Information Control in the Hold-up Problem*

Anh Nguyen† Teck Yong Tan‡

September 2017

Abstract

We study optimal information control in the hold-up problem with binary investment. A signal structure, which is publicly determined before investment, generates signals about the investment. We characterize the set of investment probability and social welfare that can be achieved in equilibrium, and the signal structure that implements them. The optimal signal structure is generically unique and takes a simple form which eliminates ex-post inefficiency arising from trade breakdown. Contrary to results suggested in the existing literature, there is *no* tradeoff between creating ex-ante investment incentive and eliminating ex-post inefficiency.

Keywords: Hold-up; Information control

JEL Classification: D21, D82, L15

*We thank Navin Kartik and Ran Spiegler for very helpful comments.
†Department of Economics, Columbia University. Email: anh.nguyen@columbia.edu.
‡Department of Economics, Nanyang Technological University. Email: teckyongtan@ntu.edu.sg.

1 Introduction

We revisit the classic hold-up problem with binary investment. Framing it as a bilateral trade, a buyer (he) can make a binary investment which increases his valuation for a good sold by a monopolist seller (she). A hold-up problem arises when the investment is not contractible: anticipating that the seller will charge a high price upon investment and expropriate all the gains, the buyer never invests.

A number of papers have since suggested to mitigate the hold-up risk by hiding the investment from the seller. The asymmetric information limits the seller's ability to extract the investment gains, thus improving *ex-ante efficiency* by restoring (at least partially) the buyer's investment incentive. However, the asymmetric information also creates the possibility of trade breaking down, thus resulting in *ex-post inefficiency.* Gibbons (1992) and Gul (2001) show that these two effects exactly cancel out when the investment outcome is completely unobservable to the seller. Lau (2008) shows that the two effects change at different rates when the probability of the seller observing the investment outcome varies; hence overall welfare increases when the seller observes the investment outcome with a probability strictly between 0 and 1. These results implicitly suggest that there is a tradeoff between creating ex-ante investment incentive and minimizing ex-post inefficiency, and welfare is maximized by balancing the two effects.

In this paper, we study optimal information control in the hold-up problem. A signal structure, which is publicly determined before investment, generates signals about the buyer's investment; the seller receives no other information besides these signals. We show that contrary to results suggested in the existing literature, there is *no* tradeoff between creating ex-ante investment incentive and eliminating ex-post inefficiency in the optimal information transmission design.

Intuitively, ex-ante investment incentive is created when the seller is unaware of the buyer's investment at least some of the time. This only concerns hiding information from the seller in the "investment state". On the other hand, ex-post inefficiency is eliminated by revealing the buyer's lack of investment so that the seller does not set the high price when the buyer's valuation is low. This only concerns information about the "non-investment state" which does not affect the ex-ante investment incentive. In turn, this separation implies that ex-post inefficiency can be eliminated without compromising the ex-ante investment incentive.

We begin the analysis by showing that when the seller cannot perfectly observe the investment, the buyer's investment decision must be mixed in equilibrium. This is because if the seller anticipates that the buyer always invests, she will charge a high price which destroys the buyer's ex-ante investment incentive. On the other hand, the seller will charge a low price if she anticipates that the buyer never invests; in turn, the buyer will want to invest for his own gains.

We then characterize the set of possible investment probability and social welfare that can be sustained in equilibrium, and the signal structure that implements them. We show that every implementable investment probability is optimally implemented by the same signal structure which is unique within an appropriate class of signal structures. This optimal signal structure also takes a simple form that can be easily replicated by practical arrangements. This addresses the usual concern in the information control literature about how one derives the ability to commit to a signal structure.

Under the optimal signal structure, the seller is perfectly certain that the buyer has invested when she sets the high price. Therefore, trade always occurs and ex-post inefficiency is zero. On the other hand, she is only sufficiently confident but not perfectly certain that the buyer has not invested when she sets the low price; this inability to perfectly detect investment is what creates the ex-ante investment incentive for the buyer. However, the need to maintain credibility for this "low signal" to the seller creates an upper bound on the buyer's probability of investment in equilibrium, thus prohibiting first-best welfare from being achieved. Since there is no ex-post inefficiency, these results effectively isolate the "true" hold-up effect due to the non-contractibility of the investment and provide a bound on the investment frequency and welfare attainable through information control.

Related Literature. This paper is primarily related to the literature on the use of asymmetric information to mitigate hold-up risks. Gul (2001) shows that the first best can be achieved in the limit if the investment is completely unobservable to the seller who can then make repeated and frequent offers after rejection by the buyer. Lau (2008) shows that welfare can be increased relative to the hold-up case if the seller observes the investment outcome with an intermediate probability; she then shows that conditional on such information transmission arrangement, welfare is always improved by allowing the seller to make repeated and frequent offers after rejection. Our results contrast Lau's results in two ways. First, we show that information control in the form of randomizing between perfect observability and perfect unobservability, as considered by Lau, is never optimal. Second, we show that conditional on the optimal information structure, allowing for repeated offers has no effect since the seller's offer will be accepted immediately.[1] But our paper also complements Lau's results in that it illustrates why repeated offers in her setup *always* improves welfare – repeated offers eliminate ex-post inefficiency due to no trade which occurs only in the "no-investment states", and our results indirectly point out that allowing for renegotiation at these states has no detrimental effect on the buyer's ex-ante investment incentive.

