

# Does making specific investments unobservable boost investment incentives?\*

Randolph Sloof,<sup>†</sup>Hessel Oosterbeek and Joep Sonnemans

February 7, 2005

## Abstract

Economic theory predicts that holdup can be alleviated by making specific investments unobservable to the other party. Private information creates an informational rent that boosts investment incentives. Empirical findings, however, indicate that holdup is attenuated by fairness and reciprocity considerations. Private information may interfere with this, as it becomes impossible to directly observe whether the investor behaved fair or not. In that way unobservability could crowd out the 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 fairness and reciprocity. But with sufficient scope for these motivational factors, unobservability does not boost specific investments.

## 1 Introduction

When a party makes a relationship-specific investment, this investment is at risk because the other party may terminate the relationship. Anticipating that she may be unable to reap the full return, the investor will invest less than the efficient level. This is the well-known holdup underinvestment problem. This problem is considered to be of central importance in a wide variety of economic contexts (cf. Klein et al. 1978, Williamson 1985). For example,

---

\*An earlier version of this paper was presented at conferences at Harvard University (ESA), in Erfuhrt (ESA) and in Madrid (EEA), and at seminars at NYU, the University of Bonn and the University of Rotterdam.

<sup>†</sup>Corresponding author; e-mail: r.sloof@uva.nl. All authors are affiliated with the Department of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands.

it serves as the cornerstone of the property rights theory of the firm; see Hart (1995) for an overview.

In an early contribution Tirole (1986, Proposition 3) shows that investment unobservability may alleviate the holdup problem. Gul (2001) obtains that private information about the specific investment made may even help solving holdup completely. The general idea underlying these theoretical results is quite intuitive. Private information yields the informed party an informational rent in the ex post bargaining. This boosts the ex ante investment incentives of this party and weakens those of the uninformed party. Therefore, by giving the investor some informational advantage, one might be able to alleviate holdup. Indeed, one objective of Gul (2001, p. 344) is:

“...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.”

Rogerson (1992), Lau (2002) and Gonzalez (2004) similarly suggest that private information rents might substitute for bargaining power in solving holdup.

Theoretically the creation of private information rents may be an effective instrument to boost investment incentives. However, a large number of experimental studies on holdup in complete information settings have found that subjects typically invest significantly more than predicted.<sup>1</sup> The experimental findings all point at the importance of fairness and reciprocity considerations as informal mechanism that alleviates holdup. Investment is typically seen as fair 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. 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 data set contains information about participation in work-related training, who (either employer or worker) 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.

---

<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) and Sonnemans et al. (2001).

The difference of 15 percentage points appears highly significant. Employers thus seem to actively rely on reciprocity as informal remedy against holdup.

When in practice holdup is less severe, there is less need for explicit solutions like private information rents. More importantly, a serious concern is that private information may interfere with the informal fairness/reciprocity mechanism in place. This holds because it becomes impossible to directly observe whether the investor has behaved fair or not. Some experimental evidence that this indeed may happen is provided in Hackett (1994), see the discussion below. Unobservability of the investment decision does not necessarily rule out fairness or reciprocity considerations altogether, but it is likely to affect the scope for these motivational factors. In the next section we develop this point more formally. We show that when parties are sufficiently motivated by considerations of fairness or reciprocity, investment levels are predicted to be independent of whether the investment is observable or not. In case such considerations are weak or absent the predictions of standard theory pertain. Whether investment unobservability really boosts investment is therefore ultimately an empirical question. This paper addresses this question by means of a controlled laboratory experiment.

The experiment concerns a two-stage game between a buyer and a seller. In the first stage the buyer can make an investment that raises her valuation of the seller's good. In the second stage the seller unilaterally determines the trading price. In one information condition the seller observes the buyer's investment decision, in another information condition he does not. 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, the investment costs are common knowledge and can take three values: low, intermediate or high. Standard theory predicts that, compared to the observable investment case, unobservability increases investment levels by the same amount independent of the costs of investment. In contrast to this, the fairness and reciprocity models predict that with lower investment costs the effect of unobservability is smaller and may even disappear.

Our results are in line with the fairness/reciprocity predictions. When investment costs are high such that there is little 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 much scope for fairness and reciprocity – mean investment

levels in the two conditions are the same.

A few other experimental studies on holdup under asymmetric information exist. Hackett (1994) considers a setting in which investments stochastically increase the surplus up for renegotiation; a higher investment increases the probability that this surplus is large. The resulting size of the surplus is always publicly observable.<sup>2</sup> The theoretically predicted investment levels are therefore independent of whether the investment itself is publicly observable or not. However, Hackett finds that investments levels are somewhat higher when they are publicly observable than when they are not. These differences are significant when bargaining power is balanced, but not when one party has much more power than the other. Oosterbeek et al. (1999) also look at a situation in which investments stochastically affect the available surplus. They find no significant differences in investment rates between the public and the private information treatment, in line with the theoretical predictions that apply in their setup.

In Ellingsen and Johannesson (2005) the investor makes an ultimatum offer about the division of the (observable) surplus created by her investment. In one of their treatments the investor is privately informed about the actual costs of investment, in two other treatments these costs are publicly observed. Standard predictions are exactly the same in the three treatments; the investor invests efficiently and obtains the complete surplus. Ellingsen and Johannesson indeed find that investment rates do not differ significantly across treatments.

The essential difference between our experiment and these previous studies 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 private information to the buyer, and she is predicted to obtain an informational rent in the bargaining stage. This theoretically boosts her investment incentives. In contrast to the earlier studies, standard predictions therefore differ between our two information conditions. Whereas previous experiments compare situations that theoretically yield the same investment results (and by and large find that this is indeed the case), here we compare two situations that are predicted to lead to different investment incentives (and we find that this is not always the case).

The remainder of this paper is organized as follows. In the next section we present the simple game on which our experiment is based. This section

---

<sup>2</sup>Ruffle (1998) considers dictator and ultimatum game experiments with endogenous pie creation. The size of the pie is determined by the recipient's/responder's relative performance in a pre-game knowledge quiz. This also corresponds to a situation where the investment itself (i.e. the effort put into getting the answers right) is unobservable to the other party, while the return to investment (i.e. the resulting pie size) is observable.

also presents the standard equilibrium predictions, as well as alternative predictions based on distributional preferences and intention-based reciprocity. Section 3 provides the details of the experimental design. Results are presented and discussed in Section 4. The final section concludes.

## 2 Theory

### 2.1 Basic setup of the model

Consider a bilateral relationship between a female buyer and a male seller who may trade one unit of a particular good. 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.

Note that in the second bargaining stage 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 above her valuation, her only option is to reject. Because the buyer has no actual choice at this stage, the seller de facto determines the terms of trade.

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 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.

Table 1: Reduced strategic form under unobservable investment

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

## 2.2 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 Standard Prediction:

**SP**  $q_{un} - q_{obs}$  is positive and independent of  $C$ ; private information always boosts investment incentives.

Risk aversion does not affect this prediction.<sup>3</sup> To see this, note that the buyer always chooses  $q_{obs}^* = 0$ , 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. It therefore only depends on the risk attitude of the seller (and not on the risk attitude of the buyer herself); the more risk averse the seller is, the larger  $q_{un}^*$  will be. If anything, risk aversion thus only strengthens the incentive effect of private information.

---

<sup>3</sup>Similarly, noisy decision making per se does not alter relative investment incentives. It can be shown that also in the logit quantal response equilibrium of McKelvey and Palfrey (1998) it necessarily holds that  $q_{un}^* > q_{obs}^*$ .

The predicted outcome in the observable investment case equals the upper-left cell in Table 1. Compared to this investment unobservability leads to more investment. This 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. The expected gain owing to more investment and the expected loss due to inefficient separations cancel out; expected net social surplus under unobservable investment also equals  $V$ . This illustrates the conclusion of Gul (2001, pp. 348-349) that “[W]hile the unobservability of the investment decision alters the nature of equilibrium behavior, it does not change the equilibrium payoffs (i.e., the extent of inefficiency). The source of the inefficiency changes (underinvestment is reduced but the possibility of disagreement is added) but the amount of inefficiency is not decreased by the ability of the buyer to conceal his investment decision.”<sup>4</sup>

Given the above conclusion, one remark is in order. In his paper Gul (2001, pp. 343-344) argues that “[N]either unobservable investment nor frequently repeated offers alleviate the holdup problem; yet, the two together completely resolve it.” His solution to holdup thus contains two instruments: (i) informational rents and (ii) efficient trade due to the Coasian effect created by frequently repeated offers. Our setup only considers the former instrument. We do so for two reasons. First, our main interest lies in isolating the effect of informational rents on investment incentives. We therefore prefer to keep the bargaining stage as simple as possible. Second, Coasian dynamics are already studied experimentally in other papers.<sup>5</sup>

The two information conditions reflect a trade-off between efficient trade decisions and ‘high-powered’ investment incentives. This trade-off is studied by a number of authors in various contexts. Riordan (1990) argues that a crucial consequence of vertical integration is a change in information structure. The downstream firm obtains better information 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 trade-off between public and private ownership.

---

<sup>4</sup>Lau (2002) notes that the latter result is an artifact of considering the two polar cases of no and full information asymmetry only. She shows that for intermediate degrees of information asymmetry the efficiency gains owing to better investment incentives outweigh the losses due to inefficient separations.

<sup>5</sup>See Cason and Reynolds (2005), Cason and Sharma (2001), Güth et al. (1995), Rapoport et al. (1995) and Reynolds (2000). These experiments are designed to test models of durable goods monopolies.

He argues that 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. In the context of our simple game, under unobservable investment the seller keeps the buyer at 'arm's length'. This enables him to commit to a single unconditional price  $P$ . Under observable investment such a commitment not to behave opportunistically is non-credible, and thus cannot be used to provide investment incentives. Without commitment the seller can always take the efficient trade decision though.

### 2.3 Predictions based on fairness and reciprocity

The above predictions are based on the assumption that agents are solely driven by their own monetary payoffs. In reality this is typically not the case. A substantial fraction of the subjects in a variety of experiments reveal a concern for fairness and reciprocity (cf. Fehr and Gächter 2000). In their overview article Fehr and Schmidt (2002) distinguish two types of theories of fairness and reciprocity. The first type assumes that (some) agents have distributional preferences and also care about the payoffs of others. The second type of theories focuses on intention-based reciprocity and assumes that agents care about the intentions of their opponents. Here agents are partly driven by the motivation to reward fair behavior and to punish unfair behavior. In this subsection we also take these alternative motivations into account. This leads to predictions different from standard theory.

Subsections 2.3.1 and 2.3.2 below present detailed predictions of two representative models of distributional preferences and intention-based reciprocity. The main conclusion that follows from the analysis is that when fairness or reciprocity considerations are sufficiently weak, the standard predictions prevail. That is, the buyer never invests when the investment is observable whereas she invests with positive probability when the investment is unobservable. However, when fairness or reciprocity considerations are sufficiently strong, the buyer will always invest independent of the investment's observability. Private information then does not affect investment incentives. These results also imply that, relative to the standard predictions, the increase in the investment rate owing to fairness and reciprocity motivations is larger when the investment is observable than when it is unobservable.

Now, the scope for both fairness and reciprocity is larger when the costs of investment  $C$  are low relative to the return on investment  $W$ . Put differently, when  $C$  decreases while keeping  $W$  constant, it becomes more likely

that fairness and reciprocity considerations are ‘sufficiently’ strong. We thus obtain the qualitative prediction that when investment costs are relatively low, the investment level under unobserved investment equals the one under observed investment. In contrast, when  $C$  is relatively high, the predictions of both alternative models are similar to the standard predictions. In sum, we have the following prediction based on Fairness and Reciprocity:

**FRP**  $q_{un} = q_{obs}$  when  $C$  is low and  $q_{un} > q_{obs}$  when  $C$  is high; private information boosts investment incentives only when  $C$  is high.

Two remarks remain. First, our formal analysis of social preferences and reciprocity assumes complete information. However, in practice subjects are heterogeneous and typically privately informed on their own preferences. Some have strong social preferences or strong reciprocal attitudes, while others are completely selfish. Hence even at a low cost level, a fraction of the subjects is likely to behave selfish. Likewise, even at a high cost level some subjects may reveal social preferences or reciprocal attitudes. Yet we expect that when we aggregate over all subjects, the above qualitative predictions will pertain.

Second, our interest lies in the impact of private information on investment incentives. We do not intend to test particular models of social preferences or intention-based reciprocity per se. Numerous studies have already established the importance of these motivational forces and a number of experiments have been purposely designed to discriminate between the various models, see e.g. Cox (2004), Falk et al. (2000) and McCabe et al. (2003). We just want to point out that when such motivations are effectively important (i.e. when  $C$  is low), private information is predicted to have less or even no impact on investment behavior, independent of exactly how these motivations are modeled.

