Bargaining under the Illusion of Transparency*

Kristóf Madarász†
London School of Economics

Draft!

Abstract

A seller sells an object to a privately informed buyer who projects information, more precisely, exaggerates the probability that the seller knows her valuation. I show that letting the buyer bargain and name her own price first improves the seller's revenue above the full commitment optimal solution. When the seller makes all price offers, then given *any* positive, potentially vanishing, prospective projection, the model implies a full reversal of the Coasian result in stationary strategies. Specifically, as bargaining becomes smooth the seller is able to extract all surplus. I relate the comparative static predictions to existing evidence.

Keywords: Biased Beliefs, Bargaining, Information Projection, Pricing, Coase Conjecture.

1 Introduction

A fundamental topic in microeconomics, underlying monopoly theory, concerns the setting where a seller wants to sell an object to a buyer privately informed about her valuation. A salesman wants to sell a new car to a buyer not knowing how much the buyer likes the car. A local gym wants to sell a contract to a potential member not knowing how much she is willing to pay. An employer wants to hire an expert for a task of fixed value facing uncertainty about the expert's reservation wage. Does the seller - the uninformed party - benefit from bargaining with the buyer over time without commitment?

Under the standard assumption is that the buyer correctly understands the extent to which her information is private *vis-a-vis* the seller, she fully appreciates the uncertainty faced by the

*I would like to thank Aislinn Bohren, Arthur Campbell, Mehmet Ekmekci, Marina Halac, Philippe Jehiel, Navin Kartik, Matthew Rabin, Luis Rayo, Balazs Szentes, Jorgen Weibull and seminar audiences at Arizona, Harvard, LSE, Stockholm, ESSET Gerzensee 2013, Royal Holloway, University College London, UCLA, Instituto Tecnologico Autonomo de Mexico, for comments. All errors are mine.

†Contact: k.p.madarasz@lse.ac.uk, or London School of Economics, WC2A 2AE, UK.

seller, the key result is a no-haggling result (e.g., Myerson 1981; Riley and Zeckhauser 1983; Skreta 2006). The optimal way to sell requires commitment and consists of making a single take-it-or-leave-it offer. Furthermore, following the celebrated insight of Coase (1972), even if the seller makes all the offers, as bargaining becomes smooth, in that the time delay between offers goes to zero, the seller loses all his static monopoly rent and sells immediately at the lowest possible reservation value of the buyer (e.g., Fudenberg, Levine and Tirole 1985; Gul, Sonnenschein and Wilson 1986).

In contrast to this standard assumption, robust evidence shows that people fail to fully appreciate the extent to which their information is private. Specifically, the typical person systematically *projects* her information onto others: she too often thinks that others act based on the information she does. When bargaining, the privately informed buyer thinks she is more transparent than she truly is. Such an illusion is potentially key as bargaining outcomes critically depend on how much information the buyer actually conceals.

This paper considers the impact of such information projection on bargaining behavior. It explores the robustness of various Bayesian predictions to its presence and shows that this phenomenon provides a potentially important channel through which the seller can benefit from haggling over time. The results shed novel light on a variety of empirical phenomena - the value of delay and commitment and the impact of private information- and have potential implications for a number of applications.

Section 2 incorporates information projection into a class of sequential bargaining games.[1] Each player understands the distribution of her own information, but a biased player exaggerates the probability that whenever her opponent acts, he does so as if he knew her private information. Specifically, the privately informed buyer exaggerates the probability that her valuation exogenously leaks to the seller. The seller is sophisticated and also understands the buyer's (mistaken) perception. Given such heterogenous perceptions of the distribution of information in the game, players choose strategies that are sequentially rational given each player's perception of the distribution of Nature's moves and the players' strategies defined both on real histories that could occur and on imaginary histories that can occur only in the buyer's imagination.

Section 3 considers a simple setting with two periods. The fact that under unbiased expectations the seller's dynamic revenue is bounded by the static monopoly profit implies a silence and a durability property: the seller's maximal revenue under a protocol where the players alternate making offers or under a protocol where the seller can make a second offer with some probability is bounded from above by the seller's static monopoly payoff. In contrast, both of these properties are violated when the buyer projects information even if such projection is constrained to happen only at the beginning of the game.

To highlight the role of the illusion of transparency over transparency *per se*, I first show

[1] I build on a companion paper, Madarasz (2014) introducing information projection into static Bayesian games.

that the above properties carry over to the case with a commonly known *ex-ante* private leakage probability. For example, when entering a car-dealership, the salesman may privately infer a buyer's valuation. The buyer now also faces uncertainty not knowing whether the salesman is informed or not, as in the case of information projection, but she is well calibrated about such uncertainty which is common knowledge. Nevertheless, the seller's expected revenue is still highest under the protocol that corresponds to a single take-it-or-leave-it offer.

To illustrate the logic for such a silence property, consider an alternating-offer protocol. Can the seller boost his static monopoly payoff by transferring bargaining power and letting the buyer speak first? For this to be true, the buyer would need to reveal information about her valuation. If the buyer revealed herself, the maximal surplus she would obtain in equilibrium would be bounded by her bargaining rent, i.e., the extent to which the seller preferred an agreement now as opposed to waiting to make a counter-offer in next round. Hence the buyer reveals herself only if such a sure bargaining rent is higher than her expected information rent. When the buyer is well-calibrated about her information rent, the amount of bargaining rent she would need to obtain to reveal herself is greater than the value of the buyer's information rent to him. In short, it is not beneficial for the seller to let the buyer speak.

In contrast, a buyer who projects information underestimates her information rent. She is willing to reveal herself too cheaply, that is in exchange for a sure bargaining rent that does not compensate for her true information rent. By adopting a bargaining protocol where it is sufficiently costly, but not too costly, for the seller to wait until he can make a counter-offer, the seller can take advantage of the buyer's mistaken perception. There needs to be sufficient friction between the bargaining rounds so that it is sufficiently costly for the seller to wait, but this cost can now be lower than the value of the buyer's information rent to the seller. Letting the buyer speak and engaging in such a 'waiting game' is a successful tactic here which boosts the seller's revenue above the static monopoly payoff. I show a similar result in the context where the seller makes all offers.

Section 4 applies the model to an infinite horizon seller-offer game with a continuum of types - the classic domain of the Coase conjecture. Given rational expectations, as bargaining becomes smooth, i.e., the time delay or noise between bargaining rounds goes to zero, the seller loses all static monopoly rent. The fact that the buyer possesses private information implies that the seller is constrained to sell at the lowest possible reservation value of the buyer.

In contrast, focusing on stationary strategies, I show that given *any* positive degree of prospective projection - where the buyer wrongly believes that with some (potentially arbitrarily small) probability her private information might leak to the seller from any one round to the next - the reverse happens. Not only is the monopolist able to protect his static profit, but dynamic haggling without commitment becomes a full-rent extracting practice allowing the seller to obtain all benefits from trade.

Projection impacts not only the revenue result, but comparative static predictions away from the limit as well. If bargaining is sufficiently smooth, a force due to projection overtakes the classical force leading to the reversal of the Bayesian comparative static predictions on the level of equilibrium prices away from the limit as well. This is true despite the fact that given any positive friction in bargaining the perceived 'mechanical' utility consequence of information projection goes to zero as such mistaken beliefs go to zero. The logic of why information projection leads to a reversal of the Coasian result is based on its impact on the relative strength of two countervailing effects. As bargaining becomes smooth both the buyer's cost of postponing a purchase and the seller's cost of postponing a sale decreases.

The sophisticated seller knows that the buyer's information will never leak to him. Hence, holding the buyer's strategy constant, his trade-offs are the same in both cases. Specifically, the seller cannot just wait for information to arrive, rather the seller still becomes more pessimistic after each rejection and reduces price over time. Given unbiased expectations, the buyer correctly understands the extent to which the seller becomes more pessimistic over time. As bargaining becomes smooth the differential willingness of different buyer types to wait decreases. This classic force makes price discrimination harder. Given information projection, the buyer underestimates the extent to which the seller becomes pessimistic over time. Specifically, she believes that there is some probability per round that the seller might figure out her valuation and hence never drop the price below her valuation. Since a higher type has more to lose from leakage than a lower type, this represents a new force which makes price-discrimination easier to achieve.

Key to the intuition is that both the classical force and the force due to projection becomes stronger as bargaining becomes more smooth. The result is then based on the fact that no matter how tiny projection is, it is sufficient to change the perceived relative strengths of the above two effects. If bargaining is sufficiently smooth, the cost for the seller to postpone a sale now decreases more quickly than the perceived cost for the buyer to postpone a deal. Since in the limit the seller cannot make significant drops from one round to the next, the seller's initial price converges to the highest possible valuation of the buyer and the buyer's willingness to accept converges to her reservation value. The seller is now able to extract the full surplus.

In Section 4.1 I further strengthen the result and show that the above reversal holds even if the degree of projection vanishes as delay vanishes. I describe an exact threshold that identifies a discontinuity. If the relative speed of how quickly projection vanishes as delay vanishes is greater than square-root, the Coasian result of full revenue loss follows. If it is lower, then full rent extraction follows. Hence even if such mistaken beliefs vanish, smooth bargaining might be a way for the seller to capture all benefits from trade.

In Section 4.2 I relate the testable comparative static predictions of the model to the existing experimental evidence - Rapoport, Erev and Zwick (1995). Consistent with the predictions of the model, and in violation of the reverse predictions without projection, the data shows that buyers accept offers too soon and that as bargaining becomes smoother the sellers raise rather than lower their prices and are able to improve their revenue.

Although there are many reasons why selling through a non-negotiable price may be desirable for the seller, haggling without commitment is the norm in many settings both historically and in the present day. It is adopted in the bazaar, when selling (more expensive) items such as furniture, mortgage, real-estate or telecommunication contracts. It is commonly adopted in the sale of standardized new cars in local dealerships, where evidence shows that this practice leads to substantial price variation uncorrelated with easily observable socio-economic characteristics of buyers, Goldberg (1996). Information projection provides a potentially powerful channel through which forming prices based on one-on-one negotiations without commitment becomes a profitable practice enhancing the seller's monopoly power. I conclude by discussing the results and further implications of the model.

1.1 Evidence and Related Literature

Evidence for information projection comes from a variety of domains. Key manifestations of this phenomenon are such robust findings as the curse-of-knowledge, Camerer, Loewenstein and Weber (1989) who document this in the context of a double auction, Loewenstein, Moore and Weber (2006), Birch and Bloom (2007); the illusion of transparency, Gilovich, Medvec and Savitsky (1998, 1999); the hindsight bias, Fischhoff (1974), Biais and Weber (2010).[2] In all these settings, people's beliefs are biased in that they too often act as if *others* knew their private information.[3]

Samuelson and Bazerman (1985), in the context of bilateral trade, provide evidence that in a common value setting privately informed subjects - sellers in their setting - act as if uninformed subjects were also informed, i.e., they set prices as if buyers knew their valuations, see also Keysar, Ginzel and Bazerman (1995).[4] Consistent with the assumptions in this paper, Danz, Madarasz and Wang (2014) provide evidence in a strategic setting showing that subjects do mistakenly project their own information onto others, but at the same time, they are sophisticated in predicting that *others* will project information onto them.

Besides those already mentioned, this paper relates to many key contributions to bargaining with one-sided private information. Riley and Zeckhauser (1983) show that the seller-optimal way to sell to a buyer with a privately known valuation is to essentially commit to a single take-it-or-leave-it offer. Skreta (2006) extends their analysis and shows that when the seller cannot

[2] Although hindsight bias is sometimes interpreted as an intrapersonal mistake, most of the evidence comes from between-subject designs, i.e., interpersonal settings.

[3] The documentation of projection information in perspective-taking tasks goes back to the work of Piaget and Inhelder (1948) studying children. Birch and Bloom (2007) has shown that the same mistake is present amongst Yale undergraduates in slightly more complex tasks. For a partial review of the rest of the evidence and some implications to inference problems see Madarasz (2012).

[4] Related but less structured evidence on exaggerated perceptions of transparency in negotiations comes from e.g., Vorauer and Claude (1998), Vorauer and Ross (1999), and Gilovich et al. (2003).For example, Vorauer and Claude (1998) show that even when the behavior of participants was completely constrained by the situations, participants too often felt that observers could accurately detect the true nature of their internal state.

commit to a price over time, the optimal protocol for the seller involves still involves making a single price offer in each round. For a general class of distributions and per-period mechanisms, she shows that the seller-optimal sequential mechanism is a seller-only offer protocol where the seller makes a take-it-or-leave-it offer in each round. The revenue from such a dynamic protocol is always bounded from above by the full commitment static monopoly payoff achieved by a single take-it-or-leave-it offer.

A large literature investigates the robustness of the Coasian logic of negative selection and the loss of static monopoly payoff, e.g., Bond and Samuelson (1986) consider the impact of good depreciation in a monopoly market, Sobel (1991) and Fuchs and Skrzypacz (2010) given an increasing cost of serving higher valuation types the arrival of new buyers. Ausebel and Deneckre (1989) studies bargaining under the so-called 'no-gap' assumption and provides a folk theorem where in addition to the Coasian equilibrium the seller can maintain any expected revenue lower than the static monopoly payoff. Board and Pycia (2014) further show that under an assumption on a privately known outside option for the buyer, monopoly pricing is the unique perfect equilibrium. In contrast, in the presence of information projection, dynamic bargaining is essential to boost the seller's revenue above the static monopoly payoff.

In terms of the informational assumptions, the closest to this paper is Feinberg and Skrzypacz (2006). They study an infinite-horizon seller-offer game with two buyer types. They assume that the buyer faces initial uncertainty as to whether the seller is informed or uninformed about the buyer's valuation. In contrast to this paper, however, they maintain the assumption that this information structure is common knowledge. They show that in an infinite horizon setting such second-order uncertainty produces positive delay even if bargaining becomes smooth. Dynamic bargaining, however, here is still a process that leads to a lower revenue than making a single take-it-or-leave it offer.

Finally this paper is related to the literature studying biased beliefs in bargaining. In particular, Yildiz (2003) considers sequential bargaining under the assumption that players optimize given overly optimistic beliefs about their bargaining power. He provides conditions where despite such over-optimism, immediate agreement follows. This paper is similar in that here players also optimize given commonly known heterogenous beliefs about the underlying environment.

2 Setup

Consider the classic setup. The seller has an object for sale valuing it at a normalized amount of 0. The buyer values it at some strictly positive amount θ such that there is common knowledge of benefits from trade. Bargaining happens sequentially over time $t = 1, 2, \ldots$. In each round one of the parties makes a price offer that the other party can accept or reject. Bargaining lasts until the parties reach an agreement or some final date T is achieved.[5]

As standard, friction in bargaining, denoted by Δ, means that there might be some noise or delay in the process. Let $e^{-\Delta}$ denote the probability that bargaining breaks down from one round to the next. Equivalently, Δ will be interpreted as the amount of delay or period length between offers given time-discounting and a normalized interest rate of one. The terminology that bargaining becomes smooth refers to the case that $\Delta \to 0$.

True Information Structure. At time $t = 0$, a bargaining protocol is fixed. The buyer then learns θ. Finally, upon meeting the seller, the buyer's valuation may privately leak to the seller - who then becomes informed about θ - with probability α. Nature makes no additional informational moves. The case where $\alpha = 0$, corresponds to the classic bargaining setting which is the main focus of this paper. The case where $\alpha > 0$, corresponds to a generalization, where the buyer also faces initial uncertainty as to whether the seller is informed or not.[6] I consider such a generalization only to facilitate the interpretation of the results in Section 3.

Information Projection Information projection corresponds to an exaggerated perception that one's opponent has access to one's private information. To introduce such beliefs into the above game, consider a perturbation of the true game. Suppose that at the beginning of *each* round t, with $t > 0$, independent of the history leading to that round, Nature plays a binary leakage lottery. The realization of this lottery determines whether the buyer's information becomes available to the seller or not. Only the seller observes this realization. A buyer who exhibits information projection of degree $\rho = \{\rho_t\}_{t=1}^{T} \in [0,1)^T$ believes that the probability of a positive leakage move in round t is ρ_t. The seller believes that this probability for all $t \geq 1$ is zero. These heterogenous beliefs about Nature's moves can be understood as common knowledge. - *building on the static framework in Madarasz (2014)*

Perfect Projection Equilibrium. Given the above beliefs, I adopt the notion of perfect $\rho-$ information projection equilibrium (equilibrium henceforth). It corresponds to a perfect equilibrium for observable move games with the only modification that players have the above described heterogenous beliefs about Nature's fixed moves. To keep the analysis focused, I present the formal definition in the Appendix, but describe it informally below.

Consider a strategy belief-system pair (σ^ρ, μ^ρ) in the perturbed game. This pair has to be such that: (i) at each information set a player's strategy is sequentially rational given *this* player's belief system, (ii) each player's belief system is consistent and is derived via Bayes' rule (whenever possible) given the players' strategies and *that* player's belief about the distribution of Nature's moves, (iii) the seller's actions never change his beliefs about the buyer's valuation. Crucially, the above conditions apply to all histories in the perturbed game, i.e., also to imaginary histories that only the buyer thinks could ever possibly be reached. [7] A perfect $\rho-$projection

[5] Female pronounces are used for the buyer and male pronounces for the seller.

[6] Such a generalization in bargaining was considered by Feinberg and Skrzypacz (2006).

[7] Note that since the players' actions are all observable, these imaginary information sets for the seller are

equilibrium of the true game is then the natural restriction of such a $(\sigma^{\rho}, \mu^{\rho})$ to the true game.

Discussion. Note that given the above solution concept, whatever happens in the game is consistent with what players think can happen in the game. The departure is solely that the biased buyer attaches positive probability to histories occurring that might never occur or occur with a different probability than perceived.[8]

To understand the extent to which classic results are robust to the presence of information projection, in the analysis below, I explore two cases. In Section 3, I constrain projection to occur only at the beginning of the bargaining process. The revenue result extend *a fortiori* to the case where projection happens in all rounds. In Section 4 I consider dynamic projection but allow it to be arbitrarily small.

