclusion

This paper addresses the question whether making specific investments unobservable boosts investment incentives, as predicted by Tirole (1986) and Gul (2001) among others. Our ex-

perimental findings indicate that this will be the case only when there is insufficient scope for social preferences, i.e. when the costs of investment are relatively high compared to the return on investment. In case the costs of investment are relatively low, social preferences are at work and these have a much more substantial impact on investment incentives when the investment decision is observable than when it is not. As a result, investment levels under the two information conditions are equal when the costs of investment are relatively low. Private information then does not improve investment incentives.

Overall, our results tentatively suggest that private information may partially crowd out the positive investment incentive effect of fairness and reciprocity motivations. Clearly our experiment just provides a first step and a number of interesting questions remain. For instance, the reduced form price-setting stage that we employ is not realistic. In reality the buyer at least has the opportunity to accept or reject the seller's price demand. Standard predictions remain unchanged for such a setup, but the predictions under social preferences change and become much more involved. In particular, the buyer may then want to reciprocate with her acceptance/rejection decision. Anticipating this, the seller may change his demand behavior, which in turn affects investment incentives. These additional strategic issues and motives that come into play will make it more difficult to interpret observed behavior. Now that we have established that in the simplest possible setup unobservability may indeed affect investment incentives (only) when the scope for social preferences is limited, future experiments can build on this and investigate whether this result generalizes to more natural bargaining settings.

## References

- Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior* 10, 122–142.
- Camerer, C. (2003). *Behavioral game theory. Experiments in strategic interaction*. Princeton University Press.
- Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117, 817–869.
- Cremer, J. (1995). Arm’s length relationships. *Quarterly Journal of Economics* 110, 275–295.
- Dufwenberg, M. and G. Kirchsteiger (2000). Reciprocity and wage undercutting. *European Economic Review* 44, 1069–1078.
- Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior* 47, 268–298.
- Ellingsen, T. and M. Johannesson (2004a). Is there a holdup problem? *Scandinavian Journal of Economics* 106, 475–494.
- Ellingsen, T. and M. Johannesson (2004b). Promises, threats, and fairness. *Economic Journal* 114, 397–420.
- Ellingsen, T. and M. Johannesson (2005). Sunk costs and fairness in incomplete information bargaining. *Games and Economic Behavior* 50, 155–177.
- Falk, A., E. Fehr, and U. Fischbacher (2000). Testing theories of fairness - Intentions matter. working paper.
- Falk, A., E. Fehr, and U. Fischbacher (2003). On the nature of fair behavior. *Economic Inquiry* 41, 20–26.
- Fehr, E. and J. List (2004). The hidden costs and returns of incentives - trust and trustworthiness among ceos. *Journal of the European Economic Association* 2, 743–771.
- Fehr, E. and K. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817–868.

- Frank, R., T. Gilovich, and D. Regan (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives* 7, 159–171.
- Gantner, A., W. Güth, and M. Königstein (2001). Equitable choices in bargaining games with advance production. *Journal of Economic Behavior and Organization* 46, 209–225.
- Gonzalez, P. (2004). Investment and screening under asymmetric endogenous information. *Rand Journal of Economics* 35, 502–519.
- Gul, F. (2001). Unobservable investment and the hold-up problem. *Econometrica* 69, 343–376.
- Hackett, S. (1993). Incomplete contracting: A laboratory experimental analysis. *Economic Inquiry* 31, 274–297.
- Hackett, S. (1994). Is relational exchange possible in the absence of reputations and repeated contract? *Journal of Law, Economics and Organization* 10, 360–389.
- Klein, B., R. Crawford, and A. Alchian (1978). Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics* 21, 279–326.
- Königstein, M. (2000). *Equity, efficiency and evolutionary stability in bargaining games with joint production*. Berlin Heidelberg: Springer Verlag.
- Konrad, K. (2001). Privacy and time-consistent optimal labor income taxation. *Journal of Public Economics* 79, 503–519.
- Lau, S. (2002). Information and bargaining in the hold-up problem. Working paper, Washington University in St. Louis.
- Leuven, E., H. Oosterbeek, R. Sloof, and C. van Klaveren (2005). Worker reciprocity and employer investment in training. *Economica* 72, 137–149.
- Oosterbeek, H., R. Sloof, and J. Sonnemans (2006). Promotion rules and skill acquisition: An experimental study. Forthcoming in: *Economica*.
- Oosterbeek, H., J. Sonnemans, and S. van Velzen (2003). The need for marriage contracts: An experimental study. *Journal of Population Economics* 16, 431–453.

- Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review* 83, 1280–1302.
- Riordan, M. (1990). What is vertical integration? In M. Aoki, B. Gustafsson, and O. Williamson (Eds.), *The firm as a nexus of treaties*, Chapter 5, pp. 94–111. London: European Sage.
- Rogerson, W. (1992). Contractual solutions to the hold-up problem. *Review of Economic Studies* 59, 777–793.
- Schmidt, K. (1996). The costs and benefits of privatization: An incomplete contracts approach. *Journal of Law, Economics and Organization* 12, 1–24.
- Sloof, R. (2006). Price-setting power versus private information: An experimental evaluation of their impact on holdup. Working Paper (May 2006).
- Sloof, R., H. Oosterbeek, and J. Sonnemans (2005). Does making specific investments unobservable boost investment incentives? Working paper, Version September 2005.
- Sloof, R., J. Sonnemans, and H. Oosterbeek (2004). Specific investments, holdup, and the outside option principle. *European Economic Review* 48, 1399–1410.
- Sonnemans, J., H. Oosterbeek, and R. Sloof (2001). On the relation between asset ownership and specific investments. *Economic Journal* 111, 791–820.
- Tirole, J. (1986). Procurement and renegotiation. *Journal of Political Economy* 94, 235–259.
- Williamson, O. (1985). *The economic institutions of free capitalism*. Free Press.

## Appendix A: Formal derivation of sequential reciprocity equilibria

In this appendix we formally state and derive the predictions based on social preferences as discussed in Subsection 3.3. Following Dufwenberg and Kirchsteiger (2004) we assume that the seller's utility is given by:

$$u_S = \pi_S + Y_S \cdot \kappa \cdot \lambda, \quad (A1)$$

with  $\pi_S$  reflecting the seller's monetary payoffs and  $Y_S \cdot \kappa \cdot \lambda$  his reciprocity payoffs. The seller's sensitivity to reciprocity is captured by parameter  $Y_S \geq 0$ . The seller's kindness  $\kappa$  of a particular price demand  $P$  is formally defined as the difference between what the seller thinks he actually gives to the buyer by choosing  $P$ , and the average of the maximum and the minimum monetary payoff that he believes he could give her in principle (i.e. the buyer's 'equitable' payoff). Note that this factor depends on the seller's (first order) beliefs about the buyer's actual investment choice. Factor  $\lambda$  represents the perceived kindness of the buyer. It equals the difference between what the seller believes the buyer believes she gives to the seller, and the average of the maximum and the minimum monetary payoff that the seller believes the buyer believes she could give to the seller in principle. To calculate  $\lambda$  we thus need the seller's second order beliefs about what the buyer believes about his pricing strategy. Using these definitions the exact specifications of  $\kappa$  and  $\lambda$  are formally derived below in the proofs of the various lemmas and propositions.

For brevity we focus on the situation considered in the experiment in which the investment more than doubles the ex post surplus, i.e.  $V < W$ . Theorem 1 below, on which Table 2 in the main text is based, characterizes all sequential reciprocity equilibria (SRE) for this case.

**Theorem 1** Suppose the seller's preferences are given by (A1) and let  $V < W$ . In the observable investment case there exists a unique SRE. In the unobservable investment case the SRE is unique when  $Y_S < \frac{2}{V+W-C}$  whereas for  $Y_S > \frac{2}{V+W-C}$  multiple SRE exist side by side. Equilibrium investment behavior is characterized by:

**(weak)** if  $Y_S < \frac{2}{V+W-C}$ , then  $1 > \frac{V}{V+W} \geq q_{un}^* > q_{obs}^* = 0$ ;

**(medium-low)** if  $\frac{2}{V+W-C} < Y_S < \frac{2}{W-C} \cdot \frac{W}{V+W}$ , then  $1 \geq q_{un}^* > q_{obs}^* = 0$ . In the unobservable investment case there exists one SRE with  $0 < q_{un}^* < \frac{V}{V+W}$  and another one with  $q_{un}^* = 1$ ;

**(medium-high)** if  $\frac{2}{W-C} \cdot \frac{W}{V+W} < Y_S < \frac{2}{W-C}$ , then  $1 \geq q_{un}^* > q_{obs}^* = 0$ . In the unobservable investment case there exists one SRE with  $0 < q_{un}^* < \frac{V}{V+W}$ , one with  $\frac{1}{2} < q_{un}^* < 1$  and another one with  $q_{un}^* = 1$ ;

**(strong)** if  $Y_S > \frac{2}{W-C}$ , then  $1 = q_{obs}^* \geq q_{un}^* > 0$ . In the unobservable investment case there exists one SRE with  $0 < q_{un}^* < \frac{V}{V+W}$ , one with  $\frac{1}{2} < q_{un}^* < 1$  and another one with  $q_{un}^* = 1$ .