Other papers that study asymmetric information in the hold-up problem include González (2004), Hermalin and Katz (2009), Hermalin (2013), Halac (2015), and Tan (2017). These papers study a variety of related issues while restricting attention to perfect observability versus perfect unobservability of the investment, but they do not consider more general forms

[1] Of course, unconditionally, allowing for repeated offers and a different information structure (in particular, the perfectly uninformative one) will improve welfare as shown in Gul (2001).

of information control as in here.

This paper is also related to the information control literature – see for example Rayo and Segal (2010), Ostrovsky and Schwarz (2010), Kamenica and Gentzkow (2011), and subsequent works on Bayesian persuasion. The difference is that these papers put no restriction on the signal structure choice (at least within the class of signal structures considered), whereas the signal structure here has to also satisfy an equilibrium condition. Consequently, we cannot appeal to the "concavification" argument (Aumann and Maschler, 1995), which is commonly used in the Bayesian persuasion settings; and (as will be discussed) the underlying logic of the optimal signal structure here will also be very different. Away from pure information control, Condorelli and Szentes (2017) study information design in bilateral trade by allowing the buyer to publicly choose the distribution of his valuation and consider how this choice affects the buyer's ex-ante expected information rent; by contrast, we study the effects of the information transmission of the buyer's realized valuation instead. Bergemann et al. (2015) study the effects of the signal structure that generates signals about the buyer's valuation to the seller; Roesler and Szentes (2017) study the effects of the signal structure that generates signals about the buyer's valuation to the buyer himself. Both of these papers do not allow the buyer to determine his own valuation as is allowed here.

2 Model

A buyer (he) has valuation $v = L$ for a good that a seller (she) can produce at a cost which is normalized to zero. Before interacting with the seller, the buyer can privately increase his valuation to $v = H$ at a cost c. Increasing the valuation is henceforth termed as an investment.[2] We assume that $H - L > c$ so that it is socially efficient to invest. But due to incomplete contracts, the investment decision is not contractible.

[2] This implies that investment is deterministic. However, the results are readily extended to the case of stochastic investment; see Section 5.2.

After the investment decision is made but before trade, the seller receives a signal s regarding v. Let the set of signals be S. For expositional clarity, we assume that S is a finite set, although this is without loss of generality. A signal structure is defined by $\{S, \pi\}$ where $\pi(s|v)$ denotes the conditional probability of $s \in S$ given valuation v. This signal structure is common knowledge to both players at the start of the game.

After observing the signal, the seller makes a take-it-or-leave-it offer p to the buyer. If accepted, the seller's payoff is p while the buyer's payoff is $v - p - \mathbb{I}c$, where $\mathbb{I}$ is an indicator function that takes the value 1 if the buyer invested, and is zero otherwise; if rejected, the seller's payoff is 0 and the buyer's payoff is $-\mathbb{I}c$.

Our equilibrium concept is the Perfect Bayesian equilibrium. Given a signal structure, the buyer optimally chooses to invest or not, taking into account the distribution of signals that his investment decision will generate and his conjecture about the seller's pricing strategy after observing each signal $s \in S$. The seller, upon observing a signal s, forms a posterior which depends on both her conjecture about the buyer's investment decision and the distribution of the signals under the signal structure, and then optimally sets a price based on her posterior. The buyer then accepts if $p \leq v$ and rejects otherwise. In equilibrium, each player's conjecture about the other player's strategy is correct. It is readily noted that the seller will only set $p = L$ or $p = H$ in equilibrium.

Throughout, we let $q \in [0, 1]$ denote the probability of the buyer investing. Thus, we say that *a signal structure implements* q if the buyer's strategy of investing with probability q can be sustained as an equilibrium under the signal structure.

Proposition 1. *The buyer's expected payoff is always zero in equilibrium. Moreover,* $q = 1$ *cannot be implemented.*

Proof. Since the seller will never set p lower than L, the buyer's payoff is zero if he does not invest (i.e. $q = 0$). If $q \in (0, 1)$ in equilibrium, the buyer must be indifferent between investing and not investing, which means that his payoff is also 0. Lastly, q cannot be 1 in equilibrium. This is because if $q = 1$, the seller will correctly conjecture it in equilibrium and always sets $p = H$, in which case the buyer will never invest in the first place. $\square$

Proposition 1 thus implies that the seller's payoff in equilibrium is also the social welfare, so there is no need to differentiate between the two. Henceforth, the term "optimality" will refer to optimality of the seller's expected payoff. Since q cannot be 1 in equilibrium, the first-best optimality cannot be achieved.