Note that there are two closely related psychological interpretations of dynamic projection. In the primary interpretation the privately informed buyer too often thinks that whenever the seller acts he acts using the information the buyer has. Relatedly, the buyer might have an irrational fear that independent of what she does her valuation leaks to her opponent even if it haven't leaked before. Both of these imply - in accordance with the existing evidence - that a biased buyer does not think that simply by acting differently she can affect the event of the seller acting on her private informational.

Finally, note that the initial stage above has two interpretations. As common in the bargaining literature, and followed in this paper, one can take any bargaining game as given and analyze the model there including comparative statics on nominal bargaining power and friction in the game. Following the approach of Riley and Zeckhauser (1983) or Skreta (2006) one can also think of the bargaining protocol - including the bargaining friction characterizing it - as one that the seller picks at the beginning of $t = 0$. Note that since at this stage no uncertainty about θ is resolved, the seller's choice cannot signal the resolution of any uncertainty to the buyer. Rather it is consistent with his prior beliefs.

Below I will denote the seller's *ex-ante* expected revenue - given the true distribution of information - that the seller can achieve by making a single-take-it-or-leave-it offer in period $t = 1$ by V_M (this depends on α). I will refer to this quantity as the monopoly payoff. I will denote by V_F the value of the full expected surplus from trade which is simply the expected value of θ. Naturally, $V_M \leq V_F$ and an equality holds iff $\alpha = 1$.

always singletons. Since there is perfect recall in the game, the specification of the imaginary informed seller's beliefs about the distribution of Nature's further moves, conditional on a positive leakage event having occurred, is inconsequential.

[8] The defining feature of this model is that players have false beliefs - as an endogenous function of the informational differences - about the information and hence the strategy of their opponent *on average*. This fact distinguishes the model qualitatively from the model of analogy-based expectations equilibrium of Jehiel (2005), and from the model of cursed equilibrium, Eyster and Rabin (2005). In both models the identifying assumption is that players have correct expectations about the distribution of their opponents' actions on *average*. Here instead the driving force is that the buyer has systematically biased expectations about the seller's strategy on average. Furthermore, in this private value setting cursed players would behave exactly as rational ones do.

3 Two-Periods

Consider a two-period setting where $\theta \in \{l, h\}$, and let the prior on the high type be q. I first re-state the fact that under rational expectations and no *ex-ante* leakage the seller's dynamic revenue is always bounded by the revenue from a single take-it-or-leave-it offer. Let $V_H(\Delta)$ denote the seller's maximal expected revenue over all perfect Bayesian equilibria that can occur in all two-period fixed sequential bargaining protocols given friction Δ.[9]

Lemma 1 *Suppose $\rho, \alpha = 0$, then $V_H(\Delta) \leq V_M$ for all Δ.*

Two consequences of the above fact are worth highlighting for this paper. Specifically, relative to making a single take-it-or-leave-it offer, it follows that

1. **Silence Property**: Letting the buyer name her own price first does not improve the seller's expected revenue,

2. **Durability Property**: A known positive chance for the seller to be able to make a second offer conditional on a rejection does not improve the seller's expected revenue.

Consider now the case where the buyer's valuation might truly leak to the seller. Since leakage is observed only by the seller, the buyer now faces uncertainty whether the seller is informed or not. Given such uncertainty, are there expected revenue gains here from haggling over time? The next result shows that as long as the buyer is well-calibrated about such uncertainty the above properties extend.

Proposition 1 *Suppose $\rho = 0$, then $V_H(\Delta) \leq V_M$ for all Δ.*

I discuss the intuition for this result when describing the results with information projection below. Note that in the static case of a single take-it-or-leave-it offer, the problem and hence V_M is fully separable: with probability α the seller is informed and collects the full surplus, with probability $1 - \alpha$ he is uninformed and collects the same rent as if there was no leakage.

As mentioned, to strengthen the results below, I restrict the buyer's projection to round 1 and hence assume that $\rho_2 = 0$. Here projection effectively implies that the buyer exaggerates the probability of the ex-ante leakage to be $\widehat{\alpha} = \alpha + (1 - \alpha)\rho_1$. All revenue results hold a fortiori when $\rho_2 > 0$.

3.1 Alternating Offers

When selling a new car to a buyer, the buyer is asked to name her own price first. If the price is not acceptable, the salesman will have to go and talk to his manager. The manager might

[9] This follows from the results of e.g., Skreta (2006).

be away that day and the salesman might then only be able to return with a counter-offer after some delay. Similarly, a firm selling a task might ask an expert to name the wage she would need to receive to accept the task. Will such a practice allow for higher sales revenue for the seller compared to quoting a single take-it-or-leave-it offer upon the initial encounter?

Let $V_A^\rho(\Delta)$ denote the seller's true maximal expected revenue given a ρ-biased buyer over the class of equilibria that can arise in the alternative-offer protocols given Δ. The next proposition claims that this revenue exceeds V_M provided the buyer is sufficiently biased and there is enough, but not too much, friction - e.g., delay - in the bargaining process.

Proposition 2 *For all $\rho > \rho_A$, and for all $\Delta \in (\Delta^\rho_{\min}, \Delta^\rho_{\max})$, it follows that $V_A^\rho(\Delta) > V_M$.*

To illustrate this result, consider first the unbiased case with no true leakage, i.e., $\alpha = 0$. Since here the buyer correctly believes that the seller cannot distinguish between a 'lie' and the 'truth', no fully separating equilibrium exists. Delay between the round where the buyer makes an offer and the round where the seller does, can only lower the seller's revenue relative to his static monopoly profit. Letting the buyer speak has no value to the seller.

Consider now a buyer who projects information and exaggerates the probability that the seller knows her valuation. Such a buyer may have an incentive to reveal herself. For this to be the case, there needs to be non-trivial friction in the bargaining process. Friction ensures that the seller prefers an early agreement to a later one, which results in positive bargaining rent for the buyer even if she reveals herself in equilibrium. Intuitively, if there is sufficient delay between rounds, and if the buyer is sufficiently worried that the seller has figured her out, then she finds it better to capitalize on her bargaining rent than to face the chance of being rejected by an informed seller and thus receiving no positive surplus in the next round. The greater is her fear, the lower is her perceived information rent and the lower the bargaining rent she needs to obtain to be willing to reveal herself.

Formally, consider a fully revealing equilibrium. Suppose the high type names a price of $p_h = e^{-\Delta}h$ which is greater than the price named by the low type p_l. Such a revealing price is always accepted by the seller. For it to be *perceived* incentive compatible the buyer's surplus needs to exceed her perceived information rent, i.e., the payoff from pretending to be a low type:

$$h - p_h \geq (1-\rho)(h - p_l). \tag{1}$$

This condition leads to a lower bound on the necessary amount of friction Δ for incentive-compatibility:

$$\Delta^\rho \geq -\ln(\rho + (1-\rho)\frac{p_l}{h}). \tag{2}$$

Such self-revelation in equilibrium requires sufficient friction where this friction is decreasing in the buyer's degree of information projection.

For the above equilibrium to generate excess revenue, the full surplus net the buyer's bargaining rent has to be greater than the static monopoly profit:

$$\ln(V_F) - \ln(V_M) \geq \Delta \tag{3}$$

This implies an upper bound on friction. Whenever the upper bound - Eq.(3) - exceeds the lower bound - Eq.(2) - the seller can boost his revenue above the static optimum.

More generally, suppose there is a true positive initial leakage probability, $\alpha > 0$. If the buyer perceives this correctly, the classic revenue result remains unaltered. This is true because the high type buyer will only reveal herself in a perfect Bayesian equilibrium if her bargaining rent $(1 - e^{-\Delta})h$ exceeds her true expected information rent $(1 - \alpha)(h - p_l)$. Given unbiased expectations, this implies that the bargaining rent of the buyer needs to exceed the value of the buyer's information rent to the seller for revelation to happen in equilibrium. Hence whenever the revelation condition is satisfied the revenue condition cannot be satisfied.

In contrast, a buyer who projects information *underestimates* her information rent. This perception relaxes her incentive compatibility constraint but not the seller's revenue constraints. Hence for the same reason as before, if ρ is sufficiently large and there is sufficient delay between the bargaining rounds, the seller's revenue now exceeds his static monopoly profit. The above logic implies the following straightforward corollary:

Corollary 1 *For all $\rho > \rho_A$, $V_A^\rho(\Delta)$ is maximal given positive but bounded friction.*

The above corollary points to a non-monotone comparative static result that is absent in the Bayesian case. In the unbiased case the seller's revenue is maximal under no friction. Given sufficient information projection, the seller's maximal revenue decreases both when delay is shorter and when it is longer than the optimal boundedly positive amount.

The above result may help explain various real-world bargaining tactics that may be harder to rationalize under standard assumptions. A seller who is aware of the buyer's tendency to suffer from the illusion of transparency might wisely engage in a "waiting game". Given non-trivial delay between offers, asking the buyer to name her own price first, the seller elicits information at a cost lower than the value of the information revealed. Anecdotal evidence suggests that such a bargaining practice may be common and successful in many settings.[10],[11]

3.2 Seller-Only Offers

Let me now briefly turn to a similar result under the protocol where the seller makes all offers. When commissioning an expert, a firm, uninformed about the expert's reservation wage, may

[10] Under the interpretation where it is time-preference and positive delay that introduces friction into bargaining, note that none of the result require the buyer to be impatient. Rather the results are based on the seller facing positive costs of delay so bargaining power can effectively be transferred to the buyer.

[11] Note that here considering only initial projection versus prospective projection is isomorphic.

first offer a lower wage to the candidate. If that offer is rejected, then the task may disappear, but with some probability the firm could return with a higher offer.

Let $V^{\rho}_{So}(\Delta)$ be the seller's true maximal revenue in equilibrium given a ρ-biased buyer and a one-sided protocol with friction Δ. The next proposition shows that if the buyer is sufficiently biased, this payoff exceeds the monopoly profit provided there is sufficient, but potentially bounded friction in the bargaining process.

Proposition 3 *For any $\rho > \rho_{So}$ and $\Delta \in (\widehat{\Delta}^{\rho}_{\min}, \widehat{\Delta}^{\rho}_{\max})$ it follows that $V^{\rho}_{So}(\Delta) > V_M$.*

The result above is based on the fact that if the buyer is sufficiently biased then there always exist an equilibrium with price-discrimination amongst the buyer types where the uninformed seller and the (fictional) informed seller *pool* together initially. Positive delay ensures that the buyer finds it credible that an informed seller would be willing to pool, as opposed to wait. This then allows the seller to pretend to be informed more often than he truly is and charge a higher initial price. Capitalizing on the buyer's mistaken belief in this fashion the seller can boost his revenue.

Specifically, consider first again the case with no true leakage, $\alpha = 0$. Suppose the uninformed seller and the (fictional) informed seller who knows that the buyer is a high type quote a price of $p_{1,h}$ initially. In order for this to be accepted by the high type, two conditions need to hold. First, the high type's perception of her continuation value conditional on rejecting the offer - her information rent - must be lower than the surplus offered:

$$h - p_{1,h} \geq (1-\rho)e^{-\Delta}(h-l), \tag{4}$$

since in equilibrium the uninformed seller will name price l in the second round and the (fictional) informed seller will offer no surplus. Second, it needs to be credible that $p_{1,h}$ could be offered by an informed seller:

$$p_{1,h} \geq e^{-\Delta}h, \tag{5}$$

ensuring that the informed seller prefers $p_{1,h}$ to waiting and collecting h in the next round. Combining the above two constraints, implies a positive lower bound on Δ. Friction thus must be positive.

The revenue result may also require fraction to be bounded from above. This is the case when $l > qh$, since here selling to the low type sufficiently early is important. In the case where $qh > l$, no such upper bound binds; for all Δ greater than the lower bound the revenue result holds.

More generally, given $\alpha > 0$, for dynamic profit to exceed the static one, one again needs to consider an equilibrium with price discrimination and initial pooling in the fashion described above. Absent such pooling, the high type buyer will always learn whether the seller is informed or not, and then revenue becomes separable across seller types leading to the standard revenue result.

Under unbiased expectations, the buyer will never accept a price that leaves her with a surplus smaller than her true discounted expected information rent. Hence the seller has to pay this rent as long as he maintains price discrimination in equilibrium. Since waiting is also costly for the seller, offering such a surplus is never beneficial for him relative to the revenue obtained from a single take-it-or-leave-it offer.

Under projection, the buyer underestimates her true information rent and is thus willing to accept a higher initial price, i.e., she accepts too soon. This is ensured whenever waiting is sufficiently costly for the seller making in credible that the initial offer could be coming from an informed seller who prefers to agree early rather than wait to reveal himself in the next bargaining round. By engaging in such a 'waiting game' the seller can pretend to be informed more often than he truly is and capitalize on the buyer's biased beliefs. The above logic again implies a straightforward corollary.

Corollary 2 *For any $\rho > \rho_{So}$, $V^{\rho}_{So}(\Delta)$ is maximal given positive but bounded friction.*

In the unbiased case the optimal revenue is achieved under no delay or infinite delay. In contrast, under information projection, positive but bounded delay will be the characteristic feature of the seller optimal bargaining protocol. Thus haggling over real time that is nominally costly for the *seller* is needed to achieve optimal revenue.

4 Infinite Horizon

Above I showed that the seller's revenue from dynamic haggling was higher than the Bayesian optimal revenue from a single take-it-or-leave-it offer provided the buyer was sufficiently biased. This was true despite the fact that projection was constrained to happen only at the beginning of the first round of bargaining, but of course extends to the case of prospective projection. I now turn to analyzing the impact of prospective projection, as described in Section 2, where I assume that $\rho_t = \rho$ for all t. I do so in the classic bargaining setup where the seller makes all the offers over a potentially infinite horizon and faces a continuum of buyer types whose valuation is distributed uniformly on $(0, 1]$ maintaining the so-called gap assumption. Bargaining lasts as long as the parties do not agree on an offer. For simplicity, below I analyze the standard case where in the true game the ex-ante leakage probability is zero.

In case prospective projection is almost full, it might be no surprise that this will change bargaining behavior in a way that the seller can obtain positive surplus even if Nature never actually reveals him the buyer's valuation. The seller might be able to capitalize on the buyer's fear that he might act on her valuation in the future even if she did not do it in the past. The results below show that the same is true even if the buyer believes that the probability that the seller will figure out her valuation by the next round is arbitrarily small.

I focus on stationary strategies.[12] These will be parametrized by a homogenous ρ and the delay Δ. Below, I describe both the imaginary and the real equilibrium paths. On the real path, the seller names an initial price of $\gamma(\rho,\Delta)$ and employs a linear pricing rule. The buyer employs a cutoff strategy and accepts price p iff her valuation θ is greater than $\lambda(\rho,\Delta)p$. Since the skimming property holds, at any time the seller always faces a left-truncation of the buyer's type distribution. On the imaginary path, the seller is believed to have a constant strategy holding the buyer to her valuation.[13] When hearing that she is 'identified', the buyer plans to accept immediately.

The above strategies imply that on the real equilibrium path, there will be a decreasing sequence of prices p_t for rounds $t \in \{1, 2, ...\}$ such that

$$p_t = \gamma(\rho,\Delta)^t \lambda(\rho,\Delta)^{t-1}$$

The following proposition re-states the well-known result for the unbiased case. As bargaining becomes smooth the seller loses all his static profit and in the limit sells immediately at a price equal to the lowest possible buyer valuation. Let $V_S^{\rho}(\Delta)$ be the seller's true expected equilibrium revenue given a ρ-biased buyer.

Proposition 4 *Suppose $\rho = 0$. There is a unique equilibrium. It is such that $\gamma(0,\Delta)$ is decreasing and $\lambda(0,\Delta)$ is increasing in Δ^{-1}. Furthermore, $V_S^0(\Delta)$ is decreasing in Δ^{-1} and $\lim_{\Delta\to 0} V_S^0(\Delta) = 0$.*

The comparative static properties of the unbiased case imply that as bargaining becomes smoother, the seller *lowers* his initial price and the buyer's demand withholding increases. Consider now the case with projection.

Proposition 5 *There exists a class of perfect ρ information projection equilibria such that,*

1. *Demand withholding $\lambda(\rho,\Delta)$ smoothly decreases in ρ for allΔ,*

2. *For any $\rho > 0$, there exists $\widetilde{\Delta}^{\rho} > 0$ such that $\gamma(\rho,\Delta)$ is increasing in Δ if $\Delta > \widetilde{\Delta}^{\rho}$, and $\gamma(\rho,\Delta)$ is decreasing in Δ if $\Delta < \widetilde{\Delta}^{\rho}$,*

3. *For any $\rho > 0$, and any $\tau > 0$, there exists $\widehat{\Delta}^{\rho}(\tau) > 0$ such that if $\Delta \in [0, \widehat{\Delta}^{\rho}(\tau)]$, then $V_S^{\rho}(\Delta)$ smoothly increases in Δ^{-1} and $\left|V_S^{\rho}(\Delta) - V_F\right| \leq \tau$.*

[12] Uniqueness of the unbiased Markov equilibrium in this setting follows from the general analysis in the literature, e.g., Sobel and Takahashi (1983), Fudenberg, Levine and Tirole (1985), Gul Sonnenschein and Wilson (1986).

[13] Except for a measure zero set of types, whose valuations correspond to prices on the real equilibrium path, where the seller offers a price slightly below the buyer's valuation.

Let me describe the above implications of the model. First, the buyer's willingness to accept smoothly increases in the degree of her projection: relative to the equilibrium prices the buyer accepts too soon. Second, the comparative static on the initial price is non-monotone. If bargaining friction is sufficiently high, the initial price decreases as friction decreases, the same way as in the unbiased case. For any positive degree of projection, however, as bargaining becomes sufficiently smooth this comparative static reverses. The initial price, as well as the entire price sequence, increases as delay or noise decreases. If the degree of projection is sufficiently large, this reverse comparative static is global. Finally, given any degree of projection, as delay or noise goes to zero, the seller's revenue smoothly increases and in the limit he extracts the full expected surplus from trade.