Note that in Table 2 in the main text we have joined the ‘medium-low’ and ‘medium-high’ categories into a single ‘medium’ category. As explained there (cf. footnote 5), we have omitted the SRE with  $\frac{1}{2} < q_{un}^* < 1$  from the general discussion, because it does not affect our main qualitative predictions. Ignoring this SRE, the two medium cases in Theorem 1 coincide. The dividing conditions on  $Y_S$  have been rewritten into conditions on  $C$  in Table 2.

In order to prove Theorem 1 we derive the SRE for the two investment conditions in two separate subsections. The theorem then directly follows from Propositions 1 and 2 below.

### A.1 Observable investment

The following additional notation is used. The seller’s pricing strategy is denoted  $(\Delta_0, \Delta_1)$ . It consists of two probability distributions over  $[0, V+W]$ , one for each investment level ( $I = 0$  and  $I = 1$ ) separately. In case the seller uses a pure pricing strategy, we use the more convenient notation  $(P_0, P_1)$ . Expected price demands are denoted  $(\overline{P}_0, \overline{P}_1)$ . Turning to beliefs,  $b$  gives the (first order) belief of the seller about the buyer’s investment strategy  $q_{obs} \equiv \Pr(I = 1)$ .  $c$  reflects the seller’s (second order) belief about the buyer’s belief about his pricing strategy  $(\Delta_0, \Delta_1)$ . Because effectively only beliefs about *expected* prices are important, we use  $(c_0, c_1)$  to reflect the second order beliefs about  $(\overline{P}_0, \overline{P}_1)$ . As explained above, the first and second order beliefs determine the factors  $\kappa$  and  $\lambda$  in the seller’s utility function (A1). In an SRE beliefs are necessarily correct:  $b = q_{obs}^*$  and  $c = (\Delta_0^*, \Delta_1^*)$ . Proposition 1 characterizes all SRE.

**Proposition 1** (Observable investment) Suppose the seller’s preferences are given by (A1).

Then the unique SRE is characterized by:

- (a)  $Y_S < \frac{2}{W-C} : q_{obs}^* = 0, P_0^* = V$  and  $P_1^* = \min\{V + \frac{2}{Y_S}, V + W\}$ ;

(b)  $Y_S > \frac{2}{W-C} : q_{obs}^* = 1, P_0^* = V$  and  $P_1^* = V + \frac{2}{Y_S}$ .

*Proof of Proposition 1.* We first derive the kindness factor  $\kappa$  in (A1) for any possible price demand  $P \in [0, V + W]$ . When the buyer chooses  $I = 1$  the seller can give her at least 0 and at most  $V + W$  (the buyer's investment costs are sunk at this stage and thus do not affect the kindness term). The equitable payoff for the buyer thus equals  $\frac{1}{2} \cdot [0 + (V + W)] = \frac{V+W}{2}$ . This implies  $\kappa(P, I = 1) = (V + W - P) - \frac{V+W}{2} = (\frac{V+W}{2} - P)$ . Similarly,  $\kappa(P, I = 0) = \max\{V - P, 0\} - \frac{V}{2}$ . For  $I = 0$  the reciprocity payoffs then equal  $-Y_S \cdot \frac{V}{2} \cdot \lambda$  whenever  $P \geq V$  and are thus independent of  $P$  (recall that factor  $\lambda$  only depends on beliefs). On the basis of monetary payoffs  $\pi_S$  the seller strictly prefers  $P = V$  after  $I = 0$  above any  $P > V$ , so the latter are never chosen in a SRE. Hence necessarily  $\overline{P}_0^* \leq V$  and thus also  $c_0 \leq V$ .