### 2.3.1 Distributional preferences

Following Charness and Rabin (2002) we assume that the utility of player  $i = B, S$  equals the following weighted average of monetary payoffs  $\pi_i$  and  $\pi_j$  ( $i \neq j$ ):<sup>6</sup>

$$\begin{aligned} u_i(\pi_i, \pi_j) &\equiv \pi_i + \rho_i \cdot (\pi_j - \pi_i) \equiv (1 - \rho_i) \cdot \pi_i + \rho_i \cdot \pi_j \text{ when } \pi_i \geq \pi_j \\ u_i(\pi_i, \pi_j) &\equiv \pi_i + \sigma_i \cdot (\pi_j - \pi_i) \equiv (1 - \sigma_i) \cdot \pi_i + \sigma_i \cdot \pi_j \text{ when } \pi_i \leq \pi_j \end{aligned} \quad (1)$$

---

<sup>6</sup>In order to focus on distributional preferences we consider the simplified version of the Charness and Rabin (2002) model, ignoring the negative reciprocity element (i.e. shift parameter  $\theta$ ) of the full model. Intention-based reciprocity is formally addressed in the next subsection.

The relative weight player  $i$  attaches to the other player's payoffs measures  $i$ 's fair-mindedness. Note that this weight depends on player  $i$  being ahead ( $\rho_i$ ) or behind ( $\sigma_i$ ). We assume that  $\sigma_i < \frac{1}{2}$ . This implies that player  $i$  is more concerned about her own payoffs than about those of the other party when she is behind. We also assume that  $\sigma_i \leq \rho_i$  and that  $\rho_i < 1$ . The latter entails that even when player  $i$  is ahead, she prefers more money to less, other things equal.

The above distributional preferences nest some particular types as special cases. For instance, competitive preferences arise when  $\sigma_i \leq \rho_i < 0$ . The inequity-aversion model of Fehr and Schmidt (1999) is obtained when we assume that  $\sigma_i < 0 < \rho_i < 1$ . Lastly, Charness and Rabin's quasi-maximin preferences require that  $1 \geq \rho_i \geq \sigma_i > 0$  (and  $\sigma_i < \frac{1}{2}$ ). This implies that players attach a positive weight to the other party's monetary payoffs, even when these payoffs exceed their own. For all these types of preferences we obtain the following result:

**Theorem 1** Suppose preferences are given by (1) and players know each others preferences:<sup>7</sup>

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

**(medium)** if  $\rho_S < \frac{1}{2}$  and  $\frac{C}{C+V+W} < \sigma_B < \frac{C}{C+W}$ , then  $1 = q_{un}^* > q_{obs}^* = 0$ ;

**(strong)** if  $\rho_S > \frac{1}{2}$  or  $\sigma_B > \frac{C}{C+W}$ , then  $q_{un}^* = q_{obs}^* = 1$ .

Theorem 1 distinguishes three cases depending on the fairness weights  $\rho_i$  and  $\sigma_i$ . In the weak case neither player is (strongly) motivated by fairness considerations. The 'strong' situation applies when either the seller or the buyer (or both) is strongly fair-minded. The remaining 'medium' case applies when the seller is not really fair-minded, while the buyer is only weakly so.

We next discuss the intuition behind Theorem 1. When  $\rho_S < \frac{1}{2}$  the seller's dislike for being better off is insufficient to justify own monetary sacrifice. He always asks for the complete actual pie under observable investment. Anticipating this, the buyer prefers to invest only if she cares sufficiently strong about the seller's payoff when the latter is ahead. Not investing yields her  $\sigma_B V$ , investing gives her  $(1 - \sigma_B)(-C) + \sigma_B(V + W)$ . We thus obtain that  $\sigma_B > \frac{C}{C+W}$  is required for  $q_{obs}^* = 1$ .

Also under unobservable investment the seller with  $\rho_S < \frac{1}{2}$  would like to claim the complete pie. Suppose that the seller is convinced that the pie is

---

<sup>7</sup>We exclude the degenerate case in which  $\rho_S = \frac{1}{2}$ , i.e. the case the seller is indifferent between any division that weakly favors himself. Also knife-edge cases for  $\sigma_B$  are excluded. The proof appears in Appendix A.1.

large. He then asks for  $P = V + W$ . In that case investing again yields the buyer  $(1 - \sigma_B)(-C) + \sigma_B(V + W)$ . But now no investment yields her a payoff of 0 rather than  $\sigma_B V$ , because the small pie is wasted. A buyer with  $\sigma_B > 0$  cares about this waste and is therefore more easily persuaded to invest for sure. Now only  $\sigma_B > \frac{C}{C+V+W}$  is needed. Roughly put, under unobservable investment the buyer's threat to abstain from investment is less credible, because she cares about the risk that the small pie is wasted. Finally, in case  $\rho_S > \frac{1}{2}$  the seller's dislike for being better off is that high that he prefers to propose an equal split of the net surplus. Anticipating this, the buyer invests for sure.

From Theorem 1 it follows that the buyer always invests weakly more under unobservable investment. In case social preferences are weak the predicted investment rates are comparable to standard theory. Also under medium social preferences investment levels are higher in the unobservable investment case. But, when either the buyer or the seller is sufficiently fair-minded, the buyer invests for sure in both information conditions.

Because the cutoff value for the buyer's fair-mindedness  $\sigma_B$  depends on the costs of investment,<sup>8</sup> these qualitative predictions can be rephrased in terms of  $C$ . If  $C$  increases some buyers will move from the 'strong' to the 'medium' or 'weak' category, and some from the 'medium' to the 'weak' category. In the observable investment case the former means fewer investments. In the unobservable investment case, more buyers with 'weak' social preferences will also decrease the probability of investment. We therefore conclude that when the investment costs are high, the qualitative predictions about investment levels mimic those from standard theory. If, on the other hand, investment costs are low, the qualitative prediction is that investment levels under unobservability are likely to equal those under observability.

---

<sup>8</sup>Instead of referring here to  $\sigma_B$  as the buyer's degree of fair-mindedness, it might be more appropriate to refer to  $\sigma_B$  as her degree of altruism. Note that the preferences in (1) represent a general model of distributional concerns. One could argue that only when  $\sigma_i < 0$  (and  $\rho_i > 0$ ) one can speak of a 'fairness' model. Now for  $\sigma_B < 0$  either the weak or the strong case of Theorem 1 applies, and the case distinction is independent of investment costs  $C$ . Only when  $\sigma_B > 0$ , i.e. when the buyer is 'altruistic', the dividing condition does depend on  $C$ . Typically, however, fairness models also assume incomplete information about other players' preferences (cf. Fehr and Schmidt 1999). And in that case equilibrium investments do vary negatively with  $C$ , even when  $\sigma_B < 0$ . To illustrate, let  $\theta_F$  be the fraction of players with  $\rho_i > \frac{1}{2}$ . From the proof of Theorem 1 follows that, in the observable investment case, the buyer invests whenever  $\sigma_B > \frac{2C - \theta_F(W+C)}{2(1-\theta_F)(W+C)}$ . The r.h.s. is an increasing function of  $C$  and negative for low values of  $C$ . So, when  $C$  increases the fraction of buyer types that are predicted to invest decreases, even when  $\sigma_B < 0$ .

### 2.3.2 Intention-based reciprocity

The reciprocity concept of Dufwenberg and Kirchsteiger (2004) is not based on a payoff comparison between players. They measure reciprocity with reference to the range of what one player could give the other player in principle. In their model utility functions take the following form:

$$\begin{aligned} u_B &= \pi_B + Y_B \cdot \kappa_{BS} \cdot \lambda_{BSB} \\ u_S &= \pi_S + Y_S \cdot \kappa_{SB} \cdot \lambda_{SBS} \end{aligned} \quad (2)$$

Again  $\pi_i$  denotes monetary payoffs of player  $i$  ( $i = B, S$ ), while  $Y_i \geq 0$  gives this player's reciprocal attitude. The higher  $Y_i$ , the more sensitive to reciprocity is  $i$ . The factor  $\kappa_{ij}$  represents  $i$ 's kindness to  $j$ . It is positive if  $i$  is kind to  $j$  and negative if  $i$  is unkind to  $j$ . The factor  $\lambda_{iji}$  gives  $i$ 's belief about how kind  $j$  is to  $i$ . It is positive when  $i$  believes that  $j$  is kind to him, and negative when  $i$  thinks that  $j$  is unkind to him. Reciprocity is captured by the incentive to match the sign of  $\kappa_{ij}$  with the sign of  $\lambda_{iji}$ . Each player attaches a non-negative payoff to being (un)kind towards the other party when the latter has been (un)kind towards him or her.

Both factors  $\kappa_{ij}$  and  $\lambda_{iji}$  depend on player  $i$ 's beliefs. Dufwenberg and Kirchsteiger (2004) provide exact definitions of how  $\kappa_{ij}$  and  $\lambda_{iji}$  are measured. Because utility now also depends on the players' beliefs, psychological game theory has to be used. 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 full equilibrium analysis is quite involved. It appears that for the buyer's equilibrium investment behavior the reciprocal attitude of the seller  $Y_S$  is decisive. In particular, irrespective of the information condition that applies,  $q^* = 1$  is possible if and only if  $Y_S$  is sufficiently high. For ease of exposition we therefore assume here that  $Y_B = 0$ . This implies that the buyer is not reciprocal at all.<sup>9</sup> In addition, we focus on the situation considered in the experiment in which the investment significantly improves the gross surplus up for division, i.e.  $V < W$ .

**Theorem 2** Let  $Y_B = 0$  and consider the case  $V < W$ . Then for any SRE:

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

---

<sup>9</sup>Dufwenberg and Kirchsteiger (2000) make a similar simplifying assumption in their analysis of employer-worker relationships. In Appendix A.2 we provide the complete equilibrium analysis for the more general case  $Y_B \geq 0$  and show that this leads to qualitatively the same results as in Theorem 2.

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

**(strong)** if  $Y_S > \frac{2}{W-C}$ , then  $q_{un}^* \leq q_{obs}^* = 1$ . In the unobservable investment case always a SRE exists in which the buyer chooses  $q_{un}^* < \frac{V}{V+W}$ , besides one in which she chooses  $q_{un}^* = 1$ .

The three different cases in Theorem 2 are separated on the basis of the seller's reciprocal attitude  $Y_S$ . In the observable investment case the equilibrium appears to be unique. In case the seller is sufficiently reciprocal ('strong' case) the buyer invests for sure. Otherwise she does not invest. In the unobservable investment case also two types of equilibria exist. In the first, *positive reciprocity* equilibrium the buyer invests for sure. This equilibrium exists whenever  $Y_S > \frac{2}{V+W-C}$ . In the second, *negative reciprocity* equilibrium the buyer invests with probability  $0 < q_{un}^* \leq \frac{V}{V+W}$ . This equilibrium exists irrespective of the value of  $Y_S$ . One important difference between the two information conditions is thus that only in the unobservable investment case the two different equilibria may exist side by side. Another important difference is that  $q_{un}^* = 1$  is possible for a larger set of values of  $Y_S$  than  $q_{obs}^* = 1$  is. The scope for positive reciprocity is thus larger under unobservable investment.

The intuition behind the negative reciprocity equilibrium is as follows. For  $V < W$  it holds that  $\frac{V}{V+W} < \frac{1}{2}$ . Therefore, would the buyer choose  $q_{un}^* = \frac{V}{V+W}$  as standard theory predicts her to do, the seller considers this as unkind. His reciprocity payoff then induces him to punish the buyer by choosing  $P = V + W$  for sure. To counterbalance this the buyer chooses  $q_{un}^* < \frac{V}{V+W}$  to let the seller prefer  $P = V$  on the basis of monetary payoffs only. Taking both the monetary and the reciprocity payoffs into account, the incentives cancel out and the seller mixes between the two prices. He does so as to make the buyer indifferent between investing or not.

From Theorem 2 it follows that when reciprocity considerations are weak, the buyer invests more under unobservable investment than under observable investment. This situation becomes more likely the higher are the costs of investment  $C$ . However, when the seller is sufficiently sensitive to intention-based reciprocity, private information does not boost investment incentives. This case is likely to apply when  $C$  is relatively low.<sup>10</sup>

<sup>10</sup>Comparing the negative reciprocity equilibrium with  $q_{obs}^* = 1$  in the 'strong' case suggests that private information may even *weaken* investment incentives. This only strengthens our observation that when  $C$  is low, private information cannot be used as an instrument to encourage investments.

### 3 Experimental design

The experiment is based on a  $2 \times 3$  design. For both the observable and unobservable investment case we consider three different levels of the investment costs:  $C \in \{20, 40, 60\}$ . We chose the other parameters to be equal to  $V = 50$  and  $W = 80$ . Within each session we kept the information condition fixed, so each subject was confronted with only one of the two information conditions. We ran six sessions in total. Three sessions considered the observable investment case, the other three the unobservable investment case. 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, i.e. subjects knew that buyers decided on their investment before sellers chose their price demand. We did so to make both information conditions fully comparable.<sup>11</sup> To exclude dominated strategies in the observable investment case, the seller could never ask for more than the actual pie. Figure 1 depicts the structure of the experimental games.