3 Benchmarks

3.1 Fully Informative π: the Hold-up Case

Consider a fully informative signal structure first, which gives the classic hold-up problem: since the seller perfectly knows v, she always sets $p = v$ and hence, the buyer never invests in equilibrium. The social welfare is thus L which is all given to the seller.

3.2 Fully Uninformative π

Consider a fully uninformative signal structure next: $\pi(s|v) = \pi(s'|v)\ \forall v \in \{L, H\}, s \in S$. The following restates the result of Gibbons (1992) and Gul (2001) that the players' payoffs under no information are the same as in the hold-up case:[3]

Proposition 2. *The equilibrium under the fully uninformative signal structure is unique: the buyer invests with probability $\frac{L}{H}$ and the seller sets $p = L$ with probability $\frac{c}{H-L}$ after every (uninformative) signal. The seller's equilibrium expected payoff is L. Therefore, the players' payoffs are the same as in the hold-up case.*

Proof. From Proposition 1, $q \neq 1$. Similarly, $q \neq 0$; if $q = 0$, the seller will also correctly conjecture that in equilibrium and always sets $p = L$, in which case the buyer will deviate

[3] See problem 2.23 in Gibbons (1992), and Proposition 1 in Gul (2001).

to choosing $q = 1$ instead. Let ρ be the probability that the seller sets price $p = L$ (ρ is independent of the signal since the signal has no information). Since $q \in (0,1)$, the buyer must be indifferent between investing and not investing, which implies that $\rho(H - L) - c = 0 \iff \rho = \frac{c}{H-L} \in (0,1)$. Thus the seller is also randomizing over H and L in equilibrium, which implies that she must be indifferent between the two prices: $L = qH \iff q = \frac{L}{H}$. $\square$

The seller's ignorance about the buyer's investment limits her ability to expropriate the gains from investment, thus improving ex-ante efficiency by (partially) restoring the buyer's ex-ante investment incentive. As a result, the buyer invests with positive probability. However, the asymmetric information at the trading stage creates ex-post inefficiency because trade breaks down when the buyer did not invest but the seller sets $p = H$. These two effects exactly cancel each other out in equilibrium.

4 Optimal Signal Structure

In this section, we first fix a $q \in (0,1)$ and solve for the signal structure that gives the highest payoffs while implementing q, assuming that such a signal structure exists. The subsequent variables will be dependent on q, but we omit the argument throughout to ease notation. Let β_s be the seller's posterior belief that $v = H$ after observing signal s under signal structure $\{S, \pi\}$. With q correctly conjectured by the seller in equilibrium,

$$\beta_s = \Pr(v = H|s) = \frac{\pi(s|H)\,q}{\pi(s|H)\,q + \pi(s|L)\,(1-q)}. \tag{4.1}$$

Conditional on belief β, the seller's expected payoff from setting $p = H$ is βH, and that from $p = L$ is L. Thus the principal sets $p = H$ when $\beta > \frac{L}{H}$, sets $p = L$ when $\beta < \frac{L}{H}$, and is indifferent between either price when $\beta = \frac{L}{H}$.

Denote $x_s := q\pi(s|H) + (1-q)\,\pi(s|L)$ as the ex-ante probability of signal s realizing.

We say that a signal structure $\{S, \pi\}$ is *almost direct* if $S = S^{ad} := \{l, n, h\}$ and π has the following properties:[4]

$$\begin{aligned} &\text{if } x_l > 0, \text{ then } \beta_l < \tfrac{L}{H}\ ; \\ &\text{if } x_n > 0, \text{ then } \beta_n = \tfrac{L}{H}\ ; \\ &\text{if } x_h > 0, \text{ then } \beta_h > \tfrac{L}{H}\ . \end{aligned} \tag{4.2}$$

An almost direct signal structure produces an incentive compatible pricing recommendation for the seller almost all the time. The seller chooses $p = L$ when she observes $s = l$, chooses $p = H$ when she observes $s = H$, but she is indifferent between either price when she receives the neutral signal n.

Lemma 1. *Suppose a signal structure $\{\pi, S\}$ implements q and the seller's expected payoff in the equilibrium is V. There exists an almost direct signal structure that also implements q and the seller's expected payoff in the equilibrium is also V.*

Proof. See Appendix. □

Lemma 1 implies that it is without loss of generality to restrict attention to signal structures that are almost direct. The intuition behind is similar to the revelation principle. For any signal structure, the seller sets the same price whenever her posterior is less (resp. more) than $\frac{L}{H}$, so all signals that generate posteriors that are less (resp. more) than $\frac{L}{H}$ can be grouped together accordingly.[5]

Restricting attention to almost direct signal structures now, we follow Kamenica and Gentzkow (2011) and frame the problem here as choosing a distribution of posteriors $\{x_s, \beta_s\}_{\sum_s x_s = 1}$

[4]Note that β_s is undefined when $x_s = 0$. For completeness, we specify the convention that $\beta_l = 0$ when $x_l = 0$, $\beta_n = \frac{L}{H}$ when $x_n = 0$, and $\beta_h = 1$ when $x_h = 0$.