Next we turn to perceived kindness  $\lambda$ . The seller's belief about how much the buyer intends to give him by choosing  $I = 1$  equals  $c_1$ . For  $I = 0$  this is  $c_0$ . Hence  $\lambda(I = 1, c_0, c_1) = c_1 - \frac{1}{2} \cdot [c_0 + c_1] = \frac{1}{2}(c_1 - c_0)$  and  $\lambda(I = 0, c_0, c_1) = \frac{1}{2}(c_0 - c_1)$ . In a SRE beliefs are correct, so that  $c_0 = \overline{P}_0^*$  and  $c_1 = \overline{P}_1^*$ . Suppose  $\overline{P}_0^* > \overline{P}_1^*$ . In that case  $\lambda(I = 1, c_0, c_1) < 0$ . The seller's reciprocity payoffs  $Y_S \cdot (\frac{V+W}{2} - P) \cdot \lambda$  after  $I = 1$  are then increasing in  $P$  (for  $Y_S > 0$ ), just like his monetary payoffs  $\pi_S$  are. Hence the seller chooses  $P = V + W$  for sure. This implies  $\overline{P}_1^* = V + W$  and contradicts the supposition that  $\overline{P}_0^* > \overline{P}_1^*$ . Therefore, necessarily  $\overline{P}_0^* \leq \overline{P}_1^*$ .

Seller's overall utility after  $I = 0$  equals  $u_S = P + Y_S \cdot (\frac{V}{2} - P) \cdot \frac{1}{2}(c_0 - c_1)$  for  $P \leq V$  (and 0 otherwise). From  $\overline{P}_0^* \leq \overline{P}_1^*$  it follows that  $\frac{1}{2}(c_0 - c_1) \leq 0$ , so  $u_S$  is strictly increasing in  $P$  (for  $P \leq V$ ). The equilibrium price after no investment therefore equals  $P_0^* = V$ . Given this price, seller's utility after  $I = 1$  equals  $u_S = P + Y_S \cdot (\frac{V+W}{2} - P) \cdot \frac{1}{2}(c_1 - V)$  for  $P \in [0, V + W]$ . We obtain  $\frac{\partial u_S}{\partial P} = 1 - \frac{Y_S}{2}(c_1 - V)$ . For  $c_1 < V + \frac{2}{Y_S}$  this is strictly positive, hence  $P < \min\{V + \frac{2}{Y_S}, V + W\}$  cannot occur. Similarly, for  $c_1 > V + \frac{2}{Y_S}$  the derivative is negative, so  $P > \min\{V + \frac{2}{Y_S}, V + W\}$  cannot occur. Necessarily then  $P_1^* = \min\{V + \frac{2}{Y_S}, V + W\}$  as equilibrium price.

Given correct beliefs about the seller's equilibrium pricing strategy, the buyer's monetary payoffs of investment strategy  $q_{obs}$  equal  $\pi_B = q_{obs} \cdot (W - \min\{\frac{2}{Y_S}, W\} - C)$ . It immediately follows that  $q_{obs}^* = 1$  [= 0] whenever  $Y_S > [<] \frac{2}{W-C}$ . In the degenerate case  $Y_S = \frac{2}{W-C}$  any  $q_{obs}^* \in [0, 1]$  is possible. Such knife-edge cases are ignored here. *QED*

## A.2 Unobservable investment

The seller's pricing strategy is now given by an unconditional probability distribution  $\Delta$  over  $[0, V + W]$ . Hence second order beliefs  $c$  are now defined with respect to this strategy. Similarly, the seller's first order beliefs  $b$  refer to  $q_{un} \equiv \Pr(I = 1)$ . Before characterizing all SRE we first present three lemmas that facilitate the equilibrium analysis.

**Lemma 1** In any SRE necessarily  $\lambda \leq \frac{1}{Y_S}$ . Moreover, it holds that:

$$(a) \quad \lambda = \frac{1}{Y_S} \iff q_{un}^* = 1;$$

$$(b) \quad \lambda < \frac{1}{Y_S} \iff 0 < q_{un}^* < 1.$$

*Proof of Lemma 1.* We first derive the kindness factor  $\kappa$  in (A1). Given his beliefs  $b$  about  $q_{un}$ , the seller thinks he can give the buyer at least  $b \cdot (-C)$  by choosing  $P = V + W$  and at most  $b \cdot (V + W - C) + (1 - b) \cdot V$  by choosing  $P = 0$ . Hence the buyer's equitable payoff equals  $\frac{V}{2} + b \cdot (\frac{W}{2} - C)$ . By choosing  $P \leq V$  the seller intends to give the buyer  $b \cdot (V + W - C - P) + (1 - b) \cdot (V - P)$ . The kindness of such a choice therefore equals  $\kappa(P \leq V, b) = \frac{V}{2} - P + b \cdot \frac{W}{2}$ . Similarly, by choosing  $P > V$  the seller intends to give the buyer a payoff of  $b \cdot (V + W - C - P)$ . In that case  $\kappa(P > V, b) = (b - \frac{1}{2}) \cdot V - b \cdot P + b \cdot \frac{W}{2}$ . Seller's expected utility thus equals:

$$\begin{aligned} u_S &= P + Y_S \cdot \lambda \cdot \left[ \frac{V}{2} - P + b \cdot \frac{W}{2} \right] \quad \text{when } P \leq V \\ &= b \cdot P + Y_S \cdot \lambda \cdot \left[ \left( b - \frac{1}{2} \right) \cdot V - b \cdot P + b \cdot \frac{W}{2} \right] \quad \text{for } P > V. \end{aligned} \tag{A2}$$

First suppose  $\lambda > \frac{1}{Y_S}$ . Then from the above expression  $\frac{\partial u_S}{\partial P} < 0$  and the seller strictly prefers  $P = 0$ . But when  $P = 0$  for sure, the buyer cannot be kind or unkind to the seller with her investment decision. This implies  $\lambda = 0$ , a contradiction. Hence necessarily  $\lambda \leq \frac{1}{Y_S}$ .

Next let  $\lambda = \frac{1}{Y_S}$ . Then  $u_S = \frac{V}{2} + b \cdot \frac{W}{2}$  when  $P \leq V$  and  $u_S = (b - \frac{1}{2}) \cdot V + b \cdot \frac{W}{2}$  for  $P > V$ . Suppose  $b < 1$ . The seller then always prefers  $P \leq V$  over  $P > V$ . Knowing that  $P \leq V$ , the selfish buyer chooses  $q_{un} = 1$ . The latter contradicts  $b < 1$  under correct equilibrium beliefs  $b = q_{un}^*$ . Hence necessarily  $b = q_{un}^* = 1$ . In sum, we have the implication  $\lambda = \frac{1}{Y_S} \implies q_{un}^* = 1$ .

Finally, consider the case where  $\lambda < \frac{1}{Y_S}$ . First suppose  $q_{un}^* = 1$ . Then under correct beliefs  $b = q_{un}^*$  we have  $u_S = P + Y_S \cdot \lambda \cdot \left[ \left( \frac{V}{2} - P \right) + \frac{W}{2} \right]$  for all  $P$ . From this we obtain  $\frac{\partial u_S}{\partial P} = 1 - Y_S \cdot \lambda > 0$ . The seller thus wants to choose  $P = V + W$  for sure. But for this price the buyer strictly prefers  $q_{un} = 0$ , contradicting  $q_{un}^* = 1$ . A similar contradiction follows from the supposition that  $q_{un}^* = 0$ . We thus obtain the implication  $\lambda < \frac{1}{Y_S} \implies 0 < q_{un}^* < 1$ .

The two derived implications, together with  $\lambda \leq \frac{1}{Y_S}$  necessarily, immediately yield the implications in the opposite direction. *QED*

**Lemma 2** In any SRE with  $q_{un}^* = 1$  the seller is indifferent between all  $P \in [0, V + W]$ . Let

$P_l \equiv E[P \mid P \leq V]$ ,  $P_h \equiv E[P \mid P > V]$  and  $\tilde{p} \equiv \Pr(P \leq V)$  for the seller's equilibrium strategy  $\Delta^*$ . A SRE with  $q_{un}^* = 1$  necessarily requires  $(1 - \tilde{p}) \cdot P_h = \frac{2}{Y_S}$  and exists iff:

$$Y_S \geq \max \left\{ \frac{2}{P_h}, \frac{2}{W - C} \cdot \frac{P_h - V}{P_h} \right\}. \quad (\text{A3})$$

*Proof of Lemma 2.* When  $q_{un}^* = 1$  we have from Lemma 1 that  $\lambda = \frac{1}{Y_S}$ , and thus  $u_S = \frac{V+W}{2}$  under correct beliefs  $b = q_{un}^* = 1$  (cf. expression (A2)). Seller's utility is thus independent of his pricing strategy and any price is a best response. The buyer's expected utility equals:

$$u_B = \pi_B = \tilde{p} \cdot (V - P_l) + q_{un} \cdot [(1 - \tilde{p})V + W - (1 - \tilde{p})P_h - C] \quad (\text{A4})$$