[ Insert Figure 1 about here ]

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. Average earnings were 27.65 euros in about one and a half hours. Earnings varied considerably, with the minimum actual earnings equal to 7.60 euros and a maximum of 51.50 euros.

Each session contained 36 rounds. We employed a block structure of rounds 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 observability case we used the ‘upward’ ordering (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

---

<sup>11</sup>According to standard game theory, the physical timing of moves does not matter in the unobservable investment case. All that matters is that the seller is uninformed about the buyer’s investment when choosing his price. Experimentally the physical timing may make a difference though. For example, in coordination game experiments typically a *positional order effect* is found. Knowledge that one player makes his decision before the other player does, leads to coordination on the equilibrium that is advantageous for the first mover, see e.g. Rapoport (1997), Müller (2001) and Güth et al. (1998).

that consider the same value of  $C$  we can test for learning effects. By comparing the two different orderings we can control for order effects.

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 actually 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>12</sup>

Subject roles' varied over the rounds. Within each block of six rounds each subject had the role of buyer exactly three times, and the role of seller also three times.<sup>13</sup> The experiment used a stranger design. Subjects were anonymously paired and their matching varied over the rounds. Within each block of six rounds subjects could meet each other only once. Subjects were explicitly informed about this. Moreover, within a session we divided the subjects into two separate groups of ten subjects. Matching of pairs only took place within these groups. We did so to generate two independent aggregate observations per session. With three sessions for the observable investment case and three sessions for the unobservable investment case, we thus have six independent observations per treatment at the aggregate group level.

The experiment was computerized. Subjects started with on-screen instructions. Before the experiment started all subjects first had to answer a number of control questions correctly. Subjects also received a summary of the instructions on paper (see Appendix A.3). At the end of the experiment subjects filled out a short questionnaire and the earned experimental points were exchanged for money.

## 4 Results

In our analyses 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 (see Appendix A.4). We also pool the results of

---

<sup>12</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>13</sup>We use 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.

sessions that differ only in the ordering of the investment costs  $C$ . Although some order effects can be detected in the data, these are only minor. Further aggregations are not possible, as it appears that behavior evolves over time. Most findings are therefore reported separately for the first and second half of the experiment.

## 4.1 Investment levels

Within each block of six rounds subjects have the role of buyer (investor) exactly 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 3 we divided the 20 subjects within a session into two groups of 10 subjects that were independently matched. Members of one group were never matched with members of the other group. 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 under unobservable investment than under observable investment.* (b) *With low investment costs, mean investment levels under unobservable investment and under observable investment are equal.*

Evidence supporting Result 1 is provided in Table 2. This table reports the mean investment levels by treatment and gives the test statistics for equality of these levels across treatments (ranksum tests). When investment costs equal  $C = 20$  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>14</sup> For  $C = 60$  the difference is significant for both halves, and largest in absolute and relative magnitude.

---

<sup>14</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 2: 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:* Theoretical predictions 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  $m = n = 60$  ( $m = n = 6$ ).

Our second result compares mean investment levels between 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 2. With 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%. With low costs of investment, mean investment levels are well above the predicted levels of  $38\frac{1}{2}\%$  and 0% respectively. With high costs of investment, mean investment levels are fairly close to these point predictions. This is especially true during the second half of the experiment. Table 3 reports the relevant  $p$ -values. Because comparisons are on a within-subject/group 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 particularly robust.

Table 3:  $p$ -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.

Results 1(a) and 2(b) are in line with standard equilibrium predictions, Results 1(b) and 2(a) are not. Standard theory predicts that for both information conditions the propensity to invest is independent of  $C$ . Reciprocity and/or fairness provide an explanation. As discussed in Section 2 the scope for these motivational factors decreases with  $C$ . Therefore, when  $C$  increases, buyers should be less willing to invest. This is exactly what we observe.<sup>15</sup> Result 2(b) demonstrates that with  $C$  large enough, the impact of alternative motivations is likely to be weak and the predictions of standard theory and fairness/reciprocity theories will coincide. Overall we conclude that unobservability of the specific investment made does boost investment incentives. But, it only does so when fairness and reciprocity motivations do not provide strong enough incentives to invest.

## 4.2 Demand decisions

Although our main interest lies in buyers' investment decisions, to understand these we have to analyze sellers' demands. In the observable investment case the seller can condition his price demand on the investment level observed.

<sup>15</sup>For the unobservable investment case in which the game theoretically reduces to a simultaneous move game with a unique mixed strategy equilibrium this is a common experimental finding. Ochs (1995), for instance, already found that when a player's payoff from a particular strategy decreases, this player is less likely to choose this particular strategy (although standard theory predicts only the other player to adapt).

Here we thus have to consider the contingencies of no-investment ( $I = 0$ ) and investment ( $I = 1$ ) separately. Figures 2 and 3 depict the frequency distributions of demands by treatment. Here separate demand decisions rather than the (individual or group) mean demands are the units of observation. In the figures 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. If only distributional preferences would play a role, we would predict no differences between the situation in which the small pie is exogenously fixed (i.e. a dictator game) and the one where it is endogenously chosen (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 in the observable investment case, 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 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. Both fairness and positive reciprocity provide an explanation for this.

In the unobservable investment case subjects typically choose between  $P = 50$  and  $P = 130$  (cf. Figure 3). This yields a clear bi-modal distribution. 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  $C = 60$ . Standard theory predicts that the seller mixes between  $P = 50$  and  $P = 130$ , where the low demand is chosen with probability  $\frac{C}{80}$ . The latter probability corresponds fairly well with the observed frequencies. This also holds when we take all demands weakly below 50 as a choice for  $P = 50$ , and all higher demands as a choice for  $P = 130$ . The relative frequency that the low demand

Table 4: 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:*  $p$ -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.

is chosen then equals 31% when  $C = 20$ , 54% in case  $C = 40$  and 73% for  $C = 60$ . Also these percentages are in line with the predicted levels of 25%, 50% and 75%. However, there are also clear indications for fair or reciprocal behavior. 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 4 and 5 present the mean demands in the various treatments, together with the predicted expected price. The lower parts present the  $p$ -values of signed-rank tests that compare the different costs situations.<sup>17</sup> For the observable investment case the mean demand appears

<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.

<sup>17</sup>In Table 4 the mean demands are conditional on either no-investment ( $I = 0$ ) or investment ( $I = 1$ ). It can occur that within a block a subject is never confronted with one of these contingencies. This individual is then left out from the corresponding signed-rank test based on individual means. The  $p$ -values obtained from the tests at the individual level are thus based on different sample sizes. Within parentheses appear the numbers of

Table 5: 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
<hr/>			
$C = 20$ 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:*  $p$ -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 investment levels.

to be 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. 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 predicted expected demand of  $P = 130 - C$ .

The findings of this subsection are summarized in Result 3.

**Result 3.** (a) *When  $I = 0$  is observed, sellers almost always demand  $P = 50$ . When  $I = 1$  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  $C$ .*

observations. A similar remark applies for the group level data. In two cases we have observations for 3 groups only, such that no sensible test statistic is available (n.a.).

### 4.3 Learning

In the previous subsections we reported the results for the first and second half of the experiment separately. As Appendix A.4.3 shows some learning effects can be detected. Learning can have two causes. First, some subjects may learn 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 the investment decisions accordingly.

First consider the observable investment case. In this case sellers practically always ask the whole pie of 50 points when no investment is made in both parts of the experiment. When the buyer invested, sellers demanded a larger part of the pie in the second half of the experiment (cf. Table 4). But, the difference is significant only when  $C = 20$ .<sup>18</sup> In the first part of the experiment buyers made a modest profit when making a low cost ( $C = 20$ ) investment. In the second half of the experiment this turned into a small loss. As a consequence, completely selfish buyers prefer not to invest in the second part of the experiment. Investments indeed decrease significantly (for all cost levels), but when  $C = 20$  still 57% of the buyers invest. Note that in this case investment is rational for buyers with a social preference parameter  $\sigma_B \geq \frac{5.53}{60}$ . On average buyers lose a few points (5.53), but the increase in earnings for the sellers (65.53) are such that even relatively weak social preferences (for example, weighing own earnings 10 times as heavy as the other player's earnings) make investment worthwhile. This suggests that playing more rounds will not bring the outcome close to complete holdup, i.e. the outcome standard theory predicts.

In the unobservable investment treatment the increase in demands is significant only when  $C = 20$ . However, mean demands stay below the standard prediction of  $130 - C$ . This indicates that also here reciprocal behavior and fairness play a role. The decrease in investment levels between the first and second half of the experiment is not statistically significant.

Overall we conclude that learning effects do not affect our main findings with respect to investment behavior, viz. Results 1 and 2.

### 4.4 Efficiency

Finally, we look at efficiency losses. Table 6 reports the mean inefficiency losses by information condition, for the first and the last 18 rounds separately. Predicted inefficiencies are between brackets. In the unobservable investment

---

<sup>18</sup>The number of buyers who invest when  $C = 60$  is too small to test the difference in demand.

Table 6: 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 [36.92]	11.39 [23.08]	32.72 [60]	25.67 [60]	.2207	.4688
$C = 40$	1-18	17.56	10	27.56	21.56	.1481	.2615
	19-36	21.78 [24.62]	13.61 [15.38]	35.39 [40]	32.44 [40]	.0192	.4201
$C = 60$	1-18	13.33	9.44	22.77	17.56	.0215	.0054
	19-36	12.33 [12.31]	7.77 [7.69]	20.11 [20]	18.44 [20]	.0031	.6242

*Remark:* Predicted inefficiencies appear in square brackets.  $p$ -values refer to Mann-whitney ranksum tests performed on group level data (with  $m = n = 6$ ).

treatment 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 then the seller can never demand more than the actual pie by design. 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. 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 by means of ranksum tests. Because efficiency losses can only be calculated for a buyer-seller pair, tests cannot be based on individual means. In this subsection we therefore only consider tests performed at the aggregate matching group level. In the observable investment case trade inefficiency is zero by design. When the investment is unobservable, the group means are always strictly positive. For

all cost levels trade inefficiency is therefore larger in the unobservable investment treatment. For the other, more interesting comparisons the  $p$ -values are reported in Table 6. The second to last column reports the test statistics of comparing investment inefficiencies across information conditions. The results reiterate our earlier conclusions about mean investment levels. The last column concerns the comparison of overall inefficiencies. Although overall inefficiencies 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 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 experimental findings indicate that this will be the case only when there is insufficient scope for reciprocity or fairness, 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, reciprocity and fairness are at work and these alternative motivations have a larger impact on investment incentives when the investment decision is observable. 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 boost investment incentives and is ineffective in alleviating holdup.

Our results tentatively suggest that private information is an effective instrument only when the efficiency gains of solving holdup are small. Clearly our experiment just provides a first step and a number of interesting questions remain. For instance, the specific bargaining stage that we employ is not particularly realistic. In practice parties can at least indicate whether they accept or reject the terms of trade. Standard predictions remain unchanged when we would have used a setup in which the buyer can actually choose whether to accept or reject the seller's price demand. Yet then additional strategic issues and motives may come into play, because the buyer can then reciprocate with here acceptance/rejection decision. Anticipating this, the seller may change his demand behavior. Now that we have established that in the simplest possible setup unobservability may indeed affect investment incentives when the relative costs of investment are high, future experiments can build on this and investigate whether this result generalizes to more involved bargaining settings.

It seems also of interest to study the impact of private information rents on investment incentives in settings with a different information structure. One interesting situation, especially in the context of investments in specific human capital, is one in which parties have private information about their outside opportunities (cf. Malcomson 1997). A recent experiment that considers such situations is reported in Sloof (2003).

## 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.
- Cason, T. and S. Reynolds (2005). Bounded rationality in laboratory bargaining with asymmetric information. *Economic Theory* 25, 553–574.
- Cason, T. and T. Sharma (2001). Durable goods, coasian dynamics, and uncertainty: Theory and experiments. *Journal of Political Economy* 109, 1311–1354.
- Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117, 817–869.
- Cox, J. (2004). How to identify trust and reciprocity. *Games and Economic Behavior* 46, 260–281.
- 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 S. Gächter (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14, 159–181.
- 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.
- Fehr, E. and K. M. Schmidt (2002). Theories of fairness and reciprocity - Evidence and economic applications. In M. Dewatripont, L. Hansen, and S. Turnovsky (Eds.), *Advances in Economics and Econometrics - 8th World Congress, Econometric Society Monographs*. Cambridge University Press.
- 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.
- Güth, W., S. Huck, and A. Rapoport (1998). The limitations of the positional order effect: can it support silent threats and non-equilibrium behavior? *Journal of Economic Behavior and Organization* 34, 313–325.
- Güth, W., P. Ockenfels, and K. Ritzberger (1995). On durable goods monopolies: An experimental study of intrapersonal price competition and price discrimination over time. *Journal of Economic Psychology* 16, 247–274.
- 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.