[5]Unlike the Bayesian persuasion literature, restricting the signal space to be the state space here is not without loss (at least at this stage). In the persuasion literature, the papers typically assume that the Receiver has a unique optimal action under each belief, or that the Receiver always takes the Sender-preferred action when the Receiver is indifferent. In this paper, the action taken at any particular belief can disrupt the equilibrium, hence it is not *a priori* clear what is a "Sender-preferred action" at each belief. In particular, when the seller is indifferent between either price, her randomization strategy can have implication on her ex-ante expected payoff.

under the restrictions of almost direct signal structures in (4.2), and subject to the Bayes plausibility constraint:

$$\sum_{s\in\{l,n,h\}} x_s\beta_s = q. \tag{4.3}$$

The original signal structure can then be backed out via $\pi(s|H) = \frac{\beta_s x_s}{q}$ and $\pi(s|L) = \frac{(1-\beta_s)x_s}{1-q}$. In addition, the signal structure must satisfy the equilibrium condition:

$$\Big(\pi(l|H) + \sigma\pi(n|H)\Big)(H-L) - c = 0, \tag{4.4}$$

where $\sigma \in [0,1]$ is the probability that the seller sets $p = L$ after observing $s = n$. The left hand side of (4.4) is the buyer's expected payoff from investing. For him to be indifferent between investing and not investing (so that $q \in (0,1)$ in equilibrium), this payoff must be the same as his payoff from not investing, which is 0. Using (4.1), (4.4) is equivalent to:

$$x_l\beta_l + \sigma x_n\beta_n = q\left(\frac{c}{H-L}\right). \tag{4.5}$$

Therefore, an almost direct signal structure implements q and σ if the resulting distribution of posteriors $\{x_s, \beta_s\}_{\sum_s x_s = 1}$ satisfies (4.2), (4.3) and (4.5). Since the seller's conditional expected payoff at belief β_n is L, her ex-ante expected payoff is:

$$(x_l + x_n)L + x_h\beta_h H = L + x_h(\beta_h H - L) \tag{4.6}$$

Lemma 2. *Suppose there exists an almost direct signal structure with $x_n > 0$ that implements q and $\sigma < 1$, and the seller's expected payoff in the equilibrium is V. There exists an almost direct signal structure that implements q and $\sigma = 1$, and the seller's expected payoff in the equilibrium is strictly higher than V.*

Proof. Suppose a distribution of posteriors $\{x_s, \beta_s\}$ supports an equilibrium with q and $\sigma < 1$.

Consider another distribution of posteriors $\{x'_s, \beta'_s\}$ where $x'_l = x_l + (1-\sigma)\, x_n \left(1 - \frac{\beta_n}{\beta_h}\right)$, $x'_n = \sigma x_n$, $x'_h = x_h + (1-\sigma)\, x_n \frac{\beta_n}{\beta_h}$; $\beta'_l = \frac{x_l}{x'_l}\beta_l$, $\beta'_n = \beta_n$, $\beta'_h = \beta_h$. It is readily verified that $\{x'_s, \beta'_s\}$ is a valid almost direct signal structure that satisfies (4.2) (4.3) and (4.5). Since $x'_h > x_h$ and $\beta'_h = \beta_h$, from (4.6), the seller's expected payoff under $\{x'_s, \beta'_s\}$ is higher. $\square$

Intuitively, the seller's expected payoff when signal n is realized is L regardless of her setting $p = L$ or $p = H$. Thus, if there is a probability of $1 - \sigma > 0$ that she will set $p = H$, her equilibrium payoff can be increased by (appropriately) shifting this probability weight $1 - \sigma$ to signal h instead where she will earn a payoff higher than L. Doing so would not affect the equilibrium condition (4.5) since the ex-ante probability of $p = L$ is not altered.

Lemma 2 implies that when searching for the (almost direct) signal structure that maximizes the sender's expected payoff, we can restrict attention to equilibria with $\sigma = 1$. In turn, signals l and n are essentially equivalent and can hence be pooled together. Thus we can restrict attention to *direct* signal structures where $S = \{l, h\}$, the resulting posteriors satisfy $\beta_l \leq \frac{L}{H}$ and $\beta_h > \frac{L}{H}$, and the seller plays $p = L$ at $s = l$ and $p = H$ at $s = h$.[6] The following proposition gives the main result of the paper.