For  $q_{un}^* = 1$  to be an equilibrium we need  $\frac{\partial u_B}{\partial q_{un}} \geq 0$ , i.e.  $(1 - \tilde{p})V + W - (1 - \tilde{p})P_h - C \geq 0$ . Moreover, the perception  $\lambda$  about the buyer's intended kindness should be correct. By choosing  $I = 0$  the buyer (correctly) believes to give the seller a monetary payoff of  $\tilde{p} \cdot P_l$ , while for  $I = 1$  this amounts to  $\tilde{p} \cdot P_l + (1 - \tilde{p}) \cdot P_h$ . The equitable payoff for the seller is thus  $\pi_S^e = \tilde{p} \cdot P_l + \frac{1}{2}(1 - \tilde{p})P_h$ . The buyer's kindness of choosing investment strategy  $q_{un}$  then equals  $\nu(q_{un}) \equiv (1 - q_{un}) \cdot \tilde{p} \cdot P_l + q_{un} \cdot [\tilde{p} \cdot P_l + (1 - \tilde{p}) \cdot P_h] - \pi_S^e = (q_{un} - \frac{1}{2})(1 - \tilde{p})P_h$ . In equilibrium the seller's belief about  $\nu(q_{un})$  should be correct and we obtain  $\lambda = \nu(1) = \frac{1}{2}(1 - \tilde{p})P_h$ .

From Lemma 1 we know that  $q_{un}^* = 1 \iff \lambda = \frac{1}{Y_S}$ . Hence necessarily  $(1 - \tilde{p}) \cdot P_h = \frac{2}{Y_S}$ . To secure  $\tilde{p} \geq 0$  this requires  $Y_S \geq \frac{2}{P_h}$  and yields the first part of condition (A3). Substituting  $(1 - \tilde{p}) \cdot P_h = \frac{2}{Y_S}$  in the above inequality  $\frac{\partial u_B}{\partial q_{un}} \geq 0$  and rewriting gives the second part. *QED*

**Lemma 3** In any SRE with  $0 < q_{un}^* < 1$  the seller necessarily *strictly* mixes between  $P = V$  and  $P = V + W$  only. Let  $p = \Pr(P = V)$ . In equilibrium then necessarily  $p^* = \frac{C}{W}$ . Moreover,  $q_{un}^*$  is characterized by the solutions of  $q_{un}$  to the following equation:

$$h(q_{un}) \equiv q_{un} \cdot \left[ 1 - Y_S \cdot \left( q_{un} - \frac{1}{2} \right) \cdot (W - C) \right] = \frac{V}{V + W}. \quad (A5)$$

*Proof of Lemma 3.* Let  $0 < q_{un}^* < 1$ . Lemma 1 then implies that  $\lambda < \frac{1}{Y_S}$  and from (A2) we obtain  $\frac{\partial u_S}{\partial P} > 0$  for all  $P \neq V$ . The seller therefore chooses between  $P = V$  and  $P = V + W$  only. Under correct beliefs about  $p \equiv \Pr(P = V)$  the buyer's expected utility equals  $u_B = q_{un} \cdot (p \cdot W - C)$ . Now  $0 < q_{un}^* < 1$  implies that necessarily  $\frac{\partial u_B}{\partial q_{un}} = 0$  at  $q_{un}^*$ . This requires  $p^* = \frac{C}{W}$ .

From the seller's expected utility given in (A2) it follows that he is willing to mix between  $P = V$  and  $P = V + W$  only if  $V + Y_S \cdot \lambda \cdot \left[ -\frac{V}{2} + b \cdot \frac{W}{2} \right] = b \cdot (V + W) + Y_S \cdot \lambda \cdot \left[ -\frac{V}{2} - b \cdot \frac{W}{2} \right]$ . Under correct beliefs  $b = q_{un}$  this reduces to  $q_{un} \cdot (V + W) - Y_S \cdot q_{un} \cdot W \cdot \lambda = V$ . The perceived kindness of the buyer  $\lambda$  is most easily obtained from the buyer's kindness  $\nu(q_{un}) = \left( q_{un} - \frac{1}{2} \right) (1 - p) (V + W)$  as derived in the proof of Lemma 2 (here we have inserted  $\tilde{p} = p$  and  $P_h = V + W$ ). We just have to move up only level in the belief hierarchy. This yields  $\lambda = \left( b - \frac{1}{2} \right) (1 - c) (V + W)$ , where  $c$  is the seller's belief about what the buyer believes about  $p$ . Under correct beliefs  $b = q_{un}$  and  $c = p^*$  we obtain that  $\lambda = \left( q_{un} - \frac{1}{2} \right) (1 - p^*) (V + W)$ . Hence  $q_{un}^*$  is characterized by the solutions to the equation:

$$q_{un} \cdot (V + W) - Y_S \cdot q_{un} \cdot W \cdot \left[ \left( q_{un} - \frac{1}{2} \right) (1 - p^*) (V + W) \right] = V.$$