- Hart, O. (1995). *Firms, Contracts and Financial Structure*. Oxford University Press.
- 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*. Springer Verlag.
- 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.
- Malcomson, J. (1997). Contracts, holdup, and labor markets. *Journal of Economic Literature* 35, 1916–1957.
- McCabe, K., M. Rigdon, and V. Smith (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization* 52, 267–275.
- McKelvey, R. and T. Palfrey (1998). Quantal response equilibria for extensive form games. *Experimental Economics* 1, 9–42.
- Müller, W. (2001). The quality of the signal matters - a note on imperfect observability and the timing of moves. *Journal of Economic Behavior and Organization* 45, 99–106.
- Ochs, J. (1995). Games with unique mixed strategy equilibria: An experimental study. *Games and Economic Behavior* 10, 202–217.
- Oosterbeek, H., R. Sloof, and J. Sonnemans (1999). Promotion rules and skill acquisition: An experimental study. Scholar working paper series WP 07/09.
- 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.
- Rapoport, A. (1997). Order of play in strategically equivalent games in extensive form. *International Journal of Game Theory* 26, 113–126.
- Rapoport, A., I. Erev, and R. Zwick (1995). An experimental study of buyer-seller negotiation with one-sided incomplete information and time discounting. *Management Science* 41, 377–394.

- Reynolds, S. (2000). Durable-goods monopoly: laboratory market and bargaining experiments. *Rand Journal of Economics* 31, 375–394.
- 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.
- Ruffle, B. (1998). More is better, but fair is fair: tipping in dictator and ultimatum games. *Games and Economic Behavior* 23, 247–265.
- Schmidt, K. (1996). The costs and benefits of privatization: An incomplete contracts approach. *Journal of Law, Economics and Organization* 12, 1–24.
- Sloof, R. (2003). Price-setting power versus private information: An experimental evaluation of their impact on holdup. Scholar working paper series WP 42/04.
- 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.

# Appendices

## A.1 Proof of Theorem 1

Theorem 1 follows from Propositions 1 and 2 below. Proposition 1 resembles Proposition 2 in Ellingsen and Johannesson (2004b) who analyze the inequality-aversion model of Fehr and Schmidt (1999) in a holdup context.

**Proposition 1** (Observable investment) Suppose preferences are given by (1) and players know each others preferences. Then the unique equilibrium is given by:

- (a)  $\rho_S < \frac{1}{2}$  and  $\sigma_B < \frac{C}{C+W}$ :  $q_{obs}^* = 0$ ,  $P_0^* = V$  and  $P_1^* = V + W$  ;
- (b)  $\rho_S < \frac{1}{2}$  and  $\sigma_B > \frac{C}{C+W}$ :  $q_{obs}^* = 1$ ,  $P_0^* = V$  and  $P_1^* = V + W$  ;
- (c)  $\rho_S > \frac{1}{2}$ :  $q_{obs}^* = 1$ ,  $P_0^* = \frac{V}{2}$  and  $P_1^* = \frac{V+W-C}{2}$ .

*Proof.* We derive all equilibria of the general incomplete information case in which players do not know each others preferences. Let  $\theta_F \equiv \Pr(\rho_S > \frac{1}{2})$ , such that  $1 - \theta_F \equiv \Pr(\rho_S < \frac{1}{2})$ . We thus assume that  $\Pr(\rho_S = \frac{1}{2}) = 0$ . The subscript  $F$  here refers to the fair-types.

Owing to  $\sigma_S < \frac{1}{2}$  the seller always demands at least half of the net surplus. Hence  $P_0^* \geq \frac{V}{2}$  and  $P_1^* \geq \frac{V+W-C}{2}$ . First consider a seller-type with  $\rho_S < \frac{1}{2}$ . This type considers own payoffs more important also when he is ahead. Hence  $P_0^* = V$  and  $P_1^* = V + W$ . Next consider a fair seller with  $\rho_S > \frac{1}{2}$ . This type always prefers an equal split of the net surplus and thus demands  $P_0^* = \frac{V}{2}$  and  $P_1^* = \frac{V+W-C}{2}$ .

Turning to the buyer, choosing  $I = 0$  yields her  $\theta_F \cdot (\frac{V}{2}) + (1 - \theta_F) \cdot (\sigma_B \cdot V)$ . Choosing  $I = 1$  gives her  $\theta_F \cdot (\frac{V+W-C}{2}) + (1 - \theta_F) \cdot (\sigma_B \cdot (V + W + C) - C)$ . She thus prefers to invest whenever  $\theta_F \cdot \{(\frac{1}{2} - \sigma_B) \cdot (W + C)\} \geq C - \sigma_B(W + C)$ . The term within  $\{\cdot\}$  is strictly positive. We obtain that the buyer invests whenever  $\theta_F \geq \frac{C - \sigma_B(W + C)}{(\frac{1}{2} - \sigma_B) \cdot (W + C)}$ . Note that the r.h.s. of this inequality is increasing in  $C$ . Hence when  $C$  increases investment becomes less likely. Parts (a) and (b) in Proposition 1 now immediately follow from setting  $\theta_F = 0$ , part (c) from taking  $\theta_F = 1$ . *QED*

**Proposition 2** (Unobservable investment) Suppose preferences are given by (1) and players know each others preferences. Let  $h(\sigma_B, \rho_B) \equiv (\rho_B - \sigma_B) \cdot \max\{(W - C) - V, 0\}$ . Then the unique equilibrium is given by:

- (a)  $\rho_S < \frac{1}{2}$  and  $\sigma_B < \frac{C}{C+V+W}$ :  $q_{un}^* = \frac{(1-\rho_S)V}{(1-\rho_S)V+(1-2\rho_S)W+h(\sigma_B,\rho_B)}$ ,  $P = V$  occurs with probability  $p^* = \frac{[C-\sigma_B \cdot (V+W+C)]}{[(1-2\sigma_B) \cdot W - \sigma_B \cdot V - h(\sigma_B,\rho_B)]}$  and  $P = V + W$  with probability  $1 - p^*$ ;
- (b)  $\rho_S < \frac{1}{2}$  and  $\sigma_B > \frac{C}{C+V+W}$ :  $q_{un}^* = 1$  and  $P^* = V + W$ ;
- (c)  $\rho_S > \frac{1}{2}$ :  $q_{un}^* = 1$  and  $P^* = \frac{V+W-C}{2}$ .

*Proof.* We first prove for the general incomplete information case with  $\theta_F \equiv \Pr(\rho_S > \frac{1}{2}) \in [0, 1]$  that necessarily  $q_{un}^* > 0$ . Suppose  $q_{un}^* = 0$ . The selfish seller with  $\rho_S < \frac{1}{2}$  then chooses  $P = V$  for sure, while the fair type of seller with  $\rho_S > \frac{1}{2}$  chooses  $P = \frac{V}{2}$ . The expected payoff for the buyer of choosing  $I = 0$  then equals  $(1 - \theta_F) \cdot (\sigma_B V) + \theta_F \cdot \frac{V}{2}$ . Choosing  $I = 1$  yields her  $(1 - \theta_F) \cdot [W - C + \sigma_B \cdot \max\{V - (W - C), 0\} - \rho_B \cdot \max\{(W - C) - V, 0\}] + \theta_F \cdot \{W - C + \frac{V}{2} - \rho_B(W - C)\}$ . Now  $\sigma_B < \frac{1}{2}$  and  $\sigma_B \leq \rho_B < 1$  together imply that  $[W - C + \sigma_B \cdot \max\{V - (W - C), 0\} - \rho_B \cdot \max\{(W - C) - V, 0\}] > \sigma_B V$ . Moreover,  $\rho_B < 1$  implies that  $W - C + \frac{V}{2} - \rho_B(W - C) > \frac{V}{2}$ . Hence  $I = 1$  yields the buyer strictly more and  $q_{un}^* = 0$  cannot be optimal. Hence necessarily  $q_{un}^* > 0$ .

The complete equilibrium analysis for the incomplete information case is tedious. In the sequel we therefore confine ourselves to a direct proof of parts (a) through (c).

*Part (a).* Suppose  $q_{un}^* = 1$ . Then the seller chooses  $P = V + W$  for sure. Given this price, choosing  $I = 1$  yields the buyer a payoff of  $u_B = -C + \sigma_B(V + W + C)$ . A choice for  $I = 0$  yields her  $u_B = 0$ . Now  $q_{un}^* = 1$  necessarily requires that  $-C + \sigma_B(V + W + C) \geq 0$ , i.e.  $\sigma_B \geq \frac{C}{V+W+C}$ . This contradicts with  $\sigma_B < \frac{C}{V+W+C}$ . Hence necessarily  $0 < q_{un}^* < 1$  in this case.

Clearly, the selfish seller with  $\rho_S < \frac{1}{2}$  necessarily chooses between  $P = V$  and  $P = V + W$ . Let  $p \equiv \Pr(P = V)$ . Then choosing  $I = 0$  yields the buyer  $p(\sigma_B \cdot V)$  in expected payoffs, while choosing  $I = 1$  gives her  $p \cdot [(W - C) + \sigma_B \cdot \max\{V - (W - C), 0\} - \rho_B \cdot \max\{(W - C) - V, 0\}] + (1 - p) \cdot [\sigma_B \cdot (V + W + C) - C]$ . From  $0 < q_{un}^* < 1$  it follows that these two payoffs must be equal. This implies that  $p \cdot [(1 - 2\sigma_B) \cdot W - \sigma_B \cdot V - h(\sigma_B, \rho_B)] = [C - \sigma_B \cdot (V + W + C)]$  must hold. From  $(1 - \rho_B) \cdot (W - C) > h(\sigma_B, \rho_B)$  it follows that  $[(1 - 2\sigma_B) \cdot W - \sigma_B \cdot V - h(\sigma_B, \rho_B)] > [C - \sigma_B \cdot (V + W + C)]$ . Hence  $\sigma_B \leq \frac{C}{V+W+C}$  is required to satisfy  $p^* = \frac{[C - \sigma_B \cdot (V + W + C)]}{[(1 - 2\sigma_B) \cdot W - \sigma_B \cdot V - h(\sigma_B, \rho_B)]} \geq 0$ . For  $\sigma_B < \frac{C}{V+W+C}$  necessarily  $0 < p^* < 1$ .

A choice for  $P = V$  yields the seller an expected payoff of  $u_S = (1 - q) \cdot (1 - \rho_S)V + q \cdot (V + \sigma_S \cdot \max\{(W - C) - V, 0\} - \rho_S \cdot \max\{V - (W - C), 0\})$ . A choice for  $P = V + W$  yields him  $u_S = q \cdot [V + W - \rho_S(V + W + C)]$ . The expected

payoffs of the two price choices are equal whenever  $(1 - q) \cdot (1 - \rho_S)V = q \cdot [(1 - 2\rho_S)W + h(\sigma_B, \rho_B)]$ .  $q_{un}^*$  equals the unique value that solves this equality.

*Part (b).* From the proof of part (a) it follows that  $\sigma_B \leq \frac{C}{V+W+C}$  is a necessary requirement for  $0 < q_{un}^* < 1$  to be possible.  $q_{un}^* = 0$  can never occur, hence necessarily  $q_{un}^* = 1$  in this case. From  $\rho_S < \frac{1}{2}$  it immediately follows that  $P^* = V + W$ . When  $\sigma_B > \frac{C}{V+W+C}$  the buyer then indeed prefers investment over no investment.

*Part (c).* From  $q_{un}^* > 0$  it follows that the seller will never choose  $P > \frac{V+W-C}{2}$ , because  $P = \frac{V+W-C}{2}$  yields him strictly more. Suppose  $V < \frac{V+W-C}{2}$ . Then the seller will never choose a price strictly between  $P = V$  and  $P = \frac{V+W-C}{2}$ , again because the latter yields him more. Finally, for  $\rho_S > \frac{1}{2}$  also an equal split of the small pie  $P = \frac{V}{2}$  is always better than any  $P < \frac{V}{2}$ . Taken together it follows that  $P$  is necessarily chosen from  $[\frac{V}{2}, V] \cup \{\frac{V+W-C}{2}\}$ .

First consider any price  $\frac{V}{2} \leq P \leq V$ . When the buyer chooses  $I = 0$ , she obtains  $u_B = (1 - \sigma_B) \cdot (V - P) + \sigma_B \cdot P$ . In case of  $I = 1$  she gets  $(1 - \sigma_B) \cdot (V + W - C - P) + \sigma_B \cdot P$  when  $P > (V + W - C - P)$ . From  $W > C$  this exceeds the payoff after  $I = 0$ . In case  $P \leq (V + W - C - P)$  the payoff after  $I = 1$  becomes  $(1 - \rho_B) \cdot (V + W - C - P) + \rho_B \cdot P$ . Given  $P \geq \frac{V}{2}$  and  $\sigma_B \leq \rho_B$  it follows that also this payoff exceeds the one after  $I = 0$ . Hence, against any price in  $\frac{V}{2} \leq P \leq V$  the buyer's best response is  $q_{un} = 1$ . Next consider the remaining case where  $P = \frac{V+W-C}{2} > V$ . Then choosing  $I = 0$  yields the buyer  $u_B = 0$  while  $I = 1$  yields her  $u_B = \frac{V+W-C}{2} > 0$ . Again the buyer's best response is  $q_{un} = 1$ . The equilibrium strategy of the buyer thus necessarily equals  $q_{un}^* = 1$ . Given this, the seller chooses  $P^* = \frac{V+W-C}{2}$  for sure. *QED*

## A.2 Proof of Theorem 2

In this section we provide the proof of Theorem 2. We again do so by deriving a proposition for each of the two separate information conditions. Theorem 2 immediately follows from these two propositions.