Proposition 3. *A signal structure that implements q exists if and only if $q \leq \frac{1}{1+\frac{c}{L}}$. For any $q \leq \frac{1}{1+\frac{c}{L}}$, the signal structure that maximizes the seller's expected payoff while implementing q is unique within the set of direct signal structures. It consists of:*

$$\pi(h|H) = 1 - \frac{c}{H-L} \quad ; \quad \pi(l|H) = \frac{c}{H-L}$$
$$\pi(h|L) = 0 \quad ; \quad \pi(l|L) = 1$$

The resulting posteriors are $\beta_h = 1$ and $\beta_l = \frac{1}{1+\frac{1-q}{q}\left(\frac{H-L}{c}\right)}$. The seller's expected payoff is

[6] To see this formally, suppose an almost direct signal structure $\{x_s, \beta_s\}$ implements $q = 1$ and $\sigma = 1$. Consider another almost direct signal structure $\{x'_s, \beta'_s\}$ such that $x'_l = x_l + x_n$, $\beta'_l = \frac{x_l\beta_l + x_n\beta_n}{x_l + x_n}$; $x'_n = 0$, $\beta'_n = \beta_n$; and $x'_h = x_h$ and $\beta'_h = \beta_h$. It is readily verified that $\{x'_s, \beta'_s\}$ also satisfies (4.2), (4.3) and (4.5) and gives the same expected payoff to the seller as $\{x'_s, \beta'_s\}$. Since $x'_n = 0$, signal n is irrelevant.

$L + q(H - L - c)$.

Proof. See Appendix. □

We emphasize that the equilibrium existence condition in Proposition 3 takes into account all possible signal structures, not just direct signal structures. This thus provides an upper bound on the possible investment frequency in equilibrium when a binary investment is not contractible ex-ante. Moreover, the optimal signal structure is independent of q (although the posteriors generated and the resulting payoff are dependent on q). This means that the (direct) signal structure that optimally implements q is the same for all implementable q.

Although this signal structure is reminiscent of the optimal signal structure in the leading "prosecutor" example in Kamenica and Gentzkow (2011) (hereafter KG), in the sense that one state is always revealed (the "L" state here and the "guilty" state in KG), the reasonings behind the two signal structures are very different. First, since the seller's payoff is convex in her belief here, the optimal signal structure in the KG world would have been the fully informative one, but this would bring us back to the hold-up case. More generally, in Bayesian persuasion, when the fully informative signal structure is not optimal but there is scope for persuasion, the optimal signal structure optimally pools "favorable" states with "unfavorable" states while maintaining the credibility of the signals. In contrast, the optimal structure here determines the conditional probabilities at each state *separately*.

In particular, the buyer's ex-ante investment incentive is provided via a probability of $\frac{c}{H-L}$ that his investment is not detected by the seller, in which case, he gets to keep the investment gains. This probability is set so that he is ex-ante indifferent between investing or not, which effectively pins down $\pi(\cdot|H)$. As for $\pi(\cdot|L)$, since $\pi(h|L)$ is the conditional probability of having ex-post inefficiency due to trade not taking place when the buyer did not invest and $\pi(h|L)$ does not affect the buyer's ex-ante investment incentive, it is set to zero to eliminate all ex-post inefficiency. The result that ex-ante investment incentive and ex-post inefficiency are taken care of separately by $\pi(\cdot|H)$ and $\pi(\cdot|L)$ respectively thus implies that there is no tradeoff between increasing ex-ante investment incentive and eliminating ex-post inefficiency.

Corollary 1. *For any implementable q, there is zero ex-post inefficiency under the optimal signal structure that implements q.*

By studying the optimal information control here, we also effectively isolate the "true" hold-up effect of the problem due to the non-contractibility of the investment. In particular, ex-ante investment incentive cannot be fully restored due to the upper bound on q. This is because when q increases, the need to maintain the credibility of signal l implies that signal l needs to detect state L more accurately and makes less mistake via (wrongly) detecting state H. However, this mistake is what creates ex-ante investment incentive for the buyer, so the need to improve the accuracy of l due to a higher q in turn destroys the buyer's ex-ante incentive to invest at all.

We close the analysis by noting from Proposition 3 that the seller's expected payoff under the optimal signal structure is strictly increasing in q. Therefore the maximum payoff is attained in the equilibrium with the highest implementable probability of investment $q = \frac{1}{1+\frac{c}{L}}$:

Corollary 2. *The set of social welfare that is achievable in equilibrium is $\left[L, \frac{HL}{c+L}\right]$.*

5 Discussions

5.1 Implementation

The simplicity of the optimal signal structure characterized in Proposition 3 implies that it can be readily implemented through practical arrangements. Viewing the signal structure as a hypothesis test for investment, the optimal signal structure makes some "false negative" type II error (i.e. fail to detect the investment) but never makes any "false positive" type I error (i.e. detect an investment when there is not). Therefore, the requirement to implement it in practice is a "technology" that can accurately detect an investment, but the "technology" has to become unavailable sometimes without the seller's knowledge.