The logic of the above result is based on how information projection affects the relative strength of two countervailing effects. As bargaining becomes more smooth the buyer's willingness to wait for a price reduction from one period to the next increases and the seller's cost of postponing a sale from one period to the next decreases. While in the unbiased case the former increases faster than the latter, given any positive degree of information projection if friction drops below a threshold level the reverse is perceived to be true.

To provide intuition, let me proceed step-by-step. Note first that since the seller always correctly understands that leakage never happens, holding the buyer's strategy constant, his trade-offs are always the same whether or not the buyer projects information. The seller cannot just wait for information to arrive, instead, since the skimming property holds, he still becomes more pessimistic after each rejection by the buyer and reduces the price over time.

For the seller's stationary intertemporal incentive-constraint on such price reduction per round to hold, one needs to compare the benefit versus the cost of reducing the price already in a given round or only in the next. The cost is the intertemporal price difference times the probability that the buyer will buy now as opposed to at a lower price tomorrow. This thus depends on the extent to which the buyer is willing to withhold demand. The benefit is the forgone interest loss on the seller's continuation value. To maintain price-discrimination, the cost has to exceed the benefit. The logic is then linked to how information projection affects this constraint.

In the unbiased case, holding the seller's strategy constant, as bargaining becomes more smooth it becomes increasingly easy for higher types to behave as lower ones do. The buyer's willingness to withhold demand increases and the extent to which the seller can extract surplus through price discrimination decreases. This is the classic force. For price discrimination to hold, it then has to be true that the ratio of the seller's initial price over the seller's continuation value must still exceed the extent to which the buyer withholds demand. As bargaining becomes smooth demand-withholding for any positive price-discrimination would be unbounded, hence the continuation value must go to zero. It follows that the seller's initial price converges to the lowest possible valuation of the buyer.

Given a small degree of information projection the same classic force is present in that the cost of waiting from one round to the next eventually disappears with a probability arbitrarily close to one. At the same time, a countervailing force is now also present. Since holding the seller's strategy constant a higher type has relatively more to lose from leakage than a lower one, this makes price discrimination easier. This force also becomes stronger as bargaining becomes more smooth.

Proposition 5 first establishes a non-monotone comparative static result. Initially, as friction decreases the classic force is stronger: the buyer's cost of waiting decreases faster than the seller's cost of waiting. This implies that the buyer's demand withholding increases and the seller lowers his initial price. There is always a critical value of Δ^{ρ}, however, such that the relative strength of the two countervailing forces switches. Past this point, which is *increasing* in the degree of information projection, the seller's *perceived* cost of waiting - as perceived by the ρ-biased buyer - decreases more slowly. Now the buyer's demand withholding starts to decrease and the seller starts to take a stronger position raising the initial price.

The above proposition establishes that while both forces become stronger as bargaining becomes smoother, eventually the force due to projection starts to dominate. As friction decreases, the extent to which different buyer types are willing to accept different prices increases due to projection. In the limit, the incentive-constraint that was binding in the unbiased case is always slack. Since the per-period price-gradient on the equilibrium path converges to null, irrespective of the degree of projection, full rent extraction follows.

Note that the mechanical utility consequence of prospective projection goes smoothly to zero as the degree of projection becomes smaller and smaller. While given prospective projection the buyer does wrongly believe that after infinitely many rounds the seller learns the buyer's valuation, the period-to-period perception is arbitrarily small, and the discounted sum of these probabilities, for any positive delay, smoothly goes to zero as the degree of projection decreases. Yet, the reversal of the comparative static above holds for positive delay bounded away from zero given *any* positive degree of projection.

Finally, Proposition 5 does rely on the fact that the seller disagrees with the buyer. The seller knows that he can only learn about the buyer's valuation from whether she accepts or not. At the same time, since he understands the buyer's illusory perception he can capitalize on it.

4.1 Vanishing Projection

The revenue result above is based on the psychologically realistic assumption that the buyer believes that there is some - no matter how tiny - probability that her opponent acts on the basis of her valuation, even if he did not do so before. Given that each bargaining round is separate, this fits tightly with the interpretation of information projection described before. The logic above suggests, however, that an even weaker assumption may suffice allowing projection to vanish as delay vanishes.

In this fashion, suppose that as delay or noise goes to zero, $\Delta \to 0$, a person's biased perception also converges to zero, $\lim_{\Delta\to 0}\rho(\Delta) = 0$. For example, as offers become more frequent, the buyer attaches a smaller and smaller probability to leakage in a given round allowing for this perception to converge to zero. Alternatively, since the result below is a limit result with a bound on the speed at which projection goes to zero, this assumption can be made consistent with a case where a biased buyer might "update" and believe that leakage is less and less likely to happen in the future if it did not happen in the past and eventually converges to zero.

The next result establishes, given the above described stationary strategies, an exact threshold on the relative speed at which projection vanishes as bargaining becomes smooth. If the relative speed is faster than this threshold, then we get back the Coasian result given stationary strategies. If it is slower, then one still obtains full rent extraction. To notationally highlight the dependence of ρ on Δ, I now make the former also an argument of the expected revenue.

Proposition 6 *Suppose $\rho = \beta(\Delta)^{\kappa}$. For all $\beta > 0$,*
If $\kappa > 0.5$, $\lim_{\Delta\to 0} V_S(\Delta, \rho(\Delta)) = 0$.
If $\kappa = 0.5$, $\lim_{\Delta\to 0} V_S(\Delta, \rho(\Delta)) = V_F\beta/(1+\beta)$.
If $\kappa < 0.5$, $\lim_{\Delta\to 0} V_S(\Delta, \rho(\Delta)) = V_F$.

For the class of functions parametrized above, Proposition 6 identifies a discontinuity. If projection vanishes slower than the square-root of the delay, then the speed at which the seller's cost of waiting decreases is faster than the speed at which the buyer's perceived cost of waiting decreases as bargaining becomes smooth. This implies full rent extraction. If projection vanishes faster than the square-root, the opposite holds. Hence the seller cannot price discriminate and is left with no profit in the limit. Finally, in the knife-edge case where discontinuity happens, $\kappa = 0.5$, the distribution of the surplus depends on the scaling parameter β. Here the seller obtains a $(\beta+1)/\beta$ fraction of the expected surplus. For example, when $\beta = 1$, his expected revenue is exactly half of the expected surplus from trade.

Note that except for the case where $\kappa = 1$, the limiting *total* undiscounted probability of perceived leakage per unit of real time, as delay goes to zero, is independent of $\kappa < 1$.[14] Hence the result also shows that discontinuity does not hinge on this limiting probability which is the same in both cases.

Finally, note that the above result does not derive simply from the buyer and the seller implicitly facing different interest rates. Even when interest rates differ, but the buyer is unbiased, the Coasian result follows.[15] I now turn to the existing evidence on the above game.

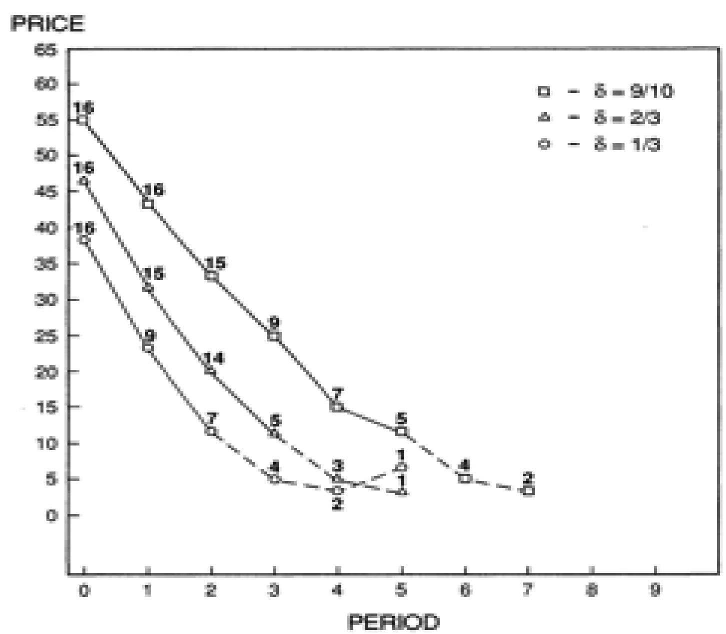
[14] For $\kappa = 1$ the total undiscounted probability of leakage per unit of time in the limit is $1 - \lim_{\Delta\to 0}(1-\beta\Delta)^{1/\Delta} = 1 - e^{-\beta}$, for *all* $\kappa < 1$, it converges to 1.

[15] The Appendix provides a proof of this statement.

4.2 Comparative Statics

The above result implies not only a different revenue result in the limit, but also different comparative static predictions. This renders the model testable away from the limit. Specifically, the comparative static predictions of the biased and the unbiased case differ. In the unbiased case as Δ decreases the seller charges lower initial prices. In contrast, under projection there is a threshold value of Δ^{ρ} such that if Δ drops below this threshold, the process reverses and the seller charges higher initial prices and obtains a larger revenue.

This allows me to relate the predictions to the existing experimental evidence on the setting studied above. Rapoport, Erev, and Zwick (1995) report an experiment with time discounting in a game that corresponds precisely to the game studied above. The buyer's private valuation is drawn uniformly from $(0, 100]$ and is never revealed to the seller. The seller makes all offers.[16] The setting is made common knowledge. The treatment variable is the discount rate $\delta = e^{-\Delta}$ varying between 0.33 and 0.9. Some of the findings are summarized in the figure below describing one of their 9 sessions:



Rapoport, Erev and Zwick (1995)

As Rapoport et al. (1995) observe, the effect of the treatment variable is the *opposite* of the unique Bayesian prediction. Their results show that: (i) buyers accept offers too soon, and

[16] The game terminated either when the buyer accepted the seller's offer or - to deal with finite experimental time - when the highest possible discounted profit became smaller than the smallest unit of the currency used, i.e, 1.

do so even after experience, (ii) sellers' initial prices increase as opposed to decrease in δ. This comparative static result is the reverse of the unbiased predictions, but is consistent with the predictions of Proposition 5. Furthermore, although in perfect Bayesian equilibrium the seller's revenue should decline in δ this prediction is also rejected by the data.

For the highest discount factor, the average initial price was above the static monopoly price, contrary also to the point prediction of the unbiased perfect equilibrium, but again consistent with a perfect equilibrium with information projection. The minimally necessary value of projection above which the reverse comparative static holds for all implemented treatment variables is $\rho \geq 0.32$. Basic calibrations based on the reported average data in Rapoport et al. show that initial prices across the higher discount factor treatments are consistent also with the point predictions of the model given significant prospective projection which is higher in the sessions without experience and smaller but still substantial in the sessions with experience.[17]

5 Conclusion

The goal of this paper was to explore the consequences of information projection to dynamic bargaining with one-sided incomplete information. I have demonstrated that this phenomenon impacts bargaining behavior. Both in a setting with only initial projection and in a setting with prospective but vanishing projection bargaining over real time without commitment was key to boost the seller's revenue above the Bayesian optimal mechanism. Since these two settings differ in important aspects, it will be worthwhile to extend the results to other settings. Given the portable way dynamic information projection is introduced in this paper, future research can generalize the results and also investigate the impact of this phenomenon in a variety of other bargaining contexts.

It might also be interesting to study the implications of information projection to mechanism design more generally. Since bargaining with private information lies at the heart of many economic settings, such as monopoly theory or dynamic contracting, e.g., Hart and Tirole (1988), the results have potential implications to other problems as well, such as the use of privacy versus personalized pricing in the repeat sale of a non-durable goods.

An immediate corollary of the above result is that this phenomenon provides a channel

[17] Rappoport et al. (1995) report average initial prices from sessions without experience (Iterations 1-3) and from sessions with experience (Iterations 7-9), (see pp. 384). For the lowest discount factor in Iterations 7-9 neither the unbiased nor the biased model may rationalize the data, as initial prices are too low. For the other discount factors the model can calibrate the data with significant information projection both for (Iterations 1-3) and (Iterations 7-9). Specifically, in sessions without experience (Iterations 1-3) assuming a $\rho \in [0.43, 0.44]$ matches the average initial prices reported for *both* discount factors, i.e., $\delta \in \{2/3, 0.9\}$. In sessions with experience (Iterations 7-9) assuming a $\rho \in [0.27, 0.32]$ matches the average initial prices reported for *both* discount factors, i.e., $\delta \in \{2/3, 0.9\}$. The above calculations rely on the fact that the equilibrium price-gradient is independent of $\rho \in [0, 1]$. Importantly, the authors find that the empirical price-gradient matches the theoretical predictions well. In these calibrated ranges of ρ, the model's comparative static predictions are the reverse of the Bayesian one, i.e., $\rho = 0$.

through which engaging in one-on-one negotiations with a buyer will be a desired practice by the monopolist. This is true because haggling without commitment - even over short periods of time -now allows the seller to take advantage of information projection and boost his revenue above the full commitment monopoly solution. In contrast to the Coasian logic, personalized haggling is a way to increase monopoly power. For example, a car dealerships that has a practice where the salesman haggles with the buyer might then perform better than non-haggling car dealerships where there is a commitment that upon meeting the buyer, the seller will always just quote a single price, make a single take-it-or-leave-it offer. In the context of selling new cars, Goldberg (1996) provides evidence that there is considerable room for real negotiations in dealerships leading to significant variation in the discounts - deviations between the actual price and the list price - received by different consumers and that such discounts are uncorrelated with easily observable socio-economic characteristics. Furthermore, evidence shows that such a practice might benefit the seller and lower the buyer's surplus relative to practices that are closer to making a single offer, e.g., Scott Morton, Zettelmeyer and Silva-Risso (2001). Future research could address the extent to which the model's predictions may hold in the field.

6 Appendix

6.1 Appendix A

I now formally define the perfect ρ projection equilibrium described in the text. Let $\Gamma = \{H, P, I_k, f_N, v_k\}$ be a sequential bargaining game with observable moves by players: H is the set of histories, Z the set of terminal histories, $P : H \backslash Z \to \{S, B, N\}$ the assignment function to the seller, the buyer, and Nature. For each player $k \in \{S, B\}$, I_k is the information partition on $\{h \in H : P(h) = k\}$ with generic element $\iota_k \in I_k$, $f_N(\cdot \mid h)$ is the probability measure describing the distribution of Nature's chance move following history h such that $P(h) = N$. In the true games studied, Nature moves only initially assigning the players' true information independently according to $f_N(\mid \emptyset)$. Finally, $v_k : Z \to \mathbb{R}$ is player k's payoff function over terminal nodes.

Consider an extended game Γ_+ derived from Γ purely by adding chance moves by Nature but leaving all other aspects of Γ unchanged. Specifically, at the beginning of each round $t = 1, 2, ..$ Nature's move is given by a binary random variable $\epsilon_t \in \{0, 1\}$. If $\epsilon_t = 1$, θ is revealed to the seller. If $\epsilon_t = 0$, it is not. The realization of ϵ_t is observed only by the seller. The realizations of ϵ_t and $\epsilon_{t'}$ are independent for any $t \neq t'$. The seller's ex-ante belief is that $\Pr(\epsilon_t = 1) = 0$ for all t. A ρ-biased buyer's ex-ante biased belief is that $\Pr(\epsilon_t = 1) = \rho_t$.

Formally, let μ_k refer to player k's system of beliefs in Γ_+ describing both the beliefs at each of player k's information set and this player's beliefs about the distribution of Nature's moves. To emphasize the dependence of the latter on ρ, let μ_B^ρ be the buyer's system of beliefs, and μ_S^0 the sellers. Let σ_k denote a behavioral strategy map of player k in Γ_+.

Let's divide the collection of the seller's information sets in Γ_+ into two disjoint subsets: I_S^0 collects the sets where all observed realizations of ϵ_t are zero, I_S^+ is the complement set. Since by time t the seller observes the realizations of all ϵ_s such that $s \leq t$, this division is well-defined. Note that at any information set in I_S^0, given μ_S^0, the seller believes that the probability of reaching any information set in I_S^+ is zero. The seller's *conditional* beliefs at such information sets are still well-defined and will be derived from the players' strategies. Note also that each information set in I_S^+ is a singleton, and since there is perfect recall in the game, the seller's beliefs about Nature's future leakage moves once at an information set in I_S^+ are inconsequential.

Note that there is a one-to-one mapping between the buyer's information sets in Γ_+ and those in Γ containing equivalent observations of the players' sequence of actions. Similarly, for the seller there is a one-to-one mapping from the information sets in I_S^0 in Γ_+ to the seller information sets in Γ containing equivalent observation of the players' sequence of actions. With a slight abuse of notation, I will refer to the restriction of $\sigma = \{\sigma_B, \sigma_S\}$ in Γ_+ given these mappings to Γ as the appropriate restriction of a σ in Γ.

Given a belief system μ_k and a strategy profile σ the value function $V_k(\cdot \mid \iota_k)$ denotes the expected utility from the lottery induced over terminal nodes for player k conditional on being at information set $\iota_k \in I_k$.