Under unobservable investment the game reduces to a simultaneous move game. For this game the SRE concept of Dufwenberg and Kirchsteiger (2004) differs slightly from the *fairness equilibrium* concept of Rabin (1993). Section 5 in the former paper provides a discussion of the differences. Here we follow the SRE specification, to make the results better comparable to those under observable investment (where the game has an explicit sequential structure).

### A.2.1 Observable investment

The following additional notation is used.  $b_{ij}$  gives the (first order) belief of player  $i$  about the strategy of player  $j$ . For instance,  $b_{SB}$  denotes the seller's belief about the buyer's investment strategy  $q_{obs}$ .  $c_{iji}$  is used to denote the second order beliefs. It reflects the belief of player  $i$  about the belief of player  $j$  about the strategy of player  $i$ . For example,  $c_{SBS} = (c_{SBS}^0, c_{SBS}^1)$  denotes the seller's belief about the buyer's belief about the seller's strategy  $(P_0, P_1)$ . The first and second order beliefs determine the factors  $\kappa_{ij}$  and  $\lambda_{iji}$  in the players' utility functions (2). In an SRE beliefs are necessarily correct:  $b_{SB} = c_{BSB} = q_{obs}^*$  and  $b_{BS} = c_{SBS} = (P_0^*, P_1^*)$ .

We make three observations that apply for  $Y_B \geq 0$ . Together they provide a full characterization of all the SRE. The results for  $Y_B = 0$  are summarized in Proposition 3. The analysis closely resembles the one of Dufwenberg and Kirchsteiger (2004) for the sequential Prisoners' dilemma (their game  $\Gamma_2$ ).

**Observation 1** In every SRE  $P_0^* = V$ .

*Proof.* Consider the contingency in which the buyer chooses  $I = 0$ . The seller then can give the buyer at least 0 and at most  $V$ . The equitable payoff for the buyer at this node thus equals  $\pi_B^e(I = 0) = \frac{1}{2} \cdot [0 + V] = \frac{1}{2}V$ . This implies  $\kappa_{SB}(P_0) = (\frac{V}{2} - P_0)$  for  $P_0 \leq V$  and  $\kappa_{SB}(P_0) = -\frac{V}{2}$  for  $P_0 > V$ . The reciprocity payoffs of the seller are the same for any  $P_0 \geq V$ . Monetary payoffs equal  $P_0$  for  $P_0 \leq V$  and 0 for  $P_0 > V$ . Hence the seller strictly prefers  $P_0 = V$  above  $P_0 > V$ . Thus necessarily  $P_0^* \leq V$  in a SRE.

Suppose  $P_0^* > P_1^*$ . The seller's belief about how much the buyer intends to give him by choosing  $I = 1$  equals  $c_{SBS}^1$ . For  $I = 0$  this is  $c_{SBS}^0$ . Hence  $\lambda_{SBS}(I = 1, c_{SBS}) = \frac{1}{2}(c_{SBS}^1 - c_{SBS}^0)$ . In a SRE beliefs are correct and we have  $c_{SBS}^0 = P_0^*$  and  $c_{SBS}^1 = P_1^*$ . This in turn implies that  $\lambda_{SBS}(I = 1, c_{SBS}) < 0$ . The seller's kindness of choosing  $P_1$  equals  $\kappa_{SB}(P_1) = (\frac{V+W}{2} - P_1)$  for  $0 \leq P_1 \leq V + W$ . On the basis of the reciprocity payoffs the seller thus prefers  $P_1 = V + W$ . This also applies for the monetary payoffs, thus  $P_1^* = V + W$ . This contradicts our assumption that  $P_0^* > P_1^*$ . Hence necessarily  $P_0^* \leq P_1^*$ .

From  $P_0^* \leq P_1^*$  and  $\lambda_{SBS}(I = 0, c_{SBS}) = \frac{1}{2}(c_{SBS}^0 - c_{SBS}^1)$  it immediately follows that  $\lambda_{SBS}(I = 0, c_{SBS}) \leq 0$  in a SRE. In words, no investment is never seen as kind. With  $\kappa_{SB}(P_0) = (\frac{V}{2} - P_0)$  it immediately follows that after  $I = 0$  seller's utility equals  $u_S = P_0 + Y_S \cdot (\frac{V}{2} - P_0) \cdot \frac{1}{2}(c_{SBS}^0 - c_{SBS}^1)$ . This is clearly maximized for  $P_0^* = V$ . *QED*

**Observation 2** In every SRE  $P_1^* = \min\{V + \frac{2}{Y_S}, V + W\}$ .

*Proof.* From Observation 1 we have that  $\lambda_{SBS}(I = 1, c_{SBS}) = \frac{1}{2}(c_{SBS}^1 - V)$  under correct beliefs  $c_{SBS}^0 = P_0^* = V$ . Overall utility after  $I = 1$  thus equals

$u_S = P_1 + Y_S \cdot \left(\frac{V+W}{2} - P_1\right) \cdot \frac{1}{2}(c_{SBS}^1 - V)$  for the relevant range  $0 \leq P_1 \leq V+W$ . We obtain  $\frac{\partial u_S}{\partial P_1} = 1 - \frac{Y_S}{2}(c_{SBS}^1 - V)$ . For  $c_{SBS}^1 < V + \frac{2}{Y_S}$  this is strictly positive, hence  $P_1^* < \min\{V + \frac{2}{Y_S}, V+W\}$  cannot occur. Similarly, for  $c_{SBS}^1 > V + \frac{2}{Y_S}$  the derivative is negative, hence  $P_1^* > \min\{V + \frac{2}{Y_S}, V+W\}$  cannot occur. *QED*

**Observation 3** The SRE is characterized by one of the three following possibilities:

- (a) if  $Y_S < \frac{2+V \cdot Y_B}{W-C}$ , then  $q_{obs}^* = 0$ ;
- (b) if  $Y_B > 0$  and either (i)  $\frac{2+V \cdot Y_B}{W-C} < Y_S < \frac{4}{W}$  and  $Y_S < \frac{2}{W-C} \cdot (1 + Y_B \cdot \frac{V}{2} - Y_B \cdot \left(\frac{W}{2} - \frac{2}{Y_S}\right))$  or (ii)  $\frac{4}{W} < Y_S < \frac{2+V \cdot Y_B}{W-C}$  and  $Y_S > \frac{2}{W-C} \cdot (1 + Y_B \cdot \frac{V}{2} - Y_B \cdot \left(\frac{W}{2} - \frac{2}{Y_S}\right))$ , then  $q_{obs}^* = \frac{Y_S}{Y_B} \cdot \frac{Y_S(W-C) - 2 - VY_B}{(4 - W \cdot Y_S)}$ ;
- (c) if  $Y_S > \frac{2}{W-C} \cdot \left(1 + Y_B \cdot \frac{V}{2} - Y_B \cdot \left(\frac{W}{2} - \frac{2}{Y_S}\right)\right)$ , then  $q_{obs}^* = 1$ .

*Proof.* To understand the buyer's investment motives, we have to determine the seller's kindness towards the buyer *at the root* of the game tree, i.e. before the buyer decides about investment. This kindness of strategy  $(P_0, P_1)$  is denoted  $\kappa_{SB}((P_0, P_1), b_{SB})$ . When the buyer is believed to use investment strategy  $b_{SB}$ , the seller can give the buyer at most  $b_{SB} \cdot (V+W-C) + (1 - b_{SB}) \cdot V$  and at least  $b_{SB} \cdot (-C)$ . The equitable payoff for the buyer at the root thus equals the average  $\pi_B^e(b_{SB}) = \frac{V}{2} + b_{SB} \cdot \left(\frac{W}{2} - C\right)$ . By using the equilibrium strategy  $(P_0^*, P_1^*) = (V, P_1^*)$  the seller gives  $b_{SB} \cdot (V+W-C - P_1^*)$  to the buyer. We therefore have  $\kappa_{SB}((P_0^*, P_1^*), b_{SB}) = b_{SB} \cdot (V+W - P_1^*) - b_{SB} \cdot \frac{W}{2} - \frac{V}{2}$ . Because in a SRE the buyer understands the seller's motivation, it is required that  $\lambda_{BSB}(b_{BS}, c_{BSB}) = \kappa_{SB}((P_0^*, P_1^*), b_{SB})$ . We thus have  $\lambda_{BSB}((P_0^*, P_1^*), c_{BSB}) = c_{BSB} \cdot (V+W - P_1^*) - c_{BSB} \cdot \frac{W}{2} - \frac{V}{2}$  under correct beliefs  $b_{BS} = (P_0^*, P_1^*)$ .

Next we determine the buyer's kindness towards the seller. The buyer can give the seller a material payoff of at least  $V$  by choosing  $I = 0$ , and at most  $P_1^* = V + \min\{\frac{2}{Y_S}, W\}$  by choosing  $I = 1$ . The equitable payoff for the seller thus equals  $\pi_S^e = V + \frac{1}{2} \cdot \min\{\frac{2}{Y_S}, W\}$ . The kindness of strategy  $q$  under correct beliefs  $b_{BS} = (P_0^*, P_1^*) = (V, P_1^*)$  is then  $\kappa_{BS}(q, (P_0^*, P_1^*)) = (q - \frac{1}{2}) \cdot \min\{\frac{2}{Y_S}, W\}$ . Overall, with monetary payoffs  $\pi_B = q \cdot (W - \min\{\frac{2}{Y_S}, W\}) - C$ , we obtain that:

$$u_B = q \cdot \left[ \left( W - \min\left\{ \frac{2}{Y_S}, W \right\} \right) - C \right] + Y_B \cdot \left( q - \frac{1}{2} \right) \cdot \min\left\{ \frac{2}{Y_S}, W \right\} \cdot \left[ c_{BSB} \cdot \left( \left( W - \min\left\{ \frac{2}{Y_S}, W \right\} \right) - \frac{W}{2} \right) - \frac{V}{2} \right]$$

It is easily seen that when  $\min\{\frac{2}{Y_S}, W\} = W$  utility  $u_B$  is strictly decreasing in  $q$ . Hence  $q_{obs}^* = 0$  for  $Y_S < \frac{2}{W}$ . Note that  $\frac{2}{W} < \frac{2+V \cdot Y_B}{W-C}$ .

Next consider the case  $Y_S > \frac{2}{W}$ . Then  $\min\{\frac{2}{Y_S}, W\} = \frac{2}{Y_S}$ . From the above expression we get  $\frac{\partial u_B}{\partial q} = \left( (W - C) - \frac{2}{Y_S} \right) + Y_B \cdot \frac{2}{Y_S} \cdot \left[ c_{BSB} \cdot \left( \frac{W}{2} - \frac{2}{Y_S} \right) - \frac{V}{2} \right]$ . Suppose  $q_{obs}^* = 1$ . Then we must have  $\frac{\partial u_B}{\partial q} \geq 0$  at  $c_{BSB} = 1$ . This requires  $W - C - \frac{2}{Y_S} \geq -Y_B \cdot \frac{2}{Y_S} \left( \frac{W}{2} - \frac{2}{Y_S} - \frac{V}{2} \right)$ , i.e.  $W - C \geq \frac{2}{Y_S} \cdot \left( 1 - Y_B \cdot \left( \frac{W}{2} - \frac{2}{Y_S} - \frac{V}{2} \right) \right)$ . Next suppose  $q_{obs}^* = 0$ . Then  $\frac{\partial u_B}{\partial q} \leq 0$  is required for  $c_{BSB} = 0$ . This comes down to  $\left( (W - C) - \frac{2}{Y_S} \right) \leq Y_B \cdot \frac{V}{Y_S}$ , i.e.  $Y_S \leq \frac{2+Y_B \cdot V}{W-C}$ . This yields parts (a) and (c).