Inserting  $p^* = \frac{C}{W}$  and rewriting yields that  $q_{un}^*$  follows from the solutions to (A5). *QED*

**Proposition 2** (Unobservable investment) Suppose the seller's preferences are given by (A1) and consider the case  $V < W$ . Let  $q_l$  ( $q_h$ ) be the smallest (largest) solution to equation (A5) in Lemma 3 above. Equilibrium behavior is then characterized by:

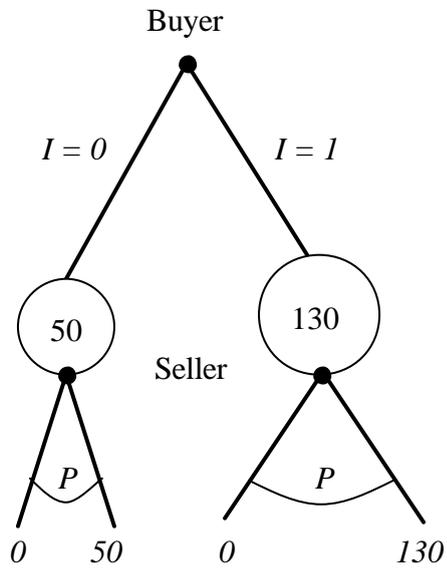
- (a)  $Y_S < \frac{2}{V+W-C}$  : there exists a unique SRE with  $0 < q_{un}^* = q_l \leq \frac{V}{V+W}$  (with a strict inequality for  $Y_S > 0$ ), and  $p^* \equiv \Pr(P = V) = \frac{C}{W}$  and  $\Pr(P = V + W) = 1 - \frac{C}{W}$ ;

- (b)  $\frac{2}{V+W-C} < Y_S < \frac{2}{W-C} \cdot \frac{W}{V+W}$  : besides the SRE described in part (a), there exists a (continuum of) SRE with  $q_{un}^* = 1$ . In the latter SRE pricing behavior necessarily satisfies  $(1 - \tilde{p}) \cdot P_h = \frac{2}{Y_S}$ , with  $P_h \equiv E[P | P > V]$  and  $\tilde{p} \equiv \Pr(P \leq V)$ ;
- (c)  $Y_S > \frac{2}{W-C} \cdot \frac{W}{V+W}$  : besides all SRE of parts (a) and (b), there exists a SRE with  $\frac{1}{2} < q_{un}^* = q_h < 1$ , and  $p^* \equiv \Pr(P = V) = \frac{C}{W}$  and  $\Pr(P = V + W) = 1 - \frac{C}{W}$ .

*Proof of Proposition 2.* From Lemma 1 we have that  $q_{un}^* = 0$  cannot occur. First, consider SRE with  $0 < q_{un}^* < 1$ . These are characterized in Lemma 3. The parabolic function  $h(q_{un})$  is concave in  $q_{un}$ , with  $h(0) = 0$  and  $h(\frac{1}{2}) = \frac{1}{2}$ . For  $V < W$  we have that  $\frac{V}{V+W} < \frac{1}{2}$ . From the intermediate value theorem it then follows that necessarily  $0 < q_l < \frac{1}{2}$  and  $q_h > \frac{1}{2}$ . Hence an equilibrium with  $q_{un}^* = q_l$  and  $p^* = \frac{C}{W}$  exists for any value of  $Y_S \geq 0$ . For  $q_{un} < \frac{1}{2}$  term  $[1 - Y_S \cdot (q_{un} - \frac{1}{2})(W - C)]$  in  $h(q_{un})$  weakly exceeds 1, and strictly so when  $Y_S > 0$ . Hence necessarily  $q_l \leq \frac{V}{V+W}$  and a strict inequality for  $Y_S > 0$ . This yields the equilibrium described in part (a). For  $q_{un} = q_h$  to be an equilibrium it is required that  $q_h \leq 1$ . Given the concavity of  $h(\cdot)$  a necessary and sufficient condition for this is that  $h(1) \leq \frac{V}{V+W}$ . Rewriting this yields that  $Y_S \geq \frac{2}{W-C} \cdot \frac{W}{V+W}$  is required. This gives part (c).