Definition 1 *The appropriate restriction of $\sigma^* = \{\sigma_B^*, \sigma_S^*\}$ to Γ is a perfect ρ projection equilibrium if there exists $\{(\sigma_B^*, \sigma_S^*), (\mu_B^\rho, \mu_S^0)\}$ in Γ_+ such that*

1. $V_B(\sigma_B^*, \sigma_S^*, \mu_B^\rho \mid \iota_B) \geq V_B(\sigma_B, \sigma_S^*, \mu_B^\rho \mid \iota_B)$ *for all σ_B and at all $\iota_B \in I_B$.*

2. $V_S(\sigma_B^*, \sigma_S^*, \mu_S^0 \mid \iota_S) \geq V_S(\sigma_B^*, \sigma_S, \mu_S^0 \mid \iota_S)$ *for all σ_S and at all $\iota_S \in I_S^0$ and all $\iota_S \in I_S^+$.*

3. *For all $i_B \in I_B$, $\mu_B^\rho(i_B)$ is derived from Bayes' rule (whenever possible) given (σ_B^*, σ_S^*) and the assumption that $\Pr(\epsilon_t = 1) = \rho_t$ for any t.*

4. *For all $i_S \in I_S^0$, $\mu_S^0(i_S)$ is derived from Bayes' rule (whenever possible) given (σ_B^*, σ_S^*) and the assumption that $\Pr(\epsilon_t = 1) = 0$ for any t.*

6.2 Appendix B

Proof of Proposition 1. The seller's expected revenue given a single take-it-or-leave-it offer in $t = 1$ is $V_M = \alpha V_F + (1 - \alpha) \max\{qh, l\}$. If $\alpha = 0$, the result follows from Skreta (2006). Consider the four possible sequential bargaining protocols. If the buyer makes all offers, the seller's payoff is zero. If they alternate and the seller makes the first offer, the informed seller cannot get a payoff greater than $(1 - e^{-\Delta})V_F$ and the uninformed seller cannot get a payoff greater than $(1 - e^{-\Delta}) \max\{qh, l\}$, because any higher offer will be rejected by the respective buyer type.

Alternating Offers. Consider the protocol where the buyer makes the first offer. Three types of equilibria can arise: separating, where different buyer types name different initial prices; semi-separating, and pooling.

Consider full separation where the seller accepts the buyer's price. The high (low) type buyer at $t = 1$ names p_h (p_l).The informed type always names θ. Note that $p_h = e^{-\Delta}h$ must hold, since any lower revealing price will be rejected given revelation, and any higher price will be accepted given the most optimistic belief the seller can have. Furthermore, $p_l \in [e^{-\Delta}l, \min\{l, p_h\}]$, since any lower price will be rejected, and any higher price will violate individual rationality or separation. A tighter upper bound on p_l may hold, but considering it will just strengthen the argument below.

For buyer separation to be incentive compatible, the following IC constraint must be satisfied:

$$(1 - e^{-\Delta})h \geq (1 - \alpha)(h - p_l) \tag{6}$$

Re-writing this, one gets a lower bound on Δ:

$$\Delta > \Delta_{\min} = \ln h - \ln(\alpha h + (1 - \alpha)p_l) \tag{7}$$

Consider now the seller's revenue. Suppose $l \geq qh$. Straightforward algebra shows that for $V_A^0(\Delta) \geq V_M$ to hold, given a binding IC constraint, it must be that:

$$\Delta \leq \Delta_{\max} = \ln(qh) - \ln(l - p_l(1 - q) + q\alpha(h - l))$$

Note, however, that since $p_l \leq l$, $\Delta_{\max} \leq \Delta_{\min}$. Suppose $qh > l$. Straightforward algebra shows again that for $V_A^0(\Delta) \geq V_M$ to hold, it must be that:

$$\Delta \leq \Delta_{\max} = \ln(qh) - \ln\left(l\alpha(1 - q) + qh - p_l(1 - q)\right)$$

Note, however, that since $(qh - l)(1 - \alpha) \geq (1 - q\alpha)(p_l - l)$ as long as $p_l \leq l$, again $\Delta_{\max} \leq \Delta_{\min}$. Note that if there was a fully revealing equilibrium where the seller did not accept the buyer's price, then by revelation the upper-bound on the seller's revenue remained unchanged.

Consider semi-separation with serious offers. The relevant case is where the high type mixes between revelation at p_h and pooling with probability y with the low type at p_l. Again $p_h = e^{-\Delta}h$ must hold where this price is accepted. If the uninformed seller does not mix following p_l the revenue result follows from the above discussion since Suppose he accepts p_l with probability z. We have two cases.

I. If $p_2 = l$, then $p_l = e^{-\Delta}l$ must hold. For the high type buyer to mix it must then be that

$$(1 - e^{-\Delta})h = (1 - \alpha)(z(h - e^{-\Delta}l) + (1 - z)e^{-\Delta}(h - l)) \tag{8}$$

Note that maximal revenue is affected by z only through the constraint on Δ. Specifically, the seller's expected revenue is:

$$\alpha e^{-\Delta}(qh + (1-q)l) + (1-\alpha)((q-qy)e^{-\Delta}h + (1-q+qy)e^{-\Delta}l)$$

The minimal necessary separating friction $\underline{\Delta}(z)$ is then given when $z = 0$. Straightforward algebra shows that even this $\underline{\Delta}(0)$ is such that $e^{-\underline{\Delta}(0)}V_F \leq V_M$, hence the same is true for all $z > 0$.

II. If $p_2 = h$, it needs to be that $qh > l$. For the high type to mix then

$$(1 - e^{-\Delta})h = (1-\alpha)z(h - p_l) \tag{9}$$

must hold, where

$$p_l = \frac{qy}{qy + (1-q)}e^{-\Delta}h$$

for the uninformed seller to mix. It is easy to see that $V_A^0(\Delta) \leq hqe^{-\Delta} + \alpha(1-q)e^{-\Delta}l \leq qh + \alpha(1-q)l = V_M$.

Consider finally a pooling equilibrium. Here, $V_A^0(\Delta) \leq \max\{l, \alpha(1-q)l + e^{-\Delta}qh\} \leq V_M$. Finally, note if there is an equilibrium where the buyer makes a non-serious offer, the revenue result holds a fortiori. .

Proof Proposition 2. Consider a fully separating equilibrium. Given the constraints in Proposition 1, one can keep α as it is in the revenue condition, and replace α with $\alpha^\rho = \alpha + (1-\alpha)\rho$ in the separation condition. Before comparing the revenue and the new separation conditions, note that in a fully separating perfect equilibrium where both announcements are accepted, the low type's announcement must satisfy:

$$p_l \leq l(1-\alpha^\rho) + \alpha^\rho l e^{-\Delta}$$

This is true because the informed seller must accept any price greater than $e^{-\Delta}l$ from the low type in equilibrium. Consider now the revelation condition:

$$(1 - e^{-\Delta})h \geq (1-\alpha^\rho)(h - p_l)$$

Re-arranging this, we get that

$$\Delta \geq \Delta^\rho_{\text{min}} = \ln h - \ln(\alpha^\rho h + (1-\alpha^\rho)p_l).$$

This constraint implies a positive lower bound for all $\rho < 1$. Furthermore, since α^ρ is increasing in ρ, Δ^ρ_{min} is decreasing in ρ with $\Delta^1_{\text{min}} = 0$. The revenue conditions are the same as before, in particular $\Delta_{\max} > 0$. It then follows that if $p_l = le^{-\Delta}$ and ρ is sufficiently high, then $\Delta^{\rho}_{\min} < \Delta_{\max}$ for all $\rho > \rho_A$.

Note that since $\Delta_{\max}$ is increasing and $\Delta^{\rho}_{\min}$ is decreasing in p_l, a higher admissible p_l will enlarge the range of frictions for which this equilibrium generates a revenue greater than V_M, yet both but remain positive and bounded for all $\rho \in (0, 1)$, given the constraint on p_l. To then see the corollary, note that above I considered fully separating equilibria. It is easy to see that under $\Delta = 0$ or $\Delta = \infty$ no pooling or semi-separating equilibrium can generate a revenue higher than V_M for any $\rho < 1$. .

Proof of Proposition 1 (Continued). Seller-Only Offers. Suppose the seller makes all offers. Two facts must hold for $V_{So}(\Delta)$ to potentially exceed V_M. First, the uninformed seller has to sell to both types at different prices with positive probability. Second, there has to be pooling between the uninformed seller and the informed seller conditional on $\theta = h$. The latter is true because otherwise the high type buyer always learns whetherr the seller is informed or not and hence the uninformed seller's pricing strategy must satisfy the same IC constraints as if $\alpha = 0$. Hence it is sufficient to consider equilibria where in $t = 1$ the informed seller (given $\theta = h$) pools with the uninformed seller. Let the pooling price named by the uninformed seller and informed seller conditional on $\theta = h$ at $t = 1$ be $p_{1,h}$.This implies that the informed seller conditional on $\theta = l$ separates. Finally, note that the informed seller has a strictly dominant strategy in $t = 2$. In $t = 1$ the informed seller must name a price between $e^{-\Delta}\theta$ and θ, or do not sell.

Consider a pure equilibria. Let the price named by the informed seller conditional on $\theta = l$ be $p_{1,l}$. At $t = 2$, the uninformed seller's price is p_2 which here must equal l. For the informed seller to pool on $p_{1,h}$,

$$p_{1,h} \geq e^{-\Delta} h \tag{10}$$

must hold. For the high type buyer to accept $p_{1,h}$ the following IC constraint must hold:

$$h - p_{1,h} \geq (1 - \alpha)e^{-\Delta}(h - l) \tag{11}$$

Furthermore, $p_{1,l} = l$ under the optimal revenue. Hence given a binding IC constraint from Eq.(11) an upper-bound on the seller's revenue in such an equilibrium is:

$$\widehat{V}^0_{So}(\Delta) = q(h - (1 - \alpha)e^{-\Delta}(h - l)) + (1 - q)((1 - \alpha)e^{-\Delta}l + \alpha l)$$

We then get that

$$\begin{aligned} \widehat{V}^0_{So}(\Delta) - V_M &= \left(e^{-\Delta} - 1\right)(1 - \alpha)(l - qh) \leq 0 \text{ if } l \geq qh \qquad (12) \\ \widehat{V}^0_{So}(\Delta) - V_M &= e^{-\Delta}(1 - \alpha)(l - qh) \leq 0 \text{ if } qh > l \end{aligned}$$

thus dynamic revenue is bounded by the static monopoly payoff for all α and Δ in this class.

Consider mixed equilibria. The IC constraint in Eq.(11) can now be relaxed if the uninformed seller mixes between h and l in $t = 2$. For this to work, $qh > l$ must hold. It then also has to be that the high type buyer mixes between accepting and rejecting $p_{1,h}$ and thus

$$h - p_{1,h} = (1 - \alpha)ke^{-\Delta}(h - l)$$

where k is the probability that the uninformed seller names l in $t = 2$. For the uninformed seller to mix in $t = 2$, it has to be true that $\frac{q(1-j)}{q(1-j)+(1-q)} = l/h$ where j is the probability that the high type buyer accepts $p_{1,h}$. This is independent of α. Hence if $k = 0$, then $p_{1,h} = h$, and straightforward calculation shows that the seller's expected dynamic revenue is bounded by V_M. Since $p_{1,h}$ is maximal for $k = 0$, and all other prices and j are independent of k, the same holds for all k provided $p_{1,h} \geq e^{-\Delta}h$ so that the informed seller is still willing to pool. Hence again revenue is bounded by V_M for all α. It is easy to see that considering mixing by the informed seller or the uninformed seller in $t = 1$ will not relax the high type's IC constraint and does not boost the seller's revenue above $\widehat{V}^0_{So}(\Delta)$.

Proof of Proposition 3. Consider a pure equilibrium with price discrimination and pooling between the uninformed and informed seller types (given $\theta = h$) in $t = 1$ as before. Let $p_2 = l$. For polling to hold it has to be true that:

$$p_{1,h} \geq \max\{e^{-\Delta}h, p_{1,l}\} \tag{13}$$

For the ρ-biased high type to accept $p_{1,h}$ the perceived IC constraint is:

$$h - p_{1,h} \geq e^{-\Delta}(1-\alpha)(1-\rho)(h - p_{1,l}) \tag{14}$$

Combining Eq.(13) and Eq.(14) one gets the following constraint:

$$\Delta \geq \widehat{\Delta}^{\rho}_{\min} = \ln(h + (1-\alpha)(1-\rho)(h - p_{1,l})) - \ln h \tag{15}$$

where $\widetilde{\Delta}^{\rho}_{\min} > 0$ for all $\rho < 1$.[18]

The seller's expected revenue here given a binding perceived IC constraint and setting $p_{1,l} = l$ is given by:

$$V^{\rho}_{So}(\Delta) \geq q(h - (1-\rho)(1-\alpha)e^{-\Delta}(h - l)) + (1-q)((1-\alpha)e^{-\Delta}l + \alpha l) \tag{16}$$

It is easy to see that if $\rho = 1$, then $V^1_{So}(0) = V_F$.

[18] Note that the sepration constraint is always satisfied since $p_{1,h} \geq l + (h - l)(\alpha + \rho - \alpha\rho)$.

Suppose $qh > l$, then $V_{So}^{\rho}(\Delta) > V_M$ holds if $\rho > \rho_{So} = (hq - l)/(hq - lq)$ and $\Delta \geq \widehat{\Delta}_{\min}^{\rho}$. Suppose $l \geq qh$, then $V_{So}^{\rho}(\Delta) > V_M$ whenever $(1 - e^{-\Delta}(1 - q\rho))l < (1 - e^{-\Delta}(1 - \rho))qh$, which is true if $\Delta \leq \widehat{\Delta}_{\max}^{\rho} = \ln(l(1 - q\rho) - hq(1 - \rho)) - \ln(l - hq)$. Hence we need that $\widehat{\Delta}_{\max}^{\rho} > \widehat{\Delta}_{\min}^{\rho}$, which is equivalent to $\rho > \rho_{So} = (l - hq)(l + hq\alpha/(1 - \alpha))^{-1}$. Hence if $\Delta \in (\widehat{\Delta}_{\min}^{\rho}, \widehat{\Delta}_{\max}^{\rho})$ and $\rho > \rho_{So}$, the revenue result holds.

Note that if $\Delta = \infty$, then the uninformed seller's maximal revenue is bounded by $\max\{qh, l\}$. If $\Delta = 0$, then the informed seller does not accept a price lower than θ. Hence a pooling would require $p_{1,h} = h$ which implies that price-discrimination cannot be incentive-compatible for any $\rho < 1$. Finally, note that in the above construction $V_{So}^{\rho}(\Delta)$ is maximal at $\widehat{\Delta}_{\min}^{\rho}$. This is true because $\partial V_{So}^{\rho}(\Delta)/\partial\Delta = e^{-\Delta}(\alpha - 1)(l - hq + hq\rho - lq\rho) < 0$ whenever $l \geq hq$ and provided that $\rho > \rho_{So}$ also when $qh > l$.

Proof of Proposition 5. Note that given the stationary strategies described the skimming property holds on the real path. This implies that in the beginning of period t, the true uninformed seller will face a uniform truncation of the type distribution, $[0, \theta_t]$. On the real path the seller will offer a price of $p_t = \gamma\theta_t$ where $\gamma \leq 1$, and the buyer of type θ accepts a price of p iff $p \leq \frac{1}{\lambda}\theta$. Given dynamic optimization by the real seller, who knows that Nature never leaks the buyer's valuation to him, under stationarity the following maximization needs to be true. Given any state variable θ_t at any round t,

$$V_s(\theta_t) = \max_{p_t}\{(\theta_t - \lambda p_t)p_t + e^{-\Delta}V_s(\lambda p_t)\}$$

since $\theta_{t+1} = \lambda p_t$. Taking the first-order condition with respect to p_t, we get that:

$$\theta_t - 2\lambda p_t + e^{-\Delta}\lambda V_s'(\lambda p_t) = 0$$

where from the envelop theorem, given the buyer's strategy and optimality, it follows that

$$V_s'(\theta_{t+1}) = \gamma\theta_{t+1}$$

hence,

$$\theta_t - 2\lambda p_t + e^{-\Delta}\lambda^2\gamma p_t = 0$$

Consider now the ρ-biased buyer's strategy on the real equilibrium path. In any round t - where the price does not equal her reservation value - the buyer will be indifferent between accepting the current price and rejecting the offer if her value is, $\theta_{t+1} = \lambda p_t$, and thus,

$$\lambda p_t - p_t = e^{-\Delta}(1 - \rho)(\lambda p_t - \gamma\lambda p_t)$$

must hold, since she believes that with probability ρ she might be detected in the next round.

We thus have two equations for the parameters λ and γ, from which we then obtain:

$$\lambda(\rho,\Delta) = \frac{1}{e^{-\Delta}(\rho-1)+1}\left(\rho(1-\sqrt{1-e^{-\Delta}}) + \sqrt{1-e^{-\Delta}}\right)$$
$$\gamma(\rho,\Delta) = \left(1-e^{-\Delta}(1-\rho)\right)\frac{1-e^{-\Delta}(1-\rho)-\sqrt{1-e^{-\Delta}}}{e^{-2\Delta}(1-\rho)^2+e^{-\Delta}(2\rho-1)}$$

Let's return to the fictional equilibrium path ensuing a positive leakage event. To specify the buyer's strategy note that by naming θ the imaginary informed seller and the uninformed seller always separate - except for an ex-ante measure zero set of types. For the set of measure zero types whose valuation is part of the support of the real seller's equilibrium price vector, the imaginary seller is believed to name a specific price that is arbitrarily smaller than the buyer's valuation b, but greater than $e^{-\Delta}b$, to separate. Since given the gap assumption bargaining terminates in finite time, this is well-defined and for all $\Delta > 0$, such a price exists. Thus in each round, the buyer knows the type of the seller in that period, hence whenever she is named a price that comes from an informed seller it is a best-response to accept.