In order to have a mixed equilibrium  $0 < q_{obs}^* < 1$ , necessarily  $\frac{\partial u_B}{\partial q} = 0$  at  $c_{BSB} = q_{obs}^*$ . This implies  $\left( (W - C) - \frac{2}{Y_S} \right) = -Y_B \cdot \frac{2}{Y_S} \cdot \left[ q_{obs}^* \cdot \left( \frac{W}{2} - \frac{2}{Y_S} \right) - \frac{V}{2} \right]$ . When  $Y_B = 0$  this can only hold for the degenerate case  $Y_S = \frac{2}{W-C}$ . In the sequel we do not consider such knife-edge cases. In case  $Y_B > 0$  we obtain:

$$q_{obs}^* = \frac{Y_S^2(W - C) - 2Y_S - VY_S Y_B}{Y_B(4 - W \cdot Y_S)} = \frac{Y_S}{Y_B} \cdot \frac{Y_S(W - C) - 2 - V \cdot Y_B}{(4 - W \cdot Y_S)}$$

The requirement that  $q_{obs}^* > 0$  comes down to either  $\frac{2+V \cdot Y_B}{W-C} < Y_S < \frac{4}{W}$  or  $\frac{4}{W} < Y_S < \frac{2+V \cdot Y_B}{W-C}$ . In the former case requiring  $q_{obs}^* < 1$  equals  $Y_S < \frac{2}{W-C} \cdot \left( 1 - Y_B \cdot \left( \frac{W}{2} - \frac{2}{Y_S} - \frac{V}{2} \right) \right)$ . In the latter case this becomes  $Y_S > \frac{2}{W-C} \cdot \left( 1 - Y_B \cdot \left( \frac{W}{2} - \frac{2}{Y_S} - \frac{V}{2} \right) \right)$ . This gives part (b). *QED*

Observations 1 through 3 characterize all possible SRE. They reveal that the probability of investment  $q_{obs}^*$  is increasing in  $Y_S$ . The more sensitive the seller is to reciprocity, the more the buyer invests. When  $Y_S$  is large enough –  $Y_S > \max\{\frac{4}{W}, \frac{2+Y_B V}{W-C}\}$  is sufficient – the buyer always invests. The seller then reciprocates by choosing  $P_1^* = V + \frac{2}{Y_S}$ . In case  $Y_S$  is low –  $Y_S \leq \min\{\frac{4}{W}, \frac{2+Y_B V}{W-C}\}$  is sufficient – the buyer never invests. Hence for investment to occur, the parameter  $Y_S$  is decisive. Yet  $Y_B$  also plays a (minor) role. This follows because in a SRE the seller always chooses  $P_0^* = V$ , and this is seen by the buyer as an unkind act. She therefore may have an incentive to punish the seller, and can do so by choosing  $q_{obs} = 0$ . Hence negative reciprocity may make no investment more attractive to the buyer. However, because investment is sufficiently rewarded for  $Y_S$  high enough ( $P_1^* = V + \frac{2}{Y_S}$  is then low), the buyer always prefers to invest when the seller is sufficiently reciprocal. Assuming  $Y_B = 0$  therefore yields qualitatively the same results. Proposition 3 below immediately follows from Observations 1 through 3.

**Proposition 3** (Observable investment) Suppose  $Y_B = 0$ . Then the unique SRE is given 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  $P_1^* = V + \frac{2}{Y_S}$ .

### A.2.2 Unobservable investment

The seller's pricing strategy is now given by a probability distribution over  $[0, V + W]$ . First and second order beliefs  $b_{BS}$  and  $c_{SBS}$  are now defined with respect to this strategy. It appears that there are multiple equilibria. In particular,  $q_{un}^* = 1$  can typically be supported by a continuum of pricing strategies of the seller. Yet in any SRE with  $q_{un}^* < 1$  the seller necessarily mixes between the two prices  $P = V$  and  $P = V + W$ . The first observation in this subsection gives a condition on the second order beliefs which structures the subsequent equilibrium analysis.

**Observation 4** In any SRE necessarily  $\lambda_{SBS} \leq \frac{1}{Y_S}$ . Moreover, it holds that:

- (a)  $\lambda_{SBS} = \frac{1}{Y_S} \iff q_{un}^* = 1$ ;
- (b)  $\lambda_{SBS} < \frac{1}{Y_S} \iff 0 < q_{un}^* < 1$ .

*Proof.* When the seller uses a mixed strategy, every price  $P$  within the support of this strategy must yield the same expected payoffs. Consider such a price  $P$ . We first specify the seller's utility belonging to  $P$ . The seller believes that the buyer uses strategy  $b_{SB}$ . He can give the buyer at least  $b_{SB}(-C)$  and at most  $b_{SB}(V + W - C) + (1 - b_{SB})V$ . Hence the equitable payoff for the buyer equals  $\pi_B^e(b_{SB}) = \frac{V}{2} + b_{SB}(\frac{W}{2} - C)$ . By choosing  $P \leq V$  the seller intends to give to the buyer a payoff of  $b_{SB}(V + W - C - P) + (1 - b_{SB})(V - P)$ . Hence the kindness of such a choice equals  $\kappa_{SB}(P \leq V, b_{SB}) = (\frac{V}{2} - P) + b_{SB} \cdot \frac{W}{2}$ . Similarly, by choosing  $P > V$  the seller intends to give to the buyer a payoff of  $b_{SB}(V + W - C - P)$ . Hence for these values of  $P$  we have  $\kappa_{SB}(P > V, b_{SB}) = b_{SB}(V - P) - \frac{V}{2} + b_{SB} \cdot \frac{W}{2}$ . The overall utility of the seller equals:

$$\begin{aligned} u_S &= P + Y_S \cdot \lambda_{SBS} \cdot \left[ \left( \frac{V}{2} - P \right) + b_{SB} \cdot \frac{W}{2} \right] \text{ when } P \leq V \\ &= b_{SB} \cdot P + Y_S \cdot \lambda_{SBS} \cdot \left[ -\frac{V}{2} - b_{SB} \cdot (P - V) + b_{SB} \cdot \frac{W}{2} \right] \text{ when } P > V \end{aligned}$$

First suppose  $\lambda_{SBS} > \frac{1}{Y_S}$ . 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  $\lambda_{SBS} = 0$ , a contradiction. Hence necessarily  $\lambda_{SBS} \leq \frac{1}{Y_S}$ .

Next consider the case  $\lambda_{SBS} = \frac{1}{Y_S}$ . Then  $u_S = \frac{V}{2} + b_{SB} \cdot \frac{W}{2}$  when  $P \leq V$  and  $u_S = (b_{SB} - \frac{1}{2}) \cdot V + b_{SB} \cdot \frac{W}{2}$  when  $P > V$ . Suppose  $b_{SB} < 1$ . Then the seller always prefers  $P \leq V$  over  $P > V$ . Knowing that  $P \leq V$ , the buyer chooses  $q_{un} = 1$  (for  $P \leq V$  the buyer cannot be kind or unkind to the seller). The latter contradicts  $b_{SB} < 1$  under correct equilibrium beliefs  $b_{SB} = q_{un}^*$ . Hence necessarily  $b_{SB} = q_{un}^* = 1$ . In sum, we have the implication  $\lambda_{SBS} = \frac{1}{Y_S} \implies q_{un}^* = 1$ .

Finally, consider the case where  $\lambda_{SBS} < \frac{1}{Y_S}$ . Suppose  $q_{un}^* = 1$ . Then under correct beliefs  $b_{SB} = q_{un}^*$  we have  $u_S = P + Y_S \cdot \lambda_{SBS} \cdot [(\frac{V}{2} - P) + \frac{W}{2}]$  for all  $P$ . From this we obtain  $\frac{\partial u_S}{\partial P} = 1 - Y_S \cdot \lambda_{SBS} > 0$ . The seller thus wants to choose  $P = V + W$  for sure. On the basis of monetary payoffs the buyer then strictly prefers  $q_{un} = 0$ . A choice for  $P = V + W$  by the seller is seen as unkind, so the buyer also prefers  $q_{un} = 0$  the basis of the reciprocity payoffs. This contradicts  $q_{un}^* = 1$ . Next, suppose  $q_{un}^* = 0$ . Then under correct beliefs  $b_{SB} = q_{un}^*$  it holds that  $u_S$  is maximized for  $P = V$ , yielding  $u_S = (1 - Y_S \cdot \lambda_{SBS}) \cdot \frac{V}{2}$ . But when  $P = V$ , the buyer strictly prefers  $q_{un} = 1$ . Again we obtain a contradiction. We obtain the implication  $\lambda_{SBS} < \frac{1}{Y_S} \implies 0 < q_{un}^* < 1$ .

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

**Observation 5** Let  $P_l \equiv E[P \mid P \leq V]$  and  $P_h \equiv E[P \mid P > V]$ . A SRE with  $q_{un}^* = 1$  exists if and only if:

- (i)  $Y_S \geq \frac{2}{P_h}$  and
- (ii)  $Y_S \geq \frac{2}{W-C} \cdot \left[ \frac{P_h - V}{P_h} - Y_B \cdot \left( \frac{V+W-2P_l}{2} - \frac{2(P_h - P_l)}{Y_S \cdot P_h} \right) \right]$

*Proof.* Let  $\tilde{p} \equiv \Pr(P \leq V)$  denote the probability that the seller chooses a price below  $V$ . Under correct (equilibrium) beliefs  $P_l$ ,  $P_h$  and  $\tilde{p}$  can be used to denote the first and second order beliefs. By choosing  $I = 0$  the buyer believes to give the seller a monetary payoff of  $\tilde{p} \cdot P_l$ , while for  $I = 1$  this amounts to  $p \cdot P_l + (1 - \tilde{p}) \cdot P_h$ . The equitable payoff for the seller under correct beliefs is thus  $\pi_S^e = \tilde{p} \cdot P_l + \frac{1}{2}(1 - \tilde{p})P_h$ . The kindness of strategy  $q$  at the correct beliefs is then  $\kappa_{BS}(q) = q \cdot [\tilde{p} \cdot P_l + (1 - \tilde{p}) \cdot P_h] + (1 - q) \cdot \tilde{p} \cdot P_l - \pi_S^e = (q - \frac{1}{2})(1 - \tilde{p}) \cdot P_h$ .

From Observation 4 we know that when  $q_{un}^* = 1 \iff \lambda_{SBS} = \frac{1}{Y_S}$ . Hence necessarily  $\lambda_{SBS} = \frac{1}{Y_S}$ . Correct second order beliefs require  $\lambda_{SBS} = \kappa_{BS}$ , and

thus  $(1 - \tilde{p}) \cdot P_h = \frac{2}{Y_S}$ . Hence in equilibrium  $\tilde{p}^* = 1 - \frac{2}{Y_S \cdot P_h}$ . This requires  $Y_S \geq \frac{2}{P_h}$ .

We next determine  $\lambda_{BSB}$ . At the correct belief that  $q_{un} = 1$  we have  $\lambda_{BSB} = (V + W - C - \tilde{p} \cdot P_l - (1 - \tilde{p}) \cdot P_h) - \frac{1}{2}[(V + W - C) + (-C)] = \frac{V+W}{2} - \tilde{p} \cdot P_l - (1 - \tilde{p}) \cdot P_h$ . Taken together, we obtain:

$$u_B = \tilde{p} \cdot (V - P_l) + q \cdot [(1 - \tilde{p})V + W - (1 - \tilde{p})P_h - C] \\ + Y_B \cdot \left( q - \frac{1}{2} \right) \frac{2}{Y_S} \cdot \left( \frac{V + W}{2} - \tilde{p} \cdot P_l - (1 - \tilde{p}) \cdot P_h \right)$$

Hence  $\frac{\partial u_B}{\partial q} \geq 0$  reduces to  $(1 - \tilde{p})V + W - (1 - \tilde{p})P_h - C \geq -Y_B \cdot \frac{2}{Y_S} \cdot \left( \frac{V+W}{2} - \tilde{p} \cdot P_l - (1 - \tilde{p}) \cdot P_h \right)$ . Substituting  $\tilde{p} = 1 - \frac{2}{Y_S \cdot P_h}$  and rewriting yields that  $Y_S \geq \frac{2}{W-C} \cdot \left[ \frac{P_h - V}{P_h} - Y_B \cdot \left( \frac{V+W-2P_l}{2} - \frac{2(P_h - P_l)}{Y_S \cdot P_h} \right) \right]$  is required. This proves the ‘only if’ part.

To prove the ‘if’ part, suppose  $Y_S$  satisfies the two stated inequalities. Then from the above it follows that for  $\tilde{p}^* = 1 - \frac{2}{Y_S \cdot P_h}$  the buyer’s strategy  $q_{un}^* = 1$  is a best response. What remains to be shown is that  $\tilde{p}^* = 1 - \frac{2}{Y_S \cdot P_h}$  is a best response against  $q_{un}^* = 1$ . From the proof of Observation 4 it follows that under correct beliefs  $b_{SB} = 1$  (and  $\lambda_{BSB} = \frac{1}{Y_S}$ ) we have that  $u_S = \frac{V+W}{2}$ . The seller’s utility is independent of his pricing strategy, hence any strategy is a best response. *QED*

**Observation 6** In any SRE with  $0 < q_{un}^* < 1$  the seller necessarily *strictly* mixes between  $P = V$  and  $P = V + W$  only.

*Proof.* From Observation 4 we know that when  $0 < q_{un}^* < 1$  we necessarily have that  $\lambda_{BSB} < \frac{1}{Y_S}$ . From the expression for  $u_S$  in the proof of Observation 4 we obtain  $\frac{\partial u_S}{\partial P} > 0$  for all  $P \neq V$ . The seller therefore only chooses between  $P = V$  and  $P = V + W$ . Let  $p \equiv \Pr(P = V)$ . We next show that  $0 < p^* < 1$ .