For example, in an organizational context with workers making firm-specific skill upgrade, the "technology" could be a monitor who shirks sometimes. In a procurement relationship where the upstream producer's investment lowers its cost to produce the downstream buyer's product, the "technology" could be the buyer's random inspection of the producer's facilities, and the "false negative" arises from the buyer's limited access to the facilities.[7] More generally, while the investor always earns zero payoff in the model, he can sell such a "technology" (e.g. granting some access to his facilities) upfront to the non-investor in exchange for a fee; the non-investor will accept the offer if it is lower than her expected payoff later.

5.2 Stochastic Investments and Multiple Investment Outcomes

Our consideration of only deterministic binary investment outcomes helps to simplify the analysis and elucidate our point, but the main results extend to stochastic investment outcomes as well. To see this, suppose that instead of a binary investment choice as in the baseline model, the buyer gets to choose an investment level $\rho \in [0,1]$ at a cost $\phi(\rho)$, where ρ is the probability that his valuation increases from L to H. We assume that $\phi(\cdot)$ is strictly increasing and convex, with $\phi'(0) = 0$ and $\lim_{q \to 1} \phi'(q) = \infty$.

Let $f(\rho) = \rho [L + \phi'(\rho)] - L$. It is readily verified that there exists unique ρ^* such that $f(\rho^*) = 0$.

Proposition 4. *Under stochastic investment, a signal structure that implements investment ρ exists if and only if $\rho \leq \rho^*$. For any $\rho \leq \rho^*$, the signal structure that maximizes the social*

[7] In both examples, the investor's investment lowers his cost and the hold-up arises when the non-investor lowers the payment after observing the investment. It is readily observed that there is an isomorphism between this and the bilateral trade context considered in the main model.

welfare while implementing ρ is unique within the set of direct signal structures. It consists of:

$$\pi(h|H) = 1 - \frac{\phi'(\rho)}{H-L} \quad ; \quad \pi(l|H) = \frac{\phi'(\rho)}{H-L}$$
$$\pi(h|L) = 0 \quad ; \quad \pi(l|L) = 1$$

The resulting posteriors are $\beta_h = 1$ and $\beta_l = \frac{1}{1+\frac{1-\rho}{\rho}\left(\frac{H-L}{\phi'(\rho)}\right)}$; hence trade always occurs. The set of social welfare that is achievable in equilibrium is $[L, L+\rho^*(H-L)-\phi(\rho^*)]$.

We only provide the arguments behind Proposition 4; the formal proof is omitted since it is mainly retracing the steps to arrive at Proposition 3. Notice that given an almost direct signal structure, the buyer's expected payoff from investment ρ is $\rho\Big[\pi(l|H)+\sigma\pi(n|H)\Big](H-L)-\phi(\rho)$. His optimal investment level is thus determined by the first order condition:

$$\Big(\pi(l|H)+\sigma\pi(n|H)\Big)(H-L)-\phi'(\rho)=0. \tag{5.1}$$

This first order condition (5.1) then replaces the equilibrium condition in (4.4). By replacing "q" with "ρ" and "c" with "$\phi'(\rho)$", the analysis in Section 4 follows through with a few cautions.

First, the analysis in Section 4 is under the assumption that $c < H-L$, hence ρ must satisfy $\phi'(\rho) < H-L$. But since $\pi(l|H)+\sigma\pi(n|H) < 1$ (if not, the seller will never set $p = H$), (5.1) implies that any implementable ρ must satisfy $\phi'(\rho) < H-L$. Second, the buyer's payoff is no longer always zero under stochastic investment, so the objective in (4.6) is now the social welfare rather than just the seller's payoff. Third, the upper bound on the implementable investment ρ^* is "analogous" to the upper bound on the probability of investment $q = \frac{1}{1+\frac{c}{L}}$ in Proposition 3; it is readily verified that $\rho^* = \frac{1}{1+\frac{\phi'(\rho^*)}{L}}$, and $\rho \leq \frac{1}{1+\frac{\phi'(\rho)}{L}}$

if and only if $\rho \leq \rho^*$.[8] Fourth, the social welfare under an implementable investment level ρ is $L + \rho(H - L) - \phi(\rho)$, and it is strictly increasing in ρ since, as observed above, any implementable ρ must satisfy $H - L > \phi'(\rho)$; this thus give an analogous set of implementable social welfare as in Corollary 2.

While the analysis has restricted attention to binary investment outcomes throughout, we postulate that our main message remains with multiple investment outcomes. In particular, under the optimal signal structure, there should be zero ex-post inefficiency – that is, the seller's posteriors satisfy $\Pr[v < p|s] = 0$ if p is played with strictly positive probability after observing signal s. The problem is then about how to optimally add uncertainty about higher valuations across the appropriate signals to create sufficient ex-ante investment incentive for the buyer.

6 Conclusion

The literature has noted that introducing information asymmetry regarding the buyer's investment can prevent the seller from abusing her bargaining power and hence alleviate the hold-up problem. Implicitly suggested in these earlier works is a tradeoff between ex-ante investment incentive and ex-post inefficiency due to the asymmetric information. In this paper, we make the point that such a tradeoff is unnecessary, because the information needed to be hidden to create ex-ante investment incentive is different from the information that creates ex-post inefficiency when being hidden. Consequently, by hiding and revealing the right information, ex-post inefficiency can be eliminated without compromising the ex-ante investment incentive.