Second, consider SRE with  $q_{un}^* = 1$ . From Lemma 2 it follows that  $Y_S \geq \max\{\frac{2}{P_h}, \frac{2}{W-C} \cdot \frac{P_h - V}{P_h}\}$  is needed. The first argument in the max-term is decreasing in  $P_h$ , the second increasing. They are equal for  $P_h = V + W - C$ . Hence  $Y_S \geq \frac{2}{V+W-C}$  is the minimum requirement. A pricing strategy with  $P_h \equiv E[P | P > V] = V + W - C$ ,  $\tilde{p} \equiv \Pr(P \leq V) = 1 - \frac{2}{Y_S(V+W-C)}$  and  $P_l \equiv E[P | P \leq V]$  arbitrary then supports  $q_{un}^* = 1$ . Lemma 2 gives the general restrictions on equilibrium pricing behavior. This gives the SRE of part (b). *QED*

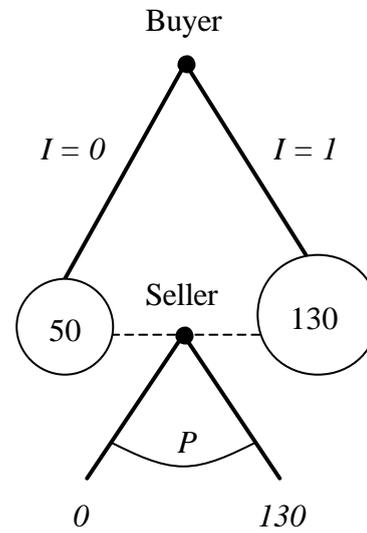
Figure 1a: Observable investment case



$$p_B = 50 + I \cdot 80 - P - I \cdot C; \quad C \in \{20, 40, 60\}$$

$$p_S = P$$

Figure 1b: Unobservable investment case



$$p_B = \max\{0, 50 + I \cdot 80 - P\} - I \cdot C; \quad C \in \{20, 40, 60\}$$

$$p_S = \begin{cases} P & \text{if } P \leq 50 + I \cdot 80 \\ 0 & \text{otherwise} \end{cases}$$

Figure 2: Frequency distribution of price demands in the observable investment case (by costs level)

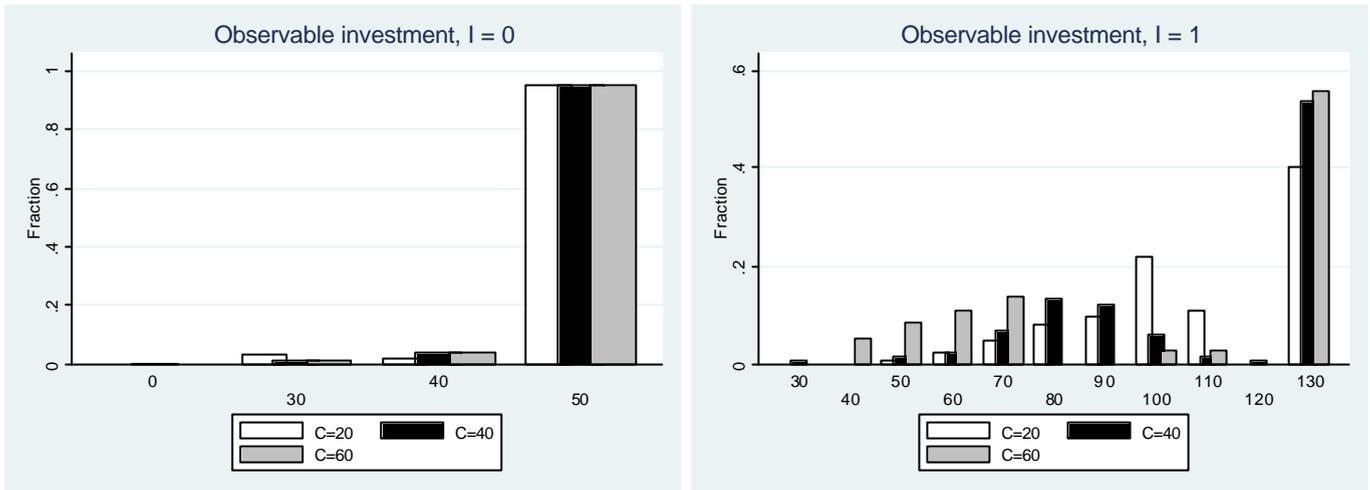


Figure 3: Frequency distribution of price demands in the unobservable investment case (by costs level)