As before, let  $b_{BS}$  ( $c_{BSB}$ ) denote the first (second) order belief about the seller’s strategy  $p$ . In the proof of Observation 5 we derived  $\kappa_{BS}(q) = \left( q - \frac{1}{2} \right) \cdot (1 - \tilde{p}) \cdot P_h$ . By making the appropriate substitutions  $P_h = V + W$  and  $\tilde{p} = b_{BS}$  we obtain  $\kappa_{BS}(q, b_{BS}) = \left( q - \frac{1}{2} \right) [(1 - b_{BS}) \cdot (V + W)]$ . From the proof of Observation 4 we have  $\pi_B^e(b_{SB}) = \frac{V}{2} + b_{SB} \left( \frac{W}{2} - C \right)$ . Hence  $\kappa_{SB}(p, b_{SB}) = p \cdot [b_{SB} \cdot (W - C)] + (1 - p) \cdot [b_{SB}(-C)] - \pi_B^e(b_{SB}) = \left( p - \frac{1}{2} \right) \cdot b_{SB} \cdot W - \frac{V}{2}$ .

The second order kindness beliefs can now most easily be obtained from the kindness functions by moving one level up in the belief hierarchy, i.e. in  $\kappa_{SB}(p, b_{SB})$  strategy  $p$  becomes  $b_{BS}$  and  $b_{SB}$  becomes  $c_{BSB}$  to obtain  $\lambda_{BSB}$ .

Overall we obtain for the utilities:

$$u_B = q(b_{BS}W - C) + Y_B \cdot \left(q - \frac{1}{2}\right) [(1 - b_{BS})(V + W)] \cdot \left[\left(b_{BS} - \frac{1}{2}\right) [c_{BSB} \cdot W] - \frac{V}{2}\right]$$

$$u_S = pV + (1 - p)b_{SB}(V + W) + Y_S \cdot \left[\left(p - \frac{1}{2}\right) [b_{SB} \cdot W] - \frac{V}{2}\right] \cdot \left(b_{SB} - \frac{1}{2}\right) [(1 - c_{SBS})(V + W)]$$

First suppose  $p^* = 0$ . Then under correct beliefs  $b_{BS} = p^* = 0$  we have  $\frac{\partial u_B}{\partial q} = -C + Y_B \cdot (V + W) \cdot \left(-\frac{1}{2} [c_{BSB} \cdot W] - \frac{V}{2}\right) < 0$ . Hence the unique best response is  $q_{un}^* = 0$ , contradicting  $0 < q_{un}^* < 1$ . Next assume  $p^* = 1$ . Then  $\frac{\partial u_B}{\partial q} = (W - C)$  under correct beliefs  $b_{BS} = 1$ . The best response for the buyer is  $q_{un}^* = 1$ , again contradicting  $0 < q_{un}^* < 1$ . Hence necessarily  $0 < p^* < 1$ . *QED*

**Observation 7** Consider the SRE with  $0 < q_{un}^* < 1$ . Let  $p \equiv \Pr(P = V)$ .

(a) If  $Y_S = 0$ , then  $p^*$  follows from solving

$$p^* = \frac{C}{W} - Y_B \cdot [(1 - p^*)] (V + W) \cdot \left[\left(p^* - \frac{1}{2}\right) \left[\frac{V \cdot W}{V + W}\right] - \frac{V}{2}\right];$$

(b) If  $Y_S > 0$ , then  $p^* = 1 + \frac{V - q_{un}^*(V + W)}{Y_S \cdot q_{un}^* W (q_{un}^* - \frac{1}{2})(V + W)}$ .

*Proof.* From Observation 6 we know that necessarily  $0 < p^* < 1$ . This implies that  $\frac{\partial u_S}{\partial p} = 0$  at  $p^*$ . For  $Y_S = 0$  we have  $\frac{\partial u_S}{\partial p} = V - b_{SB}(V + W) = 0$ . Hence  $b_{SB} = q_{un}^* = \frac{V}{V + W}$  is required. The latter implies that  $\frac{\partial u_B}{\partial q} = 0$  at  $q_{un}^* = \frac{V}{V + W}$ . We thus must have that  $\frac{\partial u_B}{\partial q} = b_{BS}W - C + Y_B \cdot [(1 - b_{BS})(V + W)] \cdot \left[\left(b_{BS} - \frac{1}{2}\right) \cdot c_{BSB} \cdot W - \frac{V}{2}\right] = 0$ . In equilibrium beliefs are correct, i.e.  $b_{BS} = p^*$ . Hence  $p^*W - C = -Y_B \cdot [(1 - p^*)] (V + W) \cdot \left[\left(p^* - \frac{1}{2}\right) \left[\frac{V \cdot W}{V + W}\right] - \frac{V}{2}\right]$ . This yields part (a).

When  $Y_S > 0$   $\frac{\partial u_S}{\partial p} = 0$  reduces to  $1 - c_{SBS} = \frac{-[V - b_{SB}(V + W)]}{Y_S \cdot [b_{SB} \cdot W] \cdot (b_{SB} - \frac{1}{2})(V + W)}$ . In equilibrium first and second order beliefs are correct:  $c_{SBS} = p^*$  and  $b_{SB} = q_{un}^*$ . This gives the formula for  $p^*$  in part (b). *QED*

**Observation 8** Consider the SRE with  $0 < q_{un}^* < 1$ .

(a) If  $Y_B = 0$ , then  $q_{un}^*$  follows from solving

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

If  $V < W$ , necessarily a solution exists for which  $q_{un}^* < \frac{V}{V+W}$ ;

(b) If  $Y_B > 0$ , then  $q_{un}^* = \frac{-[p^*W - C] + Y_B \cdot [(1-p^*)(V+W)] \cdot \frac{V}{2}}{Y_B \cdot [(1-p^*)(V+W)] \cdot (p^* - \frac{1}{2})[W]}$ .

*Proof.* From  $0 < q_{un}^* < 1$  necessarily  $\frac{\partial u_B}{\partial q} = 0$  at  $q_{un}^*$ . For  $Y_B = 0$  we have  $\frac{\partial u_B}{\partial q} = b_{BS}W - C$ . Hence  $b_{BS} = p^* = \frac{C}{W}$  is required. The latter implies  $\frac{\partial u_S}{\partial p} = 0$  at  $p^* = \frac{C}{W}$ . Now  $\frac{\partial u_S}{\partial p} = V - b_{SB}(V+W) + Y_S \cdot [b_{SB} \cdot W] \cdot (b_{SB} - \frac{1}{2}) [(1 - c_{SBS})(V+W)]$ . For  $c_{SBS} = \frac{C}{W}$  and  $b_{SB} = q$  we directly obtain the equality  $\frac{V}{V+W} = q \cdot [1 - Y_S \cdot (q - \frac{1}{2})(W - C)] \equiv h(q)$ . Note that  $h(q)$  is continuous with  $h(0) = 0$  and  $h(\frac{1}{2}) = \frac{1}{2}$ . Suppose  $V < W$ . Then  $\frac{V}{V+W} < \frac{1}{2}$ . By the intermediate value theorem then necessarily a  $q_{un}^* < \frac{1}{2}$  exists for which  $h(q_{un}^*) = \frac{V}{V+W}$ . For this  $q_{un}^*$  the term within square brackets exceeds 1. Hence necessarily  $q_{un}^* < \frac{V}{V+W}$  for this solution. This yields the first part.

When  $Y_B > 0$  a zero derivative  $\frac{\partial u_B}{\partial q} = 0$  implies

$$b_{BS}W - C = -Y_B \cdot [(1 - b_{BS})(V+W)] \cdot \left[ \left( b_{BS} - \frac{1}{2} \right) c_{SBS}W - \frac{V}{2} \right]$$

With correct beliefs in equilibrium the expression for  $q_{un}^*$  in part (b) follows. *QED*

Observations 4 through 8 characterize all possible SRE for the unobservable investment case. From Observation 5 follows that for  $q_{un}^* = 1$  to occur, parameter  $Y_S$  is decisive. Only when the seller is sufficiently motivated by reciprocity, an equilibrium exists in which the buyer invests for sure. If  $Y_S$  is low, the buyer necessarily mixes between investing and not investing. Assuming  $Y_B = 0$  (rather than  $Y_B \geq 0$ ), such that the buyer is not reciprocal at all, therefore leads to qualitatively the same results. The following proposition considers this simpler case.

**Proposition 4** (Unobservable investment) Suppose  $Y_B = 0$ . Then the SRE can be characterized as follows:

(a) if  $Y_S < \frac{2}{V+W-C}$ , then the SRE is unique with  $p^* = \frac{C}{W}$  and  $q_{un}^*$  the solution of  $\frac{V}{V+W} = q \cdot [1 - Y_S \cdot (q - \frac{1}{2})(W - C)]$ . Necessarily  $0 < q_{un}^* < 1$ , and for  $V < W$  necessarily  $0 < q_{un}^* < \frac{V}{V+W}$ ;

- (b) if  $Y_S > \frac{2}{V+W-C}$ , then a (continuum of) SRE exists with  $q_{un}^* = 1$ . The equilibrium specified in part (a) may exist at the same time. This is certainly the case when  $V < W$ .

*Proof.* Observation 5 gives the condition under which an equilibrium with  $q_{un}^* = 1$  exists. For  $Y_B = 0$  this reduces to  $Y_S > \max\{\frac{2}{W-C} \cdot \frac{P_h - V}{P_h}, \frac{2}{P_h}\}$ . The first argument in the max-term is increasing in  $P_h$ , the second argument decreasing. They are equal for  $P_h = V + W - C$ . Hence  $Y_S > \frac{2}{V+W-C}$  is the minimum requirement. Any pricing strategy with  $P_h \equiv E[P | P > V] = V + W - C$  and  $P_l \equiv E[P | P \leq V]$  arbitrary then supports  $q_{un}^* = 1$ .

Next, let  $h(q) = q \cdot [1 - Y_S \cdot (q - \frac{1}{2})(W - C)]$ . Note that  $h(q)$  is strictly concave in  $q$ . From  $h(0) = 0$  and  $h(\cdot)$  increasing for low  $q$  it follows that  $h(1) > \frac{V}{V+W}$  guarantees existence and uniqueness. Rewriting this we obtain  $Y_S < \frac{2}{V+W} \cdot \frac{W}{W-C}$ . From  $\frac{2}{V+W-C} < \frac{2}{V+W} \cdot \frac{W}{W-C}$  we then obtain that when  $Y_S < \frac{2}{V+W-C}$  the SRE is unique. Together with Observations 7 and 8 the proposition follows. *QED*

### A.3 Summary of the instructions

Besides the on-line instructions subjects received a summary of these instructions on paper. Below a direct translation of this summary sheet is given. This summary belongs to the unobservable investment case. The summary sheet for the observable investment treatment is similar.

**Summary of the instructions** This experiment consists of 36 rounds. At the start of each round the participants are paired in couples. The pairing scheme was already determined before the start of the experiment.

The division into couples is chosen such that it is impossible that you are paired with the same other participant in two consecutive rounds. It also holds that within each of the six consecutive blocks of six rounds –viz. rounds 1 up to 6, rounds 7 up to 12, rounds 13 up to 18, rounds 19 up to 24, rounds 25 up to 30 and rounds 31 up to 36– you will never be paired with the same other participant in more than one round. When you will meet the same participant again is unpredictable. With whom you are paired within a particular round is always kept secret from you.

One of the participants within a pair has role A, the other has role B. Within a round you will keep the same role. What exactly your role is, you will hear at the beginning of each round. Over the rounds your role varies. This variation is chosen such that you will be assigned the role of A in exactly half of the total number of rounds, and the role of B in the other half. It

also holds that within each block of six rounds, you are assigned the role of A three times and the role of B also three times.

Each of the 36 rounds consists of 2 stages. In stage 1 only the participant with role A takes a decision. In stage 2 the participant with role B does so. The two stages take the following form:

- 1 Participant A within a pair chooses between X and Y. A choice for X is free, a choice for Y is costly to (only) participant A. The costs of choosing Y depend on the block to which the round belongs:

rounds 1 up to 6 and 19 up to 24:	20 points
rounds 7 up to 12 and 25 up to 30:	40 points
rounds 13 up to 18 and 31 up to 36:	60 points

Participant B observes A's choice in stage 1 only **after** stage 2 has ended.

- 2 Participant B chooses the number of points, an integer between 0 and 130, s/he wants to keep for him/herself. After B made his/her choice, it is checked whether the number of points demanded can be assigned. In case A chose option X in stage 1, there are 50 points available. When A chose option Y there are 130 points available. In case the number of points demanded by B exceeds the number of points available, the available points are lost and both participants obtain nothing. Otherwise, the available points are divided according to B's demand. Apart from that, when participant A had chosen option Y, the costs of this choice has to be subtracted from his/her amount to arrive at the number of points s/he earns in this round.

At the start of experiment you will get 75 points for free. Before the start of the experiment 6 payment rounds were selected from the overall 36 rounds. From each block of six rounds, one round was selected. For every participant the same six rounds were selected. Your total number of points equals the sum of your initial amount and the number of points you obtained in the six selected payment rounds.

At end of the experiment you will be paid in euros, based on the total number of points you earned. The conversion rate is such that 10 POINTS in the experiment correspond to 1 EURO in money.