[8] This comes from noting that $\rho \leq \frac{1}{1+\frac{\phi'(\rho)}{L}}$ if and only if $f(\rho) \leq 0$, and f is strictly increasing.

A Appendix

Proof of Lemma 1

Proof. Let $S_h \subset S$ be the set of signals such that the posteriors generated are strictly greater than $\frac{L}{H}$; analogously, let S_l be the set of signals with posteriors strictly less than $\frac{L}{H}$, and S_n be the set of signals with posteriors equal to $\frac{L}{H}$. In addition, let σ_s denote the probability of the seller playing $p = L$ after observing signal s. In equilibrium, σ_s must be 1 if $s \in S_l$, and it must be 0 if $s \in S_h$, while it can be anything between 0 to 1 when $s \in S_n$. Consider the following almost direct signal structure $\left\{S^{ad}, \pi^{ad}\right\}$: $\pi^{ad}\left(s^{ad}|v\right) = \sum_{s \in S_{s^{ad}}} \pi\left(s|v\right)$ for all $s^{ad} \in S^{ad}$; let the seller play $p = L$ with probability $\sigma^{ad} = \frac{\sum_{s \in S_n} \pi[s|H]\sigma_s}{\sum_{s \in S_n} \pi[s|H]} = \frac{\sum_{s \in S_n} \pi[s|H]\sigma_s}{\pi^{ad}[n|H]}$ upon observing n.[9]

To check that this is an equilibrium with the buyer investing with probability q, first note that given q, the seller's pricing strategy is clearly a best response. For the buyer, his expected payoff after investing is

$$\begin{aligned}
\Pr\left[p = L|v = H\right]\left(H - L\right) - c &= \left(\pi^{ad}\left(l|H\right) + \pi^{ad}\left(n|H\right)\sigma^{ad}\right)\left(H - L\right) - c \\
&= \left(\sum_{s \in S_l} \pi\left[s|H\right] + \sum_{s \in S_n} \pi\left[s|H\right]\sigma_s\right)\left(H - L\right) - c,
\end{aligned}$$

where the second line is the buyer's expected payoff after investing under the original signal structure $\{S, \pi\}$. The buyer's payoff when he does not invest is 0 under both signal structures. Since the buyer is indifferent between investing or not under $\{S, \pi\}$, he is also indifferent under $\left\{S^{ad}, \pi^{ad}\right\}$. Thus q is the buyer's best response as well, and hence it is an equilibrium.

We check the payoff next. Conditional on the buyer investing, the seller's expected payoff

[9] If S_n is empty, then this is irrelevant.

under $\left\{S^{ad},\pi^{ad}\right\}$ is:[10]

$$\left(\pi^{ad}\left(l|H\right)+\pi^{ad}\left(n|H\right)\right)L+\pi\left(h|H\right)H=\left(\sum_{s\in S_l}\pi\left(s|H\right)+\sum_{s\in S_n}\pi\left(s|H\right)\right)L+H\cdot\sum_{s\in S_h}\pi\left(s|H\right)$$

where the RHS is the seller's expected payoff, conditional on the buyer investing, under $\{S,\pi\}$. Next, conditional on the buyer not investing, the seller's expected payoff under $\left\{S^{ad},\pi^{ad}\right\}$ is:

$$\left(\pi^{ad}\left(l|L\right)+\pi^{ad}\left(n|L\right)\right)L+\pi\left(h|H\right)\cdot 0=\left(\sum_{s\in S_l}\pi\left(s|L\right)+\sum_{s\in S_n}\pi\left(s|L\right)\right)L+0\cdot\sum_{s\in S_h}\pi\left(s|H\right),$$

where the RHS is the seller's expected payoff, conditional on the buyer not investing, under $\{S,\pi\}$. □

Proof of Proposition 3

Proof. We prove the "only if" direction of the existence result first. Suppose, for a contradiction, that $q>\frac{1}{1+\frac{c}{L}}$ but there exists an almost direct signal structure that implements q. From Lemma 2, there exists a direct signal structure that implements q. Let $\beta_l\leq\frac{L}{H}$ and $\beta_h>\frac{L}{H}$ be the resulting posteriors. From (4.5), $\beta_l=\frac{qc}{x_l(H-L)}$; from (4.3), $x_l=\frac{\beta_h-q}{\beta_h-\beta_l}$. Combining the two, we get:

$$\begin{aligned}\beta_l&=\frac{q\left(\frac{c}{H-L}\right)\beta_h}{\beta_h-q+q\left(\frac{c}{H-L}\right)}\\&=\frac{q\left(\frac{c}{H-L}\right)}{1-\frac{q}{\beta_h}\left(1-\frac{c}{H-L}\right)}.\end{aligned}\tag{A.1}$$

[10]Note that the seller's conditional expected payoff under posterior $\frac{L}{H}$ is always L.