The real seller would never want to deviate to such a price of the imaginary seller given that the real seller's belief never has atoms. Finally, off-equilibrium path beliefs of the buyer, following any deviation by the seller, assign full weight to these originating from the uninformed seller. The buyer here will follow her stationary cut-off strategy. This is optimal since the continuation value from rejecting remains unchanged for the buyer, since she believes that leakage still happens with probability ρ next round. This ensures that the real seller never wants to deviate. It is easy to check that given the above conditions, and since $\lambda\gamma$ is constant in ρ, the second-order condition on the value function is satisfied.[19]

To show that demand withholding is decreasing in ρ, note that

$$\lambda_\rho(\rho,\Delta) = -\frac{1}{\left(e^{-\Delta}(\rho-1)+1\right)^2}\left(\sqrt{1-e^{-\Delta}} - (1-e^{-\Delta})\right) < 0$$

with $\lim_{\rho\to1}\lambda(\rho,\Delta) = 1$. Furthermore, note that while $\lambda_\Delta(0,\Delta) < 0$ is true, for any $\rho > 0$, there exists a threshold $\Delta^{**}(\rho)$ such that if $\Delta < \Delta^{**}(\rho)$, then $\lambda_\Delta(\rho,\Delta) > 0$,4 and thus here the degree of demand withholding increases rather than decreases with the delay.

[19] To see that in the case $\rho = 0$, the results extend to different positive interests rate, let's normalize the interest rate faced by the seller to be 1 and let's denote the interest rate of the by buyer to be $b > 0$. The parameters of the stationary equilibrium are then given by the following system:

$$\theta_t - 2\lambda p_t + e^{-\Delta}\lambda^2\gamma p_t = 0$$
$$\lambda p_t - p_t = e^{-\Delta b}(\lambda p_t - \gamma\lambda p_t)$$

The solution again satisfies the second-order condition, and the initial price is given by $\gamma^b(\Delta,0) = -\left(e^{b\Delta}-1\right)\frac{e^{\Delta}+e^{b\Delta}\sqrt{e^{\Delta}\left(e^{\Delta}-1\right)}-e^{b\Delta}e^{\Delta}}{e^{\Delta}+e^{2(b\Delta)}-2e^{b\Delta}e^{\Delta}}$ and it is easy to see that $\lim_{\Delta\to0}\gamma^b(\Delta,0) = 0$ for all $b > 0$.

Consider now the impact of a change in Δ on the initial price γ. To make calculations more concise, I switch notation and denote $e^{-\Delta}$ by δ. Note that in the unbiased case, $\frac{\partial\gamma(0,\delta)}{\partial\delta} < 0$, while in the fully biased case $\frac{\partial\gamma(1,\delta)}{\partial\delta} > 0$ for all δ. To show that for any $\rho > 0$ there exists $\widetilde{\delta}^{\rho}$such that for $\delta > \widetilde{\delta}^{\rho}$ $\frac{\partial\gamma(\rho,\delta)}{\partial\delta} > 0$, note first that

$$\gamma_{\delta}(\rho,\delta) = \overbrace{\frac{1}{2\delta^2\sqrt{1-\delta}}}^{I} * \overbrace{\frac{\delta(1-\rho)+2\rho-1}{(\delta+2\rho-2\delta\rho+\delta\rho^2-1)^2}}^{II} * \overbrace{\left(\delta^2(1-\rho)^2+2(1-\delta+\delta\rho)(1-\sqrt{1-\delta})-\delta)\right)}^{III}$$

We then have that $\lim_{\delta\to 1}\frac{\partial\gamma(\rho,\delta)}{\partial\delta} > 0$ iff $\rho > 0$, and $\gamma_{\delta}(\rho,\delta) \geq 0$ iff $\delta \geq \max\{\frac{1-2\rho}{1-\rho}, 0\}$. To see the last point, note first that for any given ρ, $\gamma_{\delta}(\rho,\delta) = 0$ at $\delta = \max\{\frac{1-2\rho}{1-\rho}, 0\}$. Furthermore consider term III. It follows that

$$\frac{d\left(\delta^2(1-\rho)^2+2(1-\delta+\delta\rho)(1-\sqrt{1-\delta})-\delta)\right)}{d\delta} =$$
$$\frac{1}{\sqrt{1-\delta}}\left(2\rho\sqrt{1-\delta}-2\sqrt{1-\delta}+1\right)\left(1-\delta-\sqrt{1-\delta}+\delta\rho\right) \geq 0 \text{ for all } \rho,\delta$$

Since Terms I and III in $\gamma_{\delta}(\rho,\delta)$ are always positive, the comparative static result holds based on the sign of Term II as described above. Finally, it also follows that $\widetilde{\delta}^{\rho} = 0$ for $\rho \geq 0.5$. Hence for all $\rho > 0.5$, $\frac{\partial\gamma(\rho,\delta)}{\partial\delta} > 0$. Also for all $\rho > 0$, $\frac{\partial\gamma(\rho,\delta)}{\partial\delta} > 0$ if $\delta > \widetilde{\delta}^{\rho}$ and $\frac{\partial\gamma(\rho,\delta)}{\partial\delta} < 0$ if $\delta < \widetilde{\delta}^{\rho}$ since $\gamma_{\delta}(\rho,\delta) = 0$ is unique.

Note that the gradient of the price dynamic is always

$$\frac{p_{t+1}}{p_t} = \lambda(\rho,\Delta)\gamma(\rho,\Delta) = \frac{1}{e^{-\Delta}}\left(1-\sqrt{1-e^{-\Delta}}\right)$$

hence it is independent of ρ. Furthermore, it follows that the price-drop from one period to the next is *increasing* in Δ for all $\rho \geq 0$.

Finally, note that $\gamma(\rho,\Delta)$ converges to 1 as Δ goes to zero. Given that $\lim_{\Delta\to 0}\frac{p_{t+1}}{p_t} \to 1$, it follows that for any $\rho > 0$ and any $\tau > 0$, there exists $\widehat{\Delta}^{\rho}(\tau) > 0$ such that if $\Delta \leq \widehat{\Delta}^{\rho}(\tau)$, then $\left|V_S^{\rho}(\Delta) - V_F\right| \leq \tau$. To show this formally, note that the seller's ex-ante expected revenue can be approximated as

$$V_S^{\rho}(\Delta) = \sum_{t=1}^{\infty} e^{-\Delta(t-1)}(\lambda(\rho,\Delta)^{t-1}\gamma(\rho,\Delta)^t)(\lambda(\rho,\Delta)^{t-1}\gamma(\rho,\Delta)^{t-1} - \lambda(\rho,\Delta)^t\gamma(\rho,\Delta)^t) =$$

$$= \gamma(\rho,\Delta)(1-\frac{1}{e^{-\Delta}}(1-\sqrt{1-e^{-\Delta}})\sum_{t=1}^{\infty}(e^{-\Delta}\lambda(\rho,\Delta)^2\gamma(\rho,\Delta)^2)^{t-1}$$

It follows, that for any $\rho > 0$, $\lim_{\Delta \to 0} V_S^{\rho}(\Delta) = V_F$. Note that whether the sum converges to V_F or 0 depends solely on $\lim_{\Delta \to 0} \gamma(\rho, \Delta)$. As long as $\lim_{\Delta \to 0} \gamma(\rho, \Delta) = 1$ the revenue converges to V_F. Furthermore, simple algebra shows that since $\gamma(0, \Delta)$ is increasing in Δ, $V_S^0(\Delta)$ is increasing in Δ as well .

Proof of Proposition 6. First note that $\gamma(\rho, \Delta)$ can be re-written as

$$-\overbrace{\frac{(1-(1-\rho)\delta)^2}{\delta^2(1-\rho)^2 - \delta(1-\rho) + \rho\delta}}^{I} - \overbrace{\frac{(\sqrt{1-\delta})\,(\delta\rho - \delta + 1)}{\delta^2(1-\rho)^2 - \delta(1-\rho) + \rho\delta}}^{II} \tag{17}$$

When concentrating on the limit as $\delta \to 1$, we can ignore Term II in the above expression as it converges to zero independent of ρ. By substituting $\rho = \beta(-\ln\delta)^{\kappa}$ given $-\Delta = \ln\delta$ and setting $\beta = 1$, and re-arranging terms in Term I we get that

$$\lim_{\delta \to 1}(1 - \overbrace{\frac{\delta - 1}{\delta \ln^{\kappa}\frac{1}{\delta} + \delta^2\left(\ln^{\kappa}\frac{1}{\delta} - 1\right)^2 + \delta\left(\ln^{\kappa}\frac{1}{\delta} - 1\right)}}^{III})$$

By applying l'Hôpital's rule on Term III in this expression, after dividing both the numerator and the denominator by δ, we get that

$$\lim_{\delta \to 1} 1/\delta \left(\delta - 2\delta \ln^{\kappa}\frac{1}{\delta} + \delta \ln^{2\kappa}\frac{1}{\delta} + 2\kappa\left(\ln^{\kappa-1}\frac{1}{\delta}\right)(\delta - 1) - 2\kappa\delta \ln^{2\kappa-1}\frac{1}{\delta}\right) \tag{18}$$

It follows that terms in the denominator, except for the first and the last ones, converge to zero. The last term goes to 0 if $\kappa > 0.5$ and becomes unbounded if $\kappa < 0.5$. Hence in the former case we have that $\lim_{\Delta \to 0} \gamma(\rho(\Delta), \Delta) = 0$ and in the latter case we have that $\lim_{\Delta \to 0} \gamma(\rho(\Delta), \Delta) = 1$. In these cases the results hold for all $\beta > 0$. In the case where $\kappa = 0.5$, the revenue converges to $V_F\beta/(1+\beta)$.

References

[1] Admati, Anat and Motty Perry. (1987). "Strategic Delay in Bargaining." *Review of Economic Studies*, 54, 345–364.

[2] Ausubel, Lawrence and Raymond Deneckere. (1989). "Reputation in Bargaining and Durable Goods Monopoly." *Econometrica,* 57(3), 511–531.

[3] Biais, Bruno and Martin Weber. (2010). "Hindsight Bias, Risk Perception and Investment Performance." *Management Science*, 55(6), 1018-1029.

[4] Birch, Suan and Paul Bloom. (2007). "The Curse of Knowledge in Reasoning About False Beliefs." *Psychological Science,* 18(4), 382-386.

[5] Board, Simon and Marek Pycia. (2014). "Outside Options and the Failure of the Coase Conjecture." *American Economic Review*, 104(2): 656–671.

[6] Bond, Eric and Larry Samuelson. (1984). "Durable Good Monopolies with Rational Expectations and Replacement Sales." *Rand Journal of Economics*, 15(3), 336–345.

[7] Camerer, Colin, George Loewenstein and Martin Weber. (1989). "The Curse of Knowledge in Economic Settings: An Experimental Analysis." *Journal of Political Economy,* 97(5), 1234-1254.

[8] Coase, Ronald. (1972). "Durability and Monopoly." *Journal of Law and Economics,* 15, 143–149.

[9] Danz, David, Kristof Madarasz and Stephanie Wang. (2014). "Anticipating Information Projection: An Experimental Investigation." WZB discussion papers, Berlin.

[10] Eyster, Erik and Matthew Rabin. (2005). "Cursed Equilibrium." *Econometrica*, 73(5), 1623–1672.

[11] Feinberg, Yossi and Andrzej Skrzypacz. (2006). "Uncertainty about Uncertainty and Delay in Bargaining." *Econometrica*, 73(1), 69-91.

[12] Fuchs, William and Andrzej Skrzypacz. (2010). "Bargaining with Arrival of New Traders." *American Economic Review*, 100(3), pp. 802–36.

[13] Fudenberg, Drew, David Levine, and Jean Tirole. (1985). "Infinite-horizon Models of Bargaining with One-sided Incomplete Information." in A. E. Roth (Ed.), *Game-Theoretical Models of Bargaining,* Cambridge University Press, Cambridge, MA, 73-98.

[14] Gilovich, Thomas, Victoria Medvec and Kenneth Savitsky. (1998). "The illusion of transparency: Biased assessments of others' ability to read our emotional states." *Journal of Personality and Social Psychology,* 75, 332–346.

[15] Gilovich, Thomas, Victoria Medvec and Kenneth Savitsky. (2000). "The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance." *Journal of Personality and Social Psychology*, 78, 211–222.

[16] Gilovich, Thomas, Victoria Medvec and Leaf Van Boven. (2003). "The Illusion of Transparency in Negotioations." *Negotiation Journal*, 19, 117-131.

[17] Goldberg, Koujianou Pinelopi. (1996). "Dealer Price Discrimination in New Car Purchases: Evidence from the Consumer Expenditure Survey." *Journal of Political Economy*, 104(3), 622-654.

[18] Gul, Faruk, Hugo Sonnenschein and Robert Wilson. (1986). "Foundations of Dynamic Monopoly and the Coase Conjecture." *Journal of Economic Theory*, 39, 1.55-190.

[19] Hart, Oliver, and Jean Tirole. (1988). "Contract Renegotiation and Coasian Dynamics." *Review of Economic Studies,* 55(4), 509-540.

[20] Jehiel, Philippe. (2005). "Analogy-Based Expectations Equiluibrium." *Journal of Economic Theory,* 123, 81–104.

[21] Keysar, Boaz, Linda Ginzel and Max Bazerman. (1995). "States of Affairs and States of Mind: The Effect of Knowledge of Beliefs." *Organizational Behavior and Human Decision Processes*, 64, 283-293.

[22] Loewenstein, George, Don Moore, and Roberto Weber. (2006). "Misperceiving the Value of Information in Predicting the Performance of Others." *Experimental Economics*, 9(3), 281-295.

[23] Madarasz, Kristof. (2012). "Information Projection: Model and Applications." *Review of Economic Studies,* 79(3), 961-985.

[24] Madarasz, Kristof. (2014). "Information Projection Equilibrium: Definition and Applications to Social Investment and Persuasion." mimeo LSE.

[25] Myerson, Roger. (1981). "Optimal Auction Design.", *Mathematics of Operations Research,* 6, 58–73.

[26] Piaget, Jean and Barber Inhelder. (1948). The Child's Conception of Space. Translated (1956). London: Routledge and Kegan Paul.

[27] Pronin, Emily, Thomas Gilovich and Lee Ross. (2004). "Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others." *Psychological Review,* 111(3), 781–799.

[28] Raifa, Howard. The Art and Science of Negotiations, Harvard

[29] Rapoport, Amnon, Ido Erev and Rami Zwick. (1995). "An Experimental Study of Buyer-Seller Negotiation with One-Sided Incomplete Information and Time Discounting." *Management Science,* 41, 377–94.

[30] Riley, John and Richard Zeckhauser. (1983). "Optimal Selling Strategies: When to Haggle and When to Hold." *Quarterly Journal of Economics*, 98(2), 267-89.

[31] Samuelson, William and Max Bazerman (1985). "The Winner's Curse in Bilateral Negotiations." In Research in Experimental Economics, vol. 3, Vernon L. Smith, ed., Greenwich, CT: JAI Press.

[32] Scott-Morton, Fiona, Florian Zettelmeyer and Jorge Silva-Risso. (2001). "Internet Car Retailing." *Journal of Industrial Economics,* 49(4), 501–520.

[33] Sobel, Joel and Ichioro Takahashi. (1983). "A Multistage Model of Bargaining." *Review of Economic Studies,* 50(3), 411-426.

[34] Skreta, Vasiliki. (2006). "Sequentially Optimal Mechanisms." *Review of Economic Studies*, 73(4), 1085-1111.

[35] Stokey, Nancy. (1979), "Intertemporal Price Discrimination," *Quarterly Journal of Economics*, 93, 355-371.

[36] Yildiz, Muhamet. (2003). "Bargaining without a Common Prior: An Immediate Agreemeent Theorem." *Econometrica,* 71(3), 793–811.

[37] Vorauer, Jacquie and Stephanie-Danielle Claude. (1998). "Perceived versus Actual Transparency of Goals in Negotiation." *Personality and Social Psychology Bulletin,* 24, 371–385.

[38] Vorauer, Jacquie and Michael Ross. (1999). "Self-awareness and Feeling Transparent: Failing to Suppress one's Self." *Journal of Experimental Social Psychology*, 35, 415–440.

Anticipating Information Projection: An Experimental Investigation*

David Danz[†], Kristóf Madarász[‡], Stephanie W. Wang[§]

August 2014

Abstract

We investigate whether people predict the biases of others with an experiment that matches agents with informed or uninformed principals. While the agents' beliefs are well-calibrated regarding the average task performance, they also correctly believe that informed principals monitoring their performance project their information and overestimate the likelihood of success for agents. The agents' preference for a payoff that is independent of the principal's evaluation when matched with an informed principal is consistent with anticipating this information projection. We also use the agents' stated first- and second-order beliefs to estimate the extent of information projection and its anticipation.

Keywords: information projection, defensive agency, curse of knowledge, hindsight bias, second-order beliefs

JEL classification numbers: C91, D82, D83, D84

*We are grateful for valuable comments by Colin Camerer, Gary Charness, John Duffy, Ignacio Esponda, Dorothea Kübler, Andrew Schotter, Lise Vesterlund, and Alistair Wilson. Financial support from the Deutsche Forschungsgemeinschaft (DFG) through CRC 649 "Economic Risk" is gratefully acknowledged.

[†]WZB and Technische Universität Berlin.

[‡]London School of Economics.

[§]University of Pittsburgh.

1 Introduction

Informational differences that exist between people are key in many economic settings. The standard economic assumption when studying the consequences of asymmetric information is that people perceive informational differences correctly. However, a collection of very robust findings indicate that people systematically misperceive such differences. In particular, the evidence shows that the typical person *projects* her private information onto others in that she exaggerates the extent to which others can act on her private information.

Evidence on information projection comes from a variety of settings. One of the most well-established phenomenon in cognition is the hindsight bias, showing that people who have access to new information in the present greatly exaggerate the extent to which past actions of others must have reflected such information (e.g., Fischhoff, 1975; Bernstein et al., 2004)[1]. Research on the curse-of-knowledge, e.g., Camerer, Loewenstein, and Weber (1989), Newton (1990), Birch and Blum (2007), the illusion of transparency (Gilovich et al., 1998, 2000), and outcome bias (Baron and Hershey, 1988) all show that people fail to appreciate the extent to which others do not have access to their private information.[2]

Consider the following simple setting that describes some important consequences of this information projection phenomenon. An agent works on a task where her performance is determined jointly by her skill and the information available to her. Later her performance is evaluated by a principal who is better informed. The principal might have more experience in the task which involves access to more background information, or might have access to ex-post information that was not available ex-ante to the agent. By projecting information, the principal will exaggerate how much of her extra information should have been available to the agent. As a result, the principal will over-infer from bad performance and under-infer from good performance and will come to systematically underestimate the skill of the agent on average (Madarasz 2012).