## A.4 Tests on aggregation of data

In this appendix we report the test results on session (subject pool) effects, order effects and learning effects. These tests reveal that we can pool the data from sessions that are similar and also of those that only differ in the order of the  $C$ -levels. Yet we do find some significant learning effects.

### A.4.1 Comparing similar sessions

For both the unobservable and the observable investment case we have two sessions which use the same (upward) ordering. With only two group observations per session, no meaningful comparisons can be made at the matching group level. This holds because the smallest significance level that a two-tailed ranksum test can attain equals  $\frac{1}{3}$ . We therefore only look at tests performed at the level of individual means. Table 7 reports the results. None of the tests reaches significance at the 5%-level. The lowest  $p$ -level equals .0725, all others are well above .1. We thus conclude that we can pool the data from sessions that are similar.

### A.4.2 Order effects

Apart from two sessions with the ‘upward’ ordering, we have for each observability case one session with the ‘downward’ ordering. For each level of investment costs we compare, across orderings, the blocks in which these costs are used for the first time (rounds 1-18), and the blocks in which they are used for the second time (rounds 19-36). Again, with only two group observations for the ‘downward’ ordering and four for the ‘upward’ ordering, no meaningful comparisons can be made using group level data. We therefore confine ourselves to comparing individual means.

For the unobservable investment case we find only one significant difference, i.e. when subjects are confronted with  $C = 40$  for the second time (cf. Table 8). Here investment rates are significantly lower in the ‘upward’ ordering. Yet for demand behavior no significant differences are found. We conclude that under unobservable investment order effects are almost absent.

When the investment is observable we find five (out of 18) significant differences. Investment rates are significantly lower in the ‘upward’ ordering when subjects are confronted with  $C = 60$  for the second time. The remaining four significant differences concern demand behavior. In all these cases sellers on average demand significantly more under the ‘upward’ ordering. Two out of these four cases concern demands after  $I = 0$ . Although significant, here differences between mean demands are very small in magnitude.

Table 7: Test results for equality of similar sessions

	first: rounds 1-18			second: rounds 19-36		
	$C = 20$	$C = 40$	$C = 60$	$C = 20$	$C = 40$	$C = 60$
<b>Unobservable</b>						
investment	.3113	.4645	.4815	.2133	.9090	.5897
demand	.4622	.1151	.2922	.7014	.3936	.5891
<b>Observable</b>						
investment	.6039	.7789	.1794	.1961	.0725	.3173
demand: $I = 0$	.8454 (19)	.6388 (33)	.5405	.4793 (35)	.9537	.3173
demand: $I = 1$	.6829	.2005 (35)	.7920 (9)	.4641 (37)	.2419 (14)	n.a.

*Remark:* In each cell the  $p$ -value is given of a ranksum test using individual means. Within parentheses appear the number of observations ( $m = n$ ) when this number deviates from  $m = n = 40$ .

Sellers namely almost always demand the complete pie ( $P = 50$ ).

Overall we conclude that some order effects can be detected in the data, especially in the observable investment case. But they are typically only minor. This provides sufficient justification for pooling the results from the two different orderings.

### **A.4.3 Learning effects**

In both the ‘upward’ and the ‘downward’ ordering every cost level was represented in one block of six rounds in the first 18 rounds and in a second block in the last 18 rounds. To test for learning effects we compare the first block with the second block by means of signrank tests, for all three  $C$ -levels separately. For the unobservable investment case only three (out of 12) significant differences are found (cf. Table 9). When  $C = 20$  both the individual and the group level data indicate that demands significantly increase over time. Mean individual investment levels suggest that for  $C = 40$  investment rates decrease over time, yet no significant differences are found when we look at the group level data.

In the observable investment case we observe nine (out of 17) significant differences. There is especially strong evidence that investment rates are decreasing over time. Moreover, mean demands tend to increase over time, although differences are not always significant. Taken together these results imply that, especially when the investment is observable, behavior evolves over time. In the main text we therefore consider the first 18 rounds separately from the last 18 rounds.

Table 8: Test results for equality of different orderings

	first: rounds 1-18			second: rounds 19-36		
	$C = 20$	$C = 40$	$C = 60$	$C = 20$	$C = 40$	$C = 60$
<b>Unobservable</b>						
investment	1.000	.6261	.8809	.3723	.0318	.2986
demand	.7819	.2933	.0893	.5350	.6549	.0855
<b>Observable</b>						
investment	.3053	.5515	.0894	.1412	.5793	.0073
demand: $I = 0$	.1128 (34)	.0810 (51)	.0023 (59)	.1273 (50)	.5570	.0185
demand: $I = 1$	.9685	.0227 (51)	.8552 (18)	.0220 (57)	.1034 (25)	.1857 (12)

*Remark:* In each cell the  $p$ -value is given of a ranksum test using individual means. Within parentheses appear the number of observations ( $m = n = 60$ ) when this number deviates from  $m = n = 60$ .

Table 9: Test results on comparing first half with second half

	$C = 20$	$C = 40$	$C = 60$
Unobservable			
investment	.1118	.0148	.3357
	.2008	.1148	.7489
demand	.0032	.0604	.6261
	.0464	.3454	.9165
Observable			
investment	.0001	.0000	.0947
	.0350	.0350	.0509
demand: $I = 0$	.6315 (29)	.4035 (51)	.0020 (59)
	.8292	.6002	.0350
demand: $I = 1$	.0000 (57)	.1058 (23)	.2914 (9)
	.0277	.0431 (5)	n.a.

*Remark:* In each cell the upper (lower)  $p$ -value is given of a signrank test using individual (group) means. Within parentheses appear the number of matched pairs when this number deviates from 60 for individual data and from 6 for group level data.

Figure 1a. The experimental game in the observable investment case

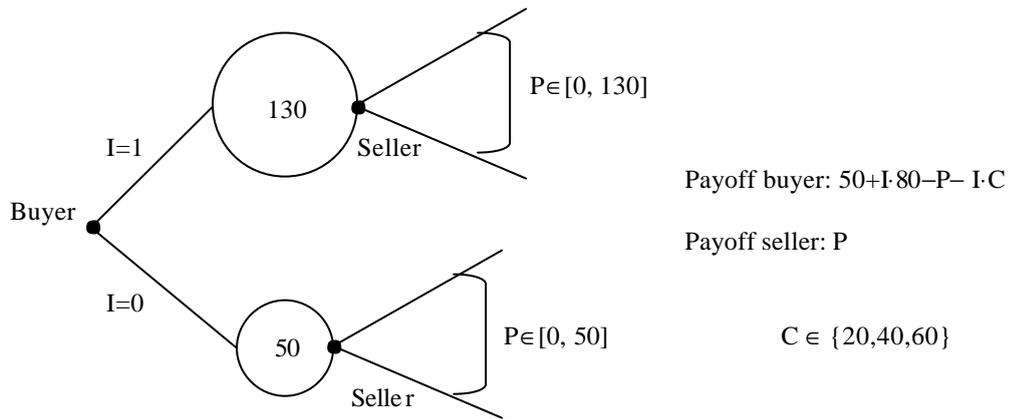


Figure 1b. The experimental game in the unobservable investment case

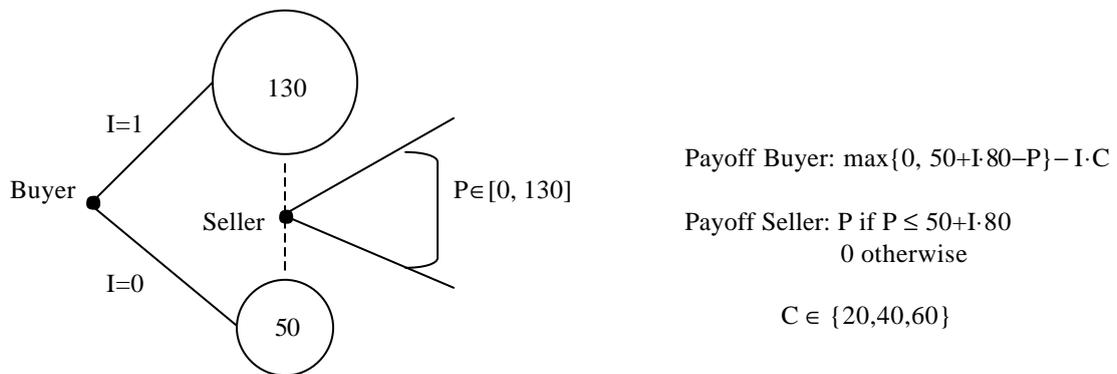


Figure 2: Frequency distribution of demands in the observable investment case (by costs level)

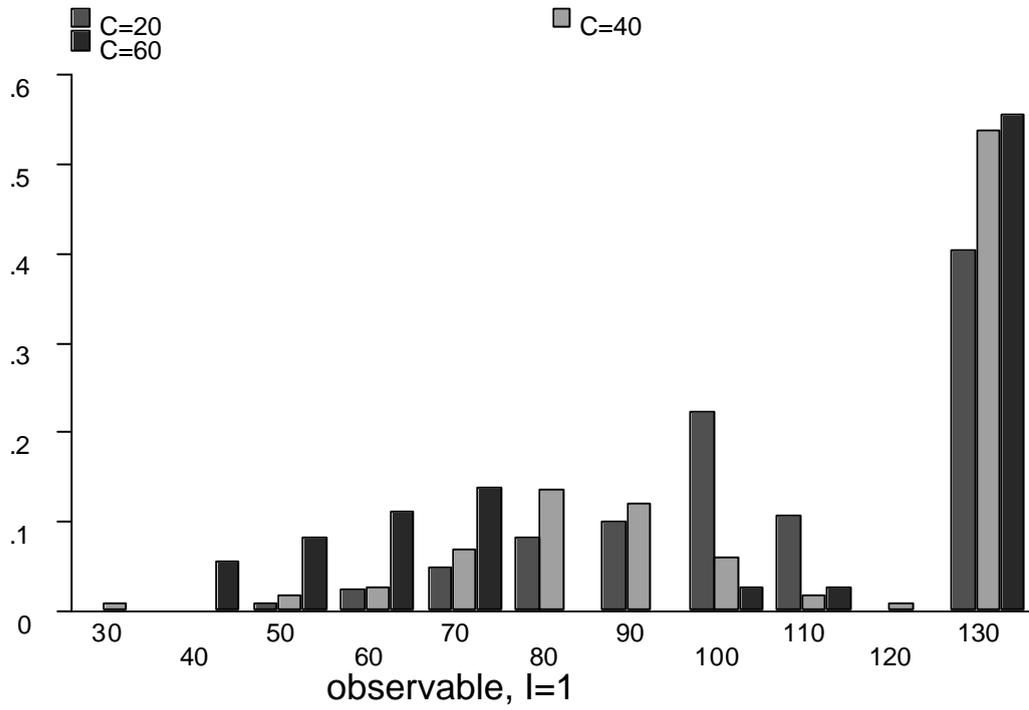
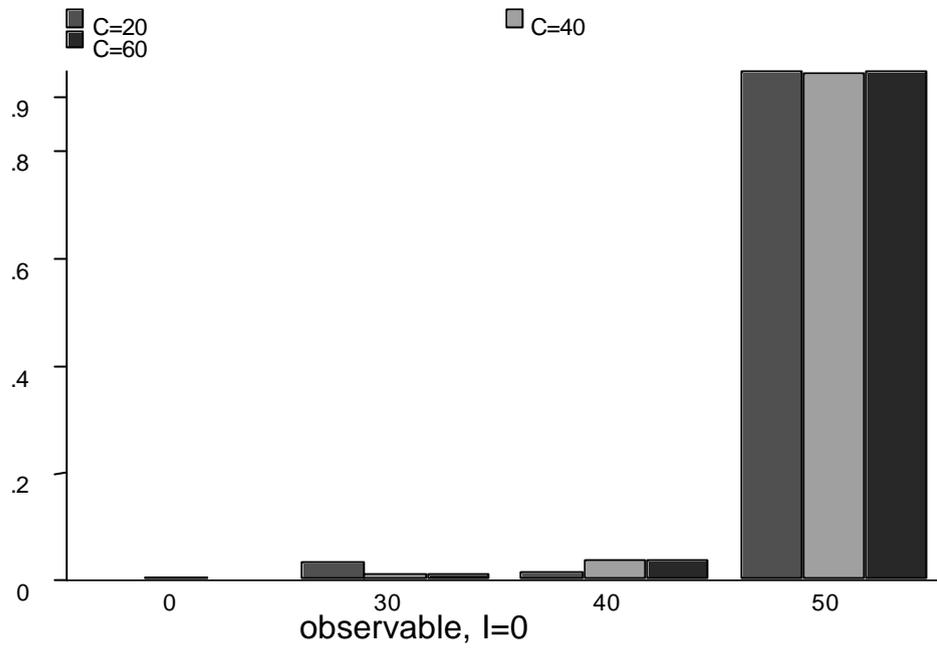


Figure 3: Frequency distribution of demands in the unobservable investment case (by costs level)