Since $\frac{c}{H-L} < 1$, β_l is decreasing in β_h. $\beta_h \leq 1$ then implies that $\beta_l \geq \frac{q\left(\frac{c}{H-L}\right)}{1-q\left(1-\frac{c}{H-L}\right)}$. When $q > \frac{1}{1+\frac{c}{L}}$, $\beta_l > \frac{L}{H}$ which contradicts $\beta_l \leq \frac{L}{H}$.

Next, for the "if" direction, the direct signal structure with $\beta_h > \frac{L}{H}$ and β_l in (A.1), which will be weakly lower than $\frac{L}{H}$ when $q \leq \frac{1}{1+\frac{c}{L}}$, is readily verified to implement q.

As for optimality, it is implied by Proposition 2 that it suffices to look within the set of direct signal structures. For any β_h, the corresponding β_l is (A.1), and the seller's expected payoff, from (4.6), is $V(\beta_h) = L + x_h\beta_h H - x_h L$. From (4.3) and (4.5),

$$x_h\beta_h H = (q - x_l\beta_l)\, H = \left[q - q\left(\frac{c}{H-L}\right)\right] H.$$

Next,

$$\begin{aligned} x_h L &= \frac{q-\beta_l}{\beta_h-\beta_l} L = \left(\frac{q - \frac{q\left(\frac{c}{H-L}\right)\beta_h}{\beta_h - q + q\left(\frac{c}{H-L}\right)}}{\beta_l - \frac{q\left(\frac{c}{H-L}\right)\beta_h}{\beta_h - q + q\left(\frac{c}{H-L}\right)}}\right) L \\ &= \left(\frac{q\left[\beta_h - q + \left(\frac{c}{H-L}\right)(q-\beta_h)\right]}{\beta_h(\beta_h - q)}\right) L \\ &= \frac{q}{\beta_h}\left(1 - \frac{c}{H-L}\right) L \end{aligned}$$

Therefore, $V(\beta_h) = L + q\left(H - \frac{L}{\beta_h}\right)\left[1 - \left(\frac{c}{H-L}\right)\right]$. Since $V(\beta_h)$ is strictly increasing, the optimal β_h is 1, and $\beta_l = \frac{q\left(\frac{c}{H-L}\right)}{1-q+q\left(\frac{c}{H-L}\right)}$. The seller's expected payoff is $V(1) = L + q(H - L - c)$. The signal structure is then backed out via $\pi(s|H) = \frac{\beta_s x_s}{q}$ and $\pi(s|L) = \frac{(1-\beta_s)x_s}{1-q}$.[11] □

References

Aumann, Robert J and Michael Maschler, *Repeated Games with Incomplete Informa-*

[11]It would be helpful to note using (4.5) that $x_h\beta_h = q - x_l\beta_l = q\left(1 - \frac{c}{H-L}\right)$; so $x_h = q\left(1 - \frac{c}{H-L}\right)$ when $\beta_h = 1$, and $x_l = 1 - q\left(1 - \frac{c}{H-L}\right)$.

tion, MIT press, 1995.

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris, "The Limits of Price Discrimination," *American Economic Review*, March 2015, *105* (3), 921–57.

Condorelli, Daniele and Balazs Szentes, "Information Design in the Hold-up Problem," 2017.

Gibbons, Robert, *Game Theory for Applied Economists*, Princeton University Press, 1992.

González, Patrick, "Investment and Screening under Asymmetric Endogenous Information," *The RAND Journal of Economics*, 2004, pp. 502–519.

Gul, Faruk, "Unobservable Investment and the Hold-Up Problem," *Econometrica*, 2001, *69* (2), 343–376.

Halac, Marina, "Investing in a relationship," *The RAND Journal of Economics*, 2015, *46* (1), 165–185.

Hermalin, Benjamin E., "Unobserved Investment, Endogenous Quality, and Trade," *The RAND Journal of Economics*, 2013, *44* (1), 33–55.

_ **and Michael L. Katz**, "Information and the Hold-up Problem," *The RAND Journal of Economics*, 2009, *40* (3), 405–423.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

Lau, Stephanie, "Information and Bargaining in the Hold-up Problem," *The RAND Journal of Economics*, 2008, *39* (1), 266–282.

Ostrovsky, Michael and Michael Schwarz, "Information Disclosure and Unraveling in Matching Markets," *American Economic Journal: Microeconomics*, 2010, *2* (2), 34–63.

Rayo, Luis and Ilya Segal, "Optimal Information Disclosure," *Journal of Political Economy*, 2010, *118* (5), 949–987.

Roesler, Anne-Katrin and Balázs Szentes, "Buyer-Optimal Learning and Monopoly Pricing," *American Economic Review*, July 2017, *107* (7), 2072–80.

Tan, Teck Yong, "The Extrinsic Motivation of Freedom at Work," 2017.