[1] Although hindsight bias is typically described as an intrapersonal phenomenon, the evidence is predominantly from interpersonal settings.

[2] This phenomenon can likely explain a host of other biases that derive from a general hypothesis of naive realism. See, for example, Nisbett and Ross (1980), whereby people engage in limited informational perspective-taking when interacting with others. Evidence also indicates that information projection is usually robust to de-biasing attempts. In the context of hindsight bias, see Wu et al. (2012).

The radiologist Leonard Berlin, in his 2003 testimony on the regulation of mammography to the U.S. Senate Committee on Health, Education, Labor, and Pensions highlighted this issue. [3] Acknowledging that given heterogeneity in skills it is important to assess dthe quality of a radiologist, he argued that making ex-post assessments of mammogram interpretations publicly available nevertheless leads to more malpractice suits and worse medical practices. Berlin describes how ex-post information causes the public to *misperceive* the ex-ante accuracy of mammograms causing juries to be "all too ready to grant great compensation." He explicitly referenced the role of hindsight bias in such ex-post assessments, which is excarbated by the fact that ex-post information makes reading old radiographs much easier: "suffice it to say that research studies performed at some of the most prestigious medical institutions of the United States reveal that as many as 90% of lung cancers and 70% of breast cancers can at least be partially observed on studies previously read as normal." In response, physicians are reluctant to follows such crucial diagnostic practices: "the end result is that more and more radiologist are refusing to perform mammography and fewer and fewer radiology residents completing their formal training are opting to take additional fellowship training in mammography. In turn mammography facilities are closing."

One of the physician letters included in the testimony explicitly called into question the evaluator's assessment of the radiologist's performance, "we all agree the screening mammogram was negative eight months prior to discovery to the cancer, except of course the plaintiff's so-called expert-witness. ... Even perfect professional performance provides no protection in Florida!" His testimony provided strong anecdotal evidence on the specter of the anticipated misperception of negligence. [4]

In another vein, a growing literature in industrial organization now focuses not only on incorporating behavioral assumptions in understanding consumer behavior, but also on managerial decisions in firms (e.g., Malmendier and Tate, 2008; Armstrong and Huck, 2010; Spiegler, 2011).[5] This approach is crucial for understanding the nature and efficiency of firms and large organizations, including how they allocate resources and their relation to

[3]www.fda.gov/ohrms/dockets/ac/03/briefing/3945b1_05_Berlin%20testimony.pdf

[4]Similar arguments were made by many others, e.g, Jackson and Righi in their book, "The Death of Mammography," argue that tort liability and ex-post assessments play a key role in doctor's reluctance to engage in otherwise very valuable diagnostic practices.

[5]For a survey, see Camerer and Malmendier (2007).

internal and external labor markets. Behavioral facts might crucially affect the principal-agent relationship and the use of information and incentives in such contexts. In order for many implications of information projection to matter in organizational or IO settings, it is not enough for the principal to be biased. The agent must anticipate this bias.

Clearly, a key question for understanding the economic implications of information projection is whether people anticipate the fact that others project information onto them. For example, an agent who understands that her principal projects information may try to engage in various practices aimed at minimizing the discrepancy between the value of the information available to her ex-ante and that valuable to the principal ex-post. Authors in the legal (e.g., Rachlinski 1998, Harley 2007) and in the medical (e.g., Studdert at al 2005) literature often discuss behavior that is attributed to people or institutions best-responding to the presence of information projection by others (e.g., bifurcation of trials that is establishing liability before hearing evidence on damages, or defensive medicine), but they provide no direct evidence. This paper is a first attempt to empirically investigate whether people - despite being subject to it themselves - anticipate the information projection of others and adjust their actions to its presence.

In our experiments, *principals* estimated the performance of *agents* in change-detection tasks in the *reference treatment*. The principals received the solution to each task in the *informed* treatment but not in the *uninformed* (control) treatment. *Agents* were matched to informed or uninformed principals, respectively, fully aware of the information available to the principal. For each task, the agents could choose between a lottery whose return positively dependent on how difficult the principal estimated the task to be and a lottery that was independent of the principal's beliefs (insurance task). Consistent with earlier findings, we find strong evidence of information projection from the principals in our experiments. Better informed principals greatly exaggerate the probability that lesser informed others should be able to act as if they had access to their superior information. That is, on average informed principals significantly overestimate the success rate of agents, while uninformed principals are well-calibrated.

Crucially, we find clear evidence that agents act as if they anticipated the information projection of principals. This evidence comes in two forms. First, agents matched with informed principals more often preferred the choice whose payoff was independent of the principal's evaluation than agents matched with uninformed principals. Second, the same anticipation is found in the incentivized elicitation of the agents' second-order beliefs. We found that while agents' beliefs are well-calibrated regarding the average performance in the task, they also understand that by having access to additional information, principals will overestimate the likelihood of success in the task. The difference between agents' second-order beliefs (their estimate of the principals' estimates) and their first-order beliefs (their estimate of the success rate of agents in the reference treatment) is significantly larger when matched with informed principals than with uninformed principals.

The fact that people anticipate the information projection of others also matters for strategic contexts more generally. In many settings such as bargaining, communication, bilateral trade, social investment, or social learning, key predictions depend on whether people are aware of the tendency of others to project information and best respond to it. For example, Madarász (2014) shows that if sellers anticipate the information projection of buyers they might use very different bargaining tactics to maximize their revenue and boost it to a level that exceeds that of the optimal mechanism under rational expectations (Myerson, 1981). The experimental design we use allows us to test an important aspect of the resulting solution concept, information projection equilibrium, whereby a person simultaneously project information onto others and to a lesser extent, where this lesser extent is governed by her own information projection, she anticipates that others will project information onto them. Consistent with the notion of information projection equilibrium, a single-parameter extension of Bayesian Nash equilibrium, we find that the exaggeration anticipated (by the agents matched with informed principals) is positive but is less than the exaggeration exhibited (by the informed principals). *Estimating this model given our...*

The paper is structured as follows. In Section 2, we present the experimental design and procedures. Section 3 enumerates the main predictions. Section 4 contains the results. In Section 5 we conclude with a discussion of some of the implications of anticipated information projection for agency and strategic settings.

2 Experimental Design and Procedures

Participants in all treatments worked on the same series of 20 change-detection tasks. In each of these tasks, the subjects had to spot the difference between two nearly identical images (see Rensink et al., 1997; Simons and Levin, 1997). Each task was presented in a 14-second video clip where the two images were displayed alternately with short interruptions.[6] Afterwards, subjects had 40 seconds to submit an answer. To this end, the image containing the object of interest was displayed together with a grid of 70 fields, where the subjects could enter one of the grid numbers as the answer. The answer was evaluated as correct as long as any part of the difference was contained in that field.

The experiment consisted of five treatments involving *agents (reference treatment)*, *principals* (*informed* or *uninformed*), and *agents* (*matched to informed principals* or *matched to uninformed principals*). All treatments followed a between-subject design.

2.1 Principals

The *agents* from the *reference treatment* performed the 20 change-detection tasks in tournaments (data taken from Danz, 2013). The *principals* worked on the same tasks and were informed that agents had performed the same tasks in previous sessions. After performing each task, the principals stated their estimate (b_t^P) of the success rate of the reference agents for the current task (π_t).

The two treatments involving principals differed as follows. *Informed principals* received the solution to each task *before* they examined the change-detection task. In the sessions with *uninformed principals*, the procedure was exactly the same except the principals were not given solutions to the tasks. Principals in both treatments observed each task exactly as the agents did. The principals did not receive feedback of any kind during the experiment.

At the end of the experiments, the principals were paid €0.50 for each correct answer in the uninformed treatment and €0.30 in the informed treatment.[7] In addition, they were paid based on the accuracy of their stated beliefs in two of the 20 tasks (randomly chosen):

[6]Each image was displayed for one second followed by a blank screen of 150 milliseconds.

[7]The principals first participated in three practice rounds to become familiar with the interface.

for each of these two tasks they were paid €12 if $b_t^P \in [\pi_t - 0.05, \pi_t + 0.05]$, that is, if the estimate was within 5 percentage points of the true success rate of the agents. We ran one session with informed principals and one with uninformed principals with 24 participants in each.

2.2 Agents

The *agents* also performed the same series of change-detection tasks as the agents in the reference treatment and the principals.[8] The agents were informed that the principals had estimated the performance of the agents in the reference treatment (being paid according to the accuracy of their estimates). The agents were further informed that they had been randomly matched to one of the principals at the outset of the experiment and that this matching would remain the same for the duration of the experiment.

The two treatments involving agents differed with respect to the kind of principal they were matched to. *Agents matched to informed principals* were told that the principals had been informed of the solution to each task prior to watching the task (that is, b_t^P comes from the treatment with *informed principals*). *Agents matched to uninformed principals* were told that the principals had observed each task just as they had done (that is, b_t^P comes from the treatment with *uninformed principals*).

After each change-detection task, the agents made a decision between a sure payoff of €4 and a payoff dependent on the principal's estimate. Specifically, s/he received €10 if the success rate on that task π_t—plus 10 percentage points—was at least as high as the principal's estimate b_t^P. If b_t^P was more than 10 percentage points higher than the success rate π_t, then the payoff was €0. [9]

In neither of the treatments did the agents receive information about the principal's estimates. Agents matched to informed principals were told that this feedback corresponded to what the principal had seen for that task. Agents matched to uninformed principals

[8]The order of the tasks was the same as for the principals.

[9]The agents participated in three practice rounds to familiarize themselves with the interface and the insurance decision.

were told that the principals had not received the solution to the task. In both information conditions, the agents did not receive information about the principal's estimates.

The agents were paid €0.50 for each correct answer to the change-detection tasks and according to one randomly selected insurance decision. We ran two sessions each of agents matched to informed principals (24 participants) and agents matched to uninformed principals (23 participants).

To explore the agents' beliefs that underlie their insurance decisions we ran additional sessions, one with agents matched to informed principals (24 participants) and one with agents matched to uninformed principals (23 participants). The sessions differed from the sessions without belief elicitation in that the insurance tasks in the first half of the experiment were replaced by belief tasks.[10] Specifically, after each of the first 10 change-detection tasks, the agents stated their belief about (i) the percentage who detected the difference in that task (first-order belief) and (ii) the estimate of their randomly matched principal's estimate of that success rate (second-order belief).

At the end of the experiment one round was randomly selected for payment. If this round involved belief tasks, then one of the agent's stated beliefs was randomly selected for payment, i.e., either their first-order belief or their second-order belief in that round. The subject received €12 if her stated belief was within five percentage points of the actual value (the actual success rate in case of a first-order belief and the randomly matched principal's estimate of that success rate in case of a second-order belief). If the round selected for payment involved an insurance decision, then the agent was paid according to her decision.

2.3 Procedures

The experimental sessions were run at the Technische Universität Berlin in 2014. Subjects were recruited with ORSEE (Greiner, 2004). The experiment was programmed and conducted with z-Tree (Fischbacher, 2007). The average duration of the principals' sessions was 67 minutes. The average earning was €15.15. The agents' sessions lasted 1 hour and 45

[10]In the sessions with belief elicitation, the agents participated in three practice rounds to familiarize themselves with the interface, the belief elicitation procedure, and the insurance decision.

minutes on average. The average payoff was €20.28.[11] Participants received printed instructions which were also read out loud and had to answer a series of comprehension questions before they were allowed to begin the experiment.[12] At the end of the experiment, but before receiving any feedback, the participants completed the four-question DOSE risk attitude assessment (Wang et al., 2010), a demographics questionnaire, the abbreviated Big-Five inventory (Rammstedt and John, 2007), and personality survey questions on perspective-taking (Davis, 1983). Table 1 summarizes the sessions.

Table 1: Overview of the sessions.

Session IDs	Treatment	Number of subjects
	Agent, reference	144
1	Principal, uninformed	24
2	Principal, informed	24
$3, 5, 8^*$	Agent, uninformed principal	$46\ (11 + 12 + 23)$
$4, 6, 7^*$	Agent, informed principal	$48\ (12 + 12 + 24)$

Note: Stars indicate sessions with belief elicitation of the agents. The data for the agents in the reference treatment was taken from Danz (2013).

3 Information Projection

The theoretical framework for our paper is based on Madarász (2014) which introduces information projection equilibrium into Bayesian games. We state the formal definition of an information projection equilibrium for our setting in the *Appendix* but present its unique predictions below.

We first need to formally express informational differences in our setting. Let there be two signals $s_{1,t}$ and $s_{2,t}$ for each task t. Signal $s_{1,t}$ represents the information revealed by watching the change-detection task. This is available to all participants. Signal $s_{2,t}$ represents the solution to the task. This signal is available to the informed principals only.

Information Projection corresponds to a player's exaggerated perception that her opponent acts as if he knew her private information and *also* what information she had. To

[11]The average duration of the sessions (the average payoff) in the treatments with and without belief elicitation was 115 and 96 minutes (€21.47 and €19.10), respectively.

[12]Two participants did not complete the comprehension questions and were excluded from the experiment.

introduce such beliefs into the analysis of strategic games, we need to consider besides the *regular* (real) version of a player a *super*(fictional) version of each player. In reality all players are regular, but fictional super versions enter into players' beliefs about each other. If player i projects information to degree $\rho \in [0, 1]$ she acts as if she believed that with probability ρ her opponent, player $-i$, is a *super version* of himself who (i) has access to the content player i's private information and (ii) also knows what information player i has. In other words, a person who projects information acts as if her information has leaked to her opponent. The regular player believes that her opponent is regular with probability $1 - \rho$ and super with probability ρ, The super version of a player believes that her opponent is regular for sure.

Now let's consider the two-player game studied in our paper. We denote the strategy of the regular version of player i by σ_i^ρ and that of the super version of player i by σ_i^+. The former is conditioned on player i's true information I_i. The latter is conditioned on the joint information of the players, $I_i \cup I_{-i}$.

A ρ information projection equilibrium (IPE) in our setting is given by a pair of strategies $\sigma^\rho = (\sigma_1^\rho, \sigma_2^\rho)$ describing the players' actual behavior and is supported by a pair of strategies $\sigma^+ = (\sigma_1^+, \sigma_2^+)$ describing the fictional super players' strategies.

Definition 1 *A strategy profile σ^ρ is a ρ information projection equilibrium if for all i,*

1. *σ_i^ρ is a best response to the belief that player $-i$ plays σ_{-i}^ρ with probability $(1 - \rho)$ and σ_{-i}^+ with probability ρ.*

2. *σ_{-i}^+ is a best-response to the belief that player i plays strategy σ_i^ρ for sure.*

The above concept is a single-parameter extension of a Bayesian Nash equilibrium (BNE). If $\rho = 0$ the set of information projection equilibria and the set of BNE coincide. If $\rho = 1$ each player puts full weight on the wrong belief that her opponent best-responds to her strategy using their joint information. It is straightforward to see that given our experimental design, the information projection equilibrium is always unique in our setting.

Predictions. The first consequence of this model for our experimental results (Claim 1) concerns the implications for the principal's predictions. In the game where the principal does not have private information, the predictions of an information projection equilibrium and a Bayesian equilibrium are the same. Both the principal and the agent receive the same signal $s_{1,t}$ so a regular strategy and a super strategy is conditioned on the same information ($I_i \cup I_{-i} = I_i$) and information projection has no bite. Hence here the principal will be well-calibrated about the agents' average success rate π_t.

In contrast, in the game where the principal does have private information, the information projection equilibrium predicts that he will exaggerate π_t. This is true because the principal who has access to both signals now acts as if she exaggerated the probability that the agent also conditions his choice on both signals. Furthermore, the theory provides a prediction about the magnitude of such exaggeration which is going to be proportional to the value of $s_{2,t}$ relative to $s_{1,t}$. To express this exaggeration, let the additional value of signal s_2, given s_1, measured in the probability of success be

$$d_t = \Pr(\text{success}_t \mid s_{1,t} \cup s_{2,t}) - \Pr(\text{success}_t \mid s_{1,t})$$

The informed principal thus exaggerates the success rate by ρd_t.

The second consequence of the model (Claim 2) concerns the implications for the agent's prediction and choice. In the game where the principal does not have private information, again information projection has no direct bite since super and regular variants of a player choose from the same set of strategies. In contrast, in the game where principal's have private information, the model implies that the agent will partially predict the principal's exaggeration. Specifically, the agent's first- and second-order beliefs about the average success rate will differ. The agent's first-order beliefs are predicted to be correct. The agent's second-order will anticipate the principal's exaggeration of the success rate but will underestimate the magnitude of this exaggeration.

To understand these predictions, consider the agent who in equilibrium projects information to degree ρ. As a consequence, she attaches probability $(1-\rho)$ to the principal being a regular version and playing the strategy the principal actually plays. This means that the agent attaches probability $(1-\rho)$ to the informed principal exaggerating the probability of the agents' success rate, i.e., the principal thinking as if the agent had access to the principal's superior information with probability ρ. At the same time, the agent attaches probability

ρ to the event that - despite being privately informed - the principal fully recognizes that the agent cannot act on signal $s_{2,t}$ and hence makes unbiased predictions about the agent's strategy. In short, due to the agent projecting herself and believing that with probability ρ the principal should recognize the agent's true information and true strategy, the agent predicts that the principal exaggerates the true success rate by $(1-\rho)\rho d_t$.

The above observations are reflected in the predictions described below that characterize the unique predictions of IPE in our setting.

Claim 1 (Principal's estimate) *In the game with an informed principal, the principal exaggerates π_t by*

$$\rho d_t \tag{1}$$

Claim 2 (Agent's estimate) *In the game with an informed principal, the difference between the agent's own estimate (first-order belief) and the agent's estimate of the principal's estimate (second-order belief) is given by*

$$(1-\rho)\rho d_t \tag{2}$$

Claim 3 (Agent's insurance choice) *The agent's propensity to choose insurance is higher when matched with an informed as opposed to with an uninformed principal.*

When $\rho = 0$, we get back the BNE predictions where the law of iterated expectations holds and all predictions of the agents and principals are correct. The principal has correct first-order belief about the success rate, and the agent has correct second-order belief about the principals first-order belief. In contrast, for all $\rho > 0$, the above differences are positive. The (informed) principal's exaggeration is increasing in ρ. The difference between the agent's (matched with the informed principal) second-order and first-order prediction is also positive, but anticipation of this exaggeration is always less than full, and is non-monotonic in the parameter ρ. If the agent is fully biased, then she expects the principal to be correct and hence anticipates no exaggeration by the principal. Limited anticipation is true, however, as long as the agent is partially biased. [13]

[13]It is straightforward to a consider the information projection equilibrium in our setting where the two-players are *differentially* biased, i.e., they exhibit differing (role-specific) degrees of information projection.

Finally, due to this anticipation effect, an agent matched with an informed principal has a higher propensity to opt for the safe payment than an agent matched with an uninformed principal, holding all other factors constant. The value the agent attaches to the lottery that depends on the principal's assessment is lower if the principal has private information about the task than if the principal does not. Furthermore, this difference is governed by the same forces as the difference between first- and second-order beliefs.

4 Results

4.1 Performance

Before analyzing the principals' beliefs or the agents' decisions, we first look at the participants' performance in the change-detection tasks in Table 2.

Table 2: Performance between treatments.

Treatment	# subjects	Success rate
Agent, reference	144	39.25
Principal, uninformed	24	31.67
Principal, informed	24	98.75
Agent, uninformed principal	46	39.89
Agent, informed principal	48	41.35

There is no significant difference in the performance between agents who were matched to informed and uninformed principals, respectively ($p = 0.57$).[14] Thus, any treatment differences in the agents' insurance decisions or beliefs cannot be attributed to differences in task performance. As expected, informed principals have significantly higher success rates than the participants in all other treatments ($p < 0.01$ for all comparisons). The performance of uninformed principals is significantly lower than the performance in the other treatments ($p < 0.01$ for all comparisons).

Here an increase in the agent's bias would simply correspond to a decrease in their estimate of how much informed principals exaggerate the probability of success.

[14] All results in this paragraph are two-sample t-tests based on the average success rates per subject over all rounds. All statistical tests are two-sided unless otherwise stated.

Next, we compare the principals' beliefs about the agents' task performance when the principals are informed about the solution versus when they are not. Recall that the primary focus of our study is on the agents' behavior, and whether or not agents anticipate information projection from the principal matters regardless of what principals actually do. Nevertheless, we ask whether the principals in our experiments exhibit information projection as has been found in previous studies.

4.2 Principals

Figure 1 depicts the performance estimates of the principals in the informed and the uninformed treatment together with the actual success rate. The average estimate of uninformed principals (39.76%) is not significantly different from the actual performance of the agents in the reference treatment (39.25%; $p = 0.82$)[15]. Informed principals significantly overestimate the performance of the agents on average (57.45%) and submit estimates that are significantly higher than those by uninformed principals ($p < 0.01$ in either case)[16]. A Kolmogorov–Smirnov test of the CDFs of average individual estimates between treatments yields $p = 0.001$[17]. This overestimation is consistent with the principals projecting their information (the answer to the change-detection task) onto the agents and underestimating the difficulty of the tasks for the agents. The CDFs also show substantial heterogeneity in beliefs in both treatments. If the informed principals simply had no idea about the success rate, we might expect to see a substantial cluster around 50% as a random guess, but this is not the case. In the uninformed treatment, essentially no principal believes that the agents' had more than 50% success rate on average, while a significant number of principals believed in an above chance success rate in the informed treatment.

Not surprisingly, informed principals (but not the uninformed principals) overestimating the actual performance by more than 17 percentage points led to significantly lower expected earnings in the informed treatment than in the uninformed treatment. The performance

[15] *t*-test of the average estimates per principal against the average success rate (over all tasks).

[16] Relatedly, principals in the uninformed treatment who spotted the difference in the task also overestimated the success of the reference agents (60.93%). Principals who have the solution think the task was easier for the agents than it actually was, regardless of whether the solution was given right away as in the informed treatment or found by the principals themselves.

[17] see Figure 6 in the Appendix for the CDF of the average estimate per principal in each treatment.

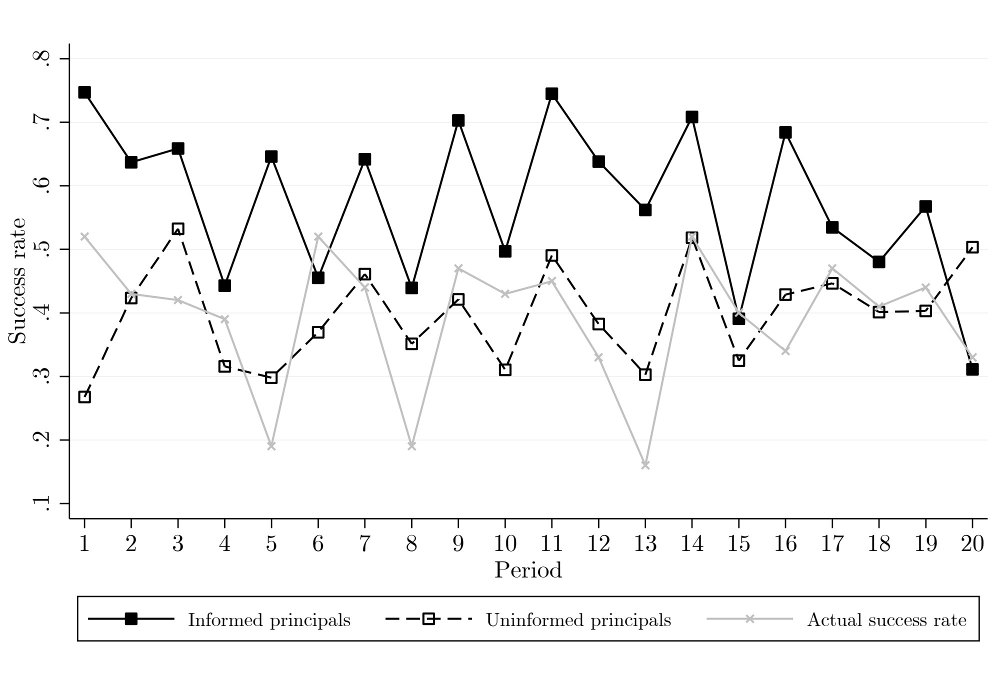


Figure 1: Average performance estimates of principals and actual success rate of the agents (reference treatment) over time.

estimates of the principals were within the 5 percentage point interval around the actual success rate in 10% of the cases in the informed treatment and in 15.2% of the cases in the uninformed treatment[18] yielding an expected payoff of €2.40 and €3.65 in the informed and the uninformed treatment, respectively (one-sided t-test: $p = 0.034$).[19]

4.3 Agents

For the agents, we first focus on the decision between the payoff that is independent of the principal's estimate, hence providing insurance, and the payoff that depends on the principal's estimate.

[18]The mean squared error of the principals' estimates in the informed treatment (0.110) is significantly higher than in the uninformed treatment (0.065; t-test: $p = 0.009$).

[19]The reported expected payoffs (and the test of the treatment difference) are based on random draws with, rather than without, replacement. The actual average payments for the belief tasks were €2.50 and €1.50 in the informed and the uninformed treatment, respectively (not significantly different).

4.3.1 Insurance decisions

Figure 2 shows the average insurance rate of the agents in both information treatments over time.[20]

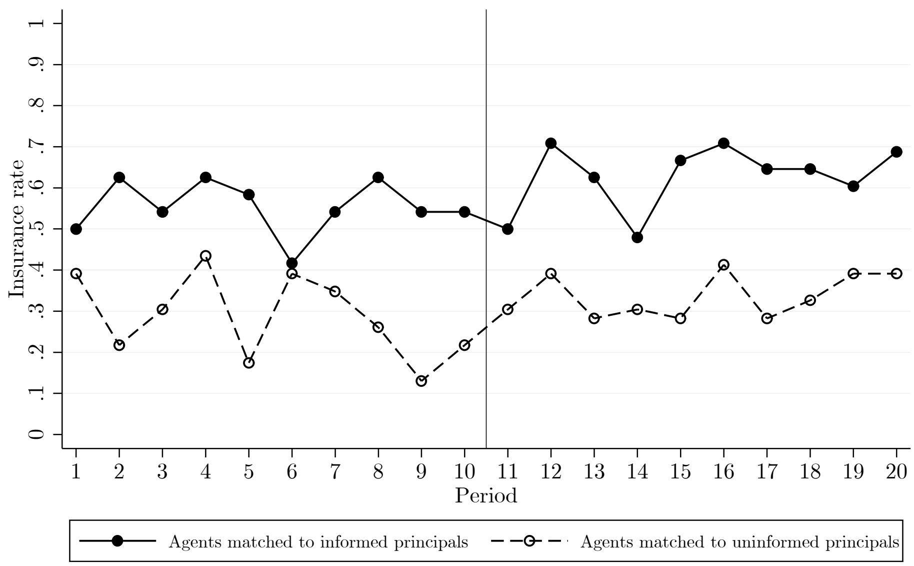


Figure 2: Insurance rates decisions over time.

Agents matched to uninformed principals insure for 32.7% of the tasks while agents matched to informed principals insure for 60.8% of the tasks. The average insurance rate per subject is significantly higher in the informed treatment than in the uninformed treatment (t-test: $p = 0.0003$). Figure 8 in the Appendix is a CDF of the insurance rate by agent. There is considerable heterogeneity across individuals in both information treatments, and the majority of agents do not choose insurance (or no insurance) for all rounds.

Table 3 reports the results of regressions of the insurance rate per subject on the treatment dummy, gender, and individual risk attitudes as measured by DOSE. There are three observations. First, the treatment effect is significant when controlling for gender and individual risk attitude. Second, female participants as well as subjects with higher degrees

[20] We pool the data of sessions with belief elicitation and those without belief elicitation. Within the informed [uninformed] treatments, the average insurance rates per agent in sessions with belief elicitation do not differ from the average insurance rates in sessions without belief elicitation (t-test, $p = 0.76$ [$p = 0.70$], see figure 7 in the Appendix for an overview of the average insurance rate for each session over time). There are no significant time trends in the insurance decisions (at the 5% level, see Table 5 in the Appendix). We thus focus on the average insurance rate per subject over all rounds in the subsequent analysis.

Table 3: Insurance rates by treatment and further controls.

Dependent variable (OLS)	Individual insurance rate (1)	(2)	(3)	(4)	(5)
Treatment (1-informed)	0.281*** (0.075)	0.279*** (0.075)	0.255** (0.102)	0.259*** (0.077)	0.254** (0.102)
Gender (1-female)		0.059 (0.075)	0.032 (0.108)		0.048 (0.109)
Treatment×Gender			0.053 (0.151)		0.009 (0.157)
Coef. risk aversion (DOSE)				0.026 (0.022)	0.024 (0.023)
Constant	0.327*** (0.053)	0.302*** (0.063)	0.313*** (0.071)	0.305*** (0.057)	0.285*** (0.076)
N	94	94	94	94	94
R^2	0.134	0.140	0.141	0.147	0.151
F	14.230	7.390	4.920	7.813	3.960

Note: Values in parentheses represent standard errors. Stars represent p-values: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

of risk aversion tend to insure more often (columns 2 and 4, respectively). However, these effects are not significant. Third, the treatment effect is significant for both male and female participants and there is no significant gender difference in the treatment effect (column 3).[21]

4.3.2 Stated beliefs

Next, we look at the beliefs that underlie the agents' strategic decision-making. To distinguish between possible differences in the agents' own beliefs about other agents' task performance across treatments and differences in beliefs induced by anticipating information projection from the principals, we elicited both first-order and second-order beliefs. The first-order beliefs of the agents, i.e., their estimates of the performance in the reference treatment (39.25%), are not significantly different between information conditions (t-test: $p = 0.96$) and are correct on average. The average estimate of the success rate is 39.72% ($p = 0.86$) for agents matched to uninformed principals and 39.91% ($p = 0.77$) for agents matched to

[21] There is no significant interaction between the treatment and having completed a task successfully. The treatment effect on insurance decisions is significant both for periods where the agents solved the task and for periods where the agents did not solve the task. For further details see Table 6 in the Appendix.

informed principals. Figure 3 shows the average second-order beliefs of the agents together with their first-order beliefs in both information conditions over time.

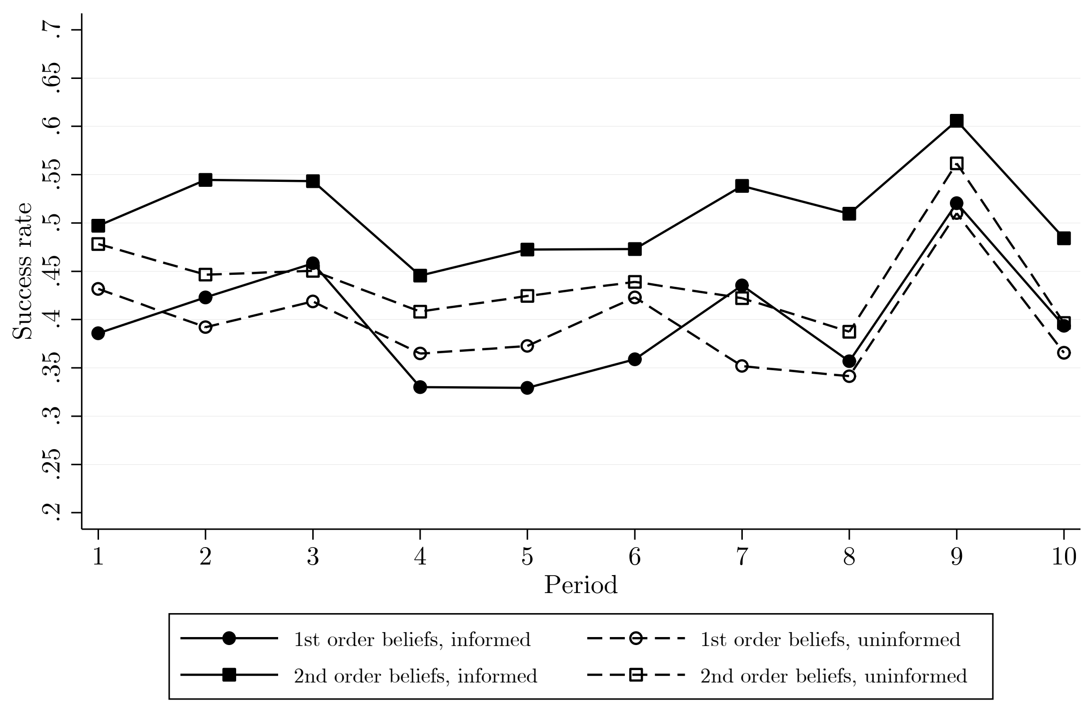


Figure 3: Agents' 1st-order beliefs (estimates of the success rates of the subjects in the reference treatment) and 2nd-order beliefs (estimates of the principals' estimate) over time conditional on being matched with informed or uninformed principals.

The second-order beliefs of agents (i.e., their estimates of the principals' estimates) who were matched to informed principals (51.14%) are significantly higher than those of agents matched to uninformed principals (44.15%; one-sided t-test: $p = 0.031$).[22] Accordingly, mean individual differences between second- and first-order beliefs are significantly higher for agents matched to informed principals than for agents matched to uninformed principals ($p < 0.001$)[23]. While the second-order belief for agents in the informed treatment is significantly higher than the actual success rate, it is also significantly lower ($p < 0.001$) than the informed principals' estimates (57.45%). That is, the agents appear to anticipate the infor-

[22]The t-test is applied to the average second-order belief per agent (over all periods) between treatments.

[23]A Kolmogorov–Smirnov test between treatments of the agents' average difference between their second-order beliefs and their first-order beliefs yields $p < 0.001$. A (two-sided) Kolmogorov–Smirnov test of the agents' average second-order beliefs between treatments yields $p = 0.064$.

mation projection of the informed principals, but not to its full extent, which is consistent with the predictions of the Information Projection Equilibrium.

Table 4 reports the results of regressions of average individual differences $(b^A_{2,i} - b^A_{1,i}) = T^{-1}\sum_t(b^A_{2,i,t} - b^A_{1,i,t})$ between agents' second-order beliefs (estimate of the principal's estimate) and their first-order beliefs (estimate of the success rate) on the treatment dummy, gender, and individual risk attitudes as measured by DOSE. Again, the treatment effect is significant when controlling for gender and individual risk attitude. In the uninformed condition, the average individual differences between second- and first-order beliefs tend to be larger for female participants than for male participants, and the treatment effect tends to be smaller for female than for male participants (columns 3 and 5). However, these differences are not significant at the 10% level. Higher degrees of individual risk aversion are slightly related to smaller differences in second- and first-order beliefs, but this relation is not significant either (columns 4 and 5).[24]

4.4 Information projection

We have presented evidence consistent with information projection and the predictions presented in section 3. In particular, consistent with information projection equilibrium, informed principals, but not uninformed ones, exaggerated the success rate of agents (Claim 1). In addition, agents' insurance choices were consistent with the fact that they anticipated such an exaggeration (Claim 3). Now we can also use the principals' beliefs and the agents' second-order beliefs to quantitatively estimate the extent of the information projection and the anticipation of such information projection (ρ in Claims 1 and 2).

Principals Let us first turn to the principals' estimation of the success rates. To have comparable estimates across treatments, we estimate an over-estimation term for each principal in each treatment by OLS. This will be denoted by α_i. The estimation for each principal is based on the 20 experimental tasks, where we substitute π_t by the actual success rate of the

[24]There is no significant interaction between the treatment and having completed a task successfully. The treatment effect on the individual differences between first-order and second-order beliefs is significant both for periods where the agents solved the task and for periods where the agents did not solve the task. For further details see Table 7 in the Appendix.

Table 4: Mean individual differences in second-order beliefs (estimate of the principal's estimate) and first-order beliefs $b^A_{1,i}$ (estimate of success rate) by treatment and further controls.

Dependent variable			$(b^A_{2,i} - b^A_{1,i})$		
(OLS)	(1)	(2)	(3)	(4)	(5)
Treatment	0.068***	0.067***	0.089***	0.073***	0.090***
(1-informed)	(0.019)	(0.019)	(0.024)	(0.020)	(0.024)
Gender		0.013	0.047		0.045
(1-female)		(0.020)	(0.029)		(0.030)
Treatment×Gender			−0.062		−0.056
			(0.040)		(0.041)
Coef. risk aversion				−0.006	−0.004
(DOSE)				(0.006)	(0.006)
Constant	0.044***	0.040**	0.030*	0.048***	0.034*
	(0.014)	(0.015)	(0.016)	(0.014)	(0.017)
N	47	47	47	47	47
R^2	0.220	0.228	0.270	0.236	0.278
F	12.720	6.490	5.289	6.787	4.035

Note: Values in parentheses represent standard errors. Stars represent p-values: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

agents per task, and $\widetilde{\Pr}(S_t|I_j)$ by the principals' stated beliefs in each period t.[25] That is, for each principal we estimate

$$b^P_{1,t,i} = (1 - \alpha_i)\pi_t + \alpha_i + \epsilon_{it},$$

where $b^P_{1,t,i}$ denotes the first-order belief of principal i and π_t the success rate of the agents in the reference treatment in task t.

We now obtain an average estimated over-estimation term in the uninformed treatment by averaging the coefficient estimates of principals in this treatment, let this be denoted by $\overline{\widehat{\alpha}}_{un}$. We also obtain an analogous average estimated over-estimation term in the informed treatment, let this be denoted by $\overline{\widehat{\alpha}}_{in}$.

[25]Note that the treatment difference does not need special indexation since the index i denotes a specific principal and we use a between-subject design.

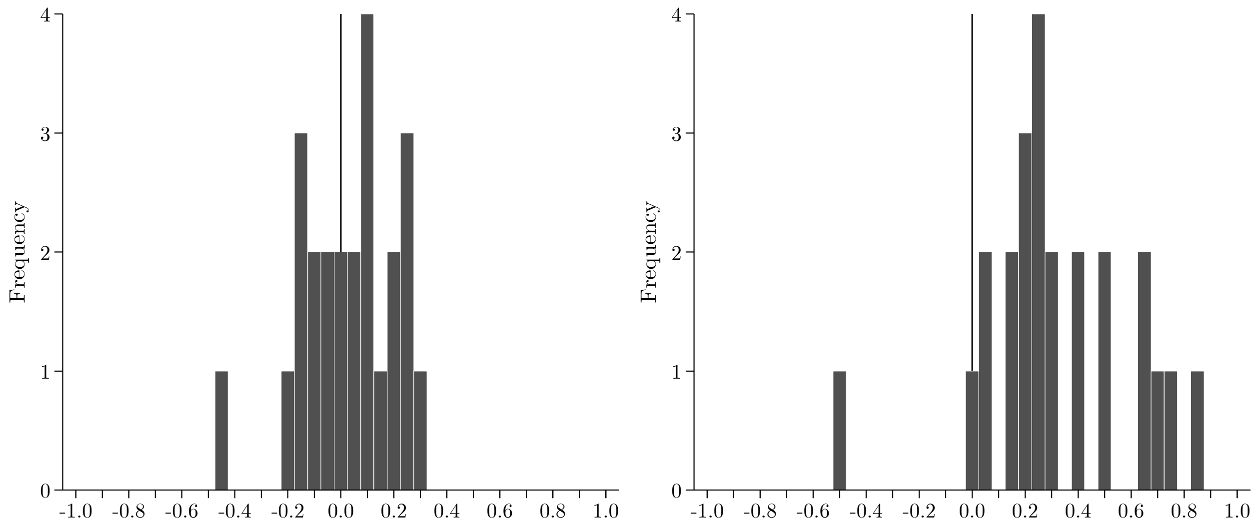


Figure 4: Histogram of individual estimates of the parameter of overestimation in the uninformed treatment (left panel) and the informed treatment (right panel).

Figure 4 provides an overview of the individual estimates in both treatments. We first note that the estimated parameters of over-estimation of informed principals are significantly higher than those of uninformed principals on average ($p < 0.001$). The average estimated parameter of over-estimation is significantly greater than zero in the informed treatment ($\overline{\widehat{\alpha}}_{in} = 0.31$; t-test: $p < 0.0001$) and is not significantly different from zero in the uninformed treatment ($\overline{\widehat{\alpha}}_{un} = 0.03$; $p = 0.42$). Likewise, the proportion of (strictly) positive estimates of α_i is significantly larger than expected by chance (0.5) in the informed treatment ($95.8\%, p < 0.0001$), but is not significantly different from 0.5 in the uninformed treatment ($54.16\%, p = 0.68$). To summarize, on average informed principals behave as if believing that the solution to the tasks was available to the agents with roughly 30% probability, while the corresponding probability for uninformed principals is statistically indistinguishable from zero.

Agents Do agents anticipate the information projection of the informed principals? To answer this question we perform the same exercise as in the previous subsection, but replace the principals' first order beliefs by the agents' second order beliefs. That is, for each agent we estimate:

$$b^A_{2,t,j} = (1 - \beta_j)\pi_t + \beta_j + \epsilon_{jt},$$

where $b^A_{2,t,j}$ denotes the second order belief of agent j and π_t the success rate in task t.

We then compare the difference in the average parameter of anticipated over-estimation of the agents between treatments with the actual difference in over-estimation by the principals between treatments and the corresponding degrees of anticipated and true information projection. Figure 5 provides an overview of the individual estimates of the agents' anticipated degree of over-estimation of the principals in both information conditions.

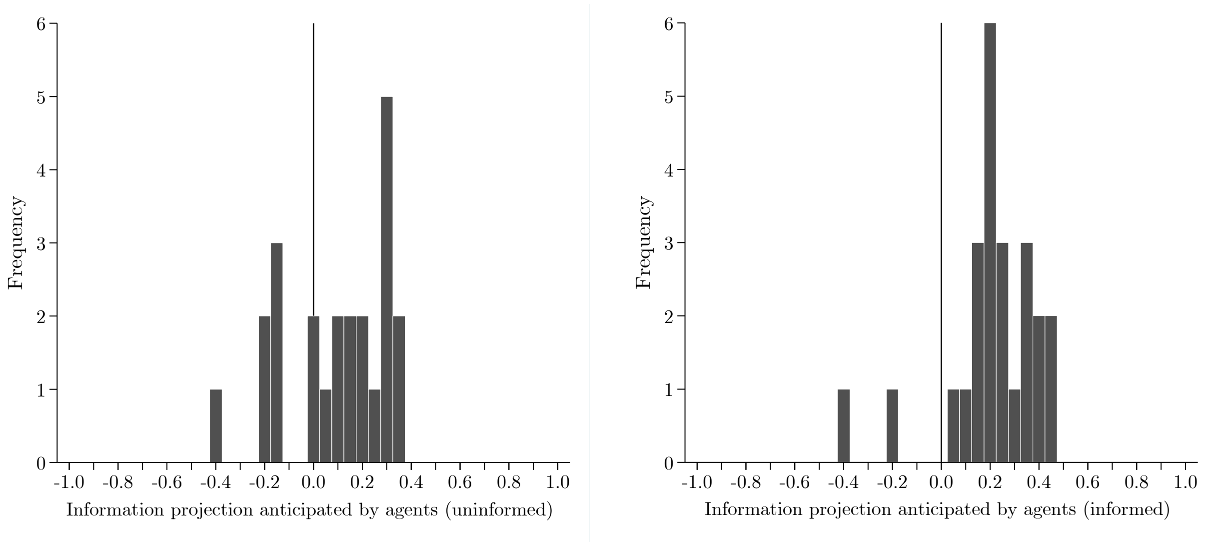


Figure 5: Histogram of individual estimates of the parameter of anticipated overestimation in the uninformed treatment (left panel) and the informed treatment (right panel).

The estimated parameters of anticipated over-estimation of agents matched to informed principals are significantly higher than those of agents matched to uninformed principals on average (one-sided t-test: $p = 0.0285$).[26]

We obtain the maximum likelihood estimate of ρ using the equations from Claims 1 and 2. We calculate d_t, the additional informational value of having the solution, by subtracting the empirical success rate of the reference agents for each task from 1. The difference between the agents' second-order beliefs and first-order beliefs in the informed treatment is calculated by matching participants by task. Recall that ρ is zero in Bayesian Nash Equilibrium. We instead estimate ρ to be 0.31 and significantly greater than zero ($p < 0.01$).

[26] The proportion of (strictly) positive estimates of anticipated over-estimation is significantly larger than expected by chance (0.5) in the informed treatment (91.67%, $p < 0.0001$), but not significantly different (at the 5% level) from 0.5 in the uninformed treatment (69.57%, $p = 0.0606$). The average estimated parameter of anticipated over-estimation is significantly greater than zero in the informed treatment ($\overline{\widehat{\beta}}_{in} = 0.21$; $p < 0.0001$), but also in the uninformed treatment ($\overline{\widehat{\beta}}_{un} = 0.094$; $p = 0.0444$).

4.5 Expected payoffs

Finally, we look at the empirical optimality of the agents' decisions, that is, whether or not to insure against the risk of information-projection-driven overestimation. Payoff comparisons underscore the economic relevance of this bias and anticipation of it in decision-making. Over all rounds and possible matchings with the principals, the average probability that buying the insurance pays off[27] is 33.13% when agents are matched with an uninformed principal and 61.88% when agents are matched with an informed principal. Therefore, expected payoffs from not buying insurance (in each period) are significantly higher when agents are matched with uninformed principals (€6.69) than when they are matched with informed principals (€3.81; t-test: $p = 0.0001$). Note that buying insurance guarantees a payoff of €4. Thus, given the principals' estimates, agents matched to uninformed principals have a clear incentive not to insure while agents matched to informed principals have a (weak) incentive to insure on average. Remarkably, the average insurance rates for agents who were matched to uninformed principals (32.72%) and those who were matched to informed principals (60.83%) are not significantly different from the average probability that insuring pays off in either treatment (t-test: $p = 0.54$ and $p = 0.38$, respectively).

A similar pattern is observed when looking at the empirically optimal insurance decision per period. For each period, we compute the fraction of principals who overestimate the success rate by more than 10 percentage points. We then assess whether insurance per period is empirically optimal by comparing the expected payoffs from not buying insurance per period with the certain payoff of €4 gained from insuring.[28] Agents matched to uninformed principals should insure in 10% of the tasks (2/20 rounds) only, while agents matched with informed principals should insure in 60% of the tasks (12/20 rounds).[29]

Do agents matched to uninformed principals earn more from their insurance choices than agents matched to informed principals? Yes, they do. In the treatments without belief

[27]These are the cases where agents are matched with a principal who (strictly) overestimates the actual success rate in a particular round by more than 10 percentage points. In these cases, not buying insurance pays 0.

[28]Given the principals' estimates, the expected payoffs from not insuring are always strictly different from €4 in each round of both treatments.

[29]Fisher's exact test rejects independence of insurance being payoff-maximizing between treatments ($p = 0.001$).

elicitation,[30] the average actual earning from the insurance choices was €6.96 when agents were matched to an uninformed principal and €4.25 when they were matched to an informed principal.[31] Similarly, the expected earnings given the actual insurance choices (given all possible matchings with principals in the same information condition) is €5.69 when agents were matched to uninformed principals and €3.95 when matched to informed principals (t-test: $p < 0.001$).

Would the payoffs from insurance differ if agents in the informed treatment behaved as agents in the uninformed treatment (and vice versa)? If agents who were matched to uninformed principals behaved like the agents who were matched to informed principals (but are still matched to uninformed principals), they would earn significantly less on average (€5.00 rather than €5.69; t-test: $p = 0.0004$). Conversely, if agents who were matched to informed principals behaved as agents who were matched to uninformed principals (but are still matched to informed principals), they would earn slightly less on average (€3.91 rather than €3.95). However, this difference is not significantly different ($p = 0.39$) and might be part of the explanation as to why the insurance decisions of agents matched to uninformed principals are empirically optimal in 64.45% of the cases, while agents matched to informed principals insure optimally in 49.58% of the cases only.

5 Conclusion

Previous experiments have identified a host of robust biases stemming from people's failure of sufficient informational perspective-taking. This study is the first to document people's anticipation of the resulting information projection by others. We find that not only do principals who receive the answer to change-detection tasks underestimate the difficulty of the task for agents who did not have the answer, agents also anticipate the principals' bias as evinced by their decision to insure against the principals' overestimation of success rates. While information projection biases one's belief about the belief of others, anticipation shows

[30] In these sessions, one of the insurance decision was selected for payment with certainty.

[31] Fisher's exact test rejects independence of actual payoffs from insurance choices between treatments, $p = 0.018$).

that a person is aware that *others'* beliefs about her own beliefs are biased. This fact lends support for the notion of an information projection equilibrium.

The fact that lesser informed agents anticipate the information projection of better informed principals has a number of potentially important economic implications. Our insurance task can be interpreted as the agents' valuation of the relationship with the principal. The more they are worried about the quality of their performance being underestimated by the principals, the less of their resource they are willing to bet on their relationship with the principal. By anticipating information projection, the greater the information gap between an agent and a principal evaluating this agent, the less the agent values the relationship. In the context where agents have to sort into various jobs in the labor market, such anticipation will then affect sorting decisions. The results imply that people might inefficiently choose self-employment and work without principals, even in a context where the gains from working under principals would lead to larger economic gains. If different agents initially have differential informational distance from a principal, this can also lead to misperception-based discrimination against agents who share less of the principals' private information.

The anticipation of information projection has a number of potentially key economic implications to agency settings. In the case of medicine for example, it will lead doctors to distort the ex-ante production of information to minimize the informational gap between the time of the diagnosis and the time of the evaluation. They want to avoid ex-post misperceptions of the ex-ante probability of success which can result in false liability or unduly bad reputation. These concerns can motivate the adoption of severely inefficient medical practices, highlighted by Berlin (2003) and many others (e.g, Studdert et al, 2005; Jackson and Rigli, 2003).

Even more broadly, anticipation of information projection affects optimal contracting and thus the use of incentives and monitoring in organizations. Under rational expectations, better monitoring of outcomes will allow for better sorting of workers and will typically allow the firm to improve efficiency (Holmstrom, 1979). In contrast, under information projection, monitoring with extra information can have the opposite effect. Importantly, the type of reasons identified in the literature under rational expectations whereby more information backfires in agency settings with dynamic career concerns, e.g. Schwartzenstein and Stein

(1990), Cremer (1995), Prat (2005), require different contractual solutions than those used for the anticipation of information projection. As our experiment demonstrated, the latter is present in pure learning problems without *any* private information (action or type) by the agent and thus provides a very basic force. This also has further implications for optimal tort law and regulations of negligence and disclosure.

Our insurance task can be interpreted as the agents' valuation of the relationship with the principal. We then show that a greater information gap between the parties leads to a lower willingness by the agent to invest resources into the relationship on average. When agents have to sort into various jobs in the labor market, such anticipation will then lead to inefficient sorting decisions whereby workers choose inefficient forms of employment or tasks.

Anticipation also leads to defensive choices more generally in order to avoid mistaken preference attributions. Suppose, for example, that a hiring panel needs to choose between two candidates, one coming from a group that has been traditionally discriminated against and one who was not. If the hiring committee - anticipating information projection - wants to avoid the perception of having discriminatory preferences it might too often opt for the candidate from the former group even if the object is select the candidate who is predicted to perform the best. If the minority candidate appears weaker ex-ante but turns out to be better ex-post, biased observers will attribute too much discriminatory preference to the hiring panel.

Future work can shed more light into the determinants of the extent of information projection and the anticipation thereof such as experience, role-switching, incentives, etc. Our findings also highlight the potential effects of anticipated information projection in a variety of economic contexts that have yet to be empirically explored. For example, in bilateral bargaining with asymmetric information, the lack of anticipation of the bias could lead to impasses and breakdowns. The bargaining shares, should a deal be struck, and optimal selling-strategies can also depend upon the extent of information projection as well as the level of anticipation. We can continue to build a better understanding of people's perceptions of this phenomenon to design institutions and mechanisms for economic interactions in the presence of information projection.

References

[1] Armstrong, M. and S. Huck. (2010) Behavioral Economics as Applied to Firms: A Primer. Competition Policy International, 6:3-45.

[2] Baron, J. and J. Hershey. (1988) Outcome Bias in Decision Evaluation. Journal of Personality and Social Psychology, 54:569-579.

[3] Bernstein, D. M., Atance, C., Loftus, G. R., and Meltzoff, A. (2004) We Saw It All Along: Visual Hindsight Bias in Children and Adults. Psychological Science, 15(4): 264-267.

[4] Birch, S. A. J. and Bloom, P. (2007) The Curse of Knowledge in Reasoning About False Beliefs. Psychological Science, 18(5): 382-386.

[5] Camerer, C., Loewenstein, G., and Weber, M. (1989) The curse of knowledge in economic settings: An experimental analysis. Journal of Political Economy, 97(5):1232–1254.

[6] Camerer, C. F. and U. Malmendier. (2007) Behavioral Economics of Organizations. In: P. Diamond and H. Vartiainen (eds.), Behavioral Economics and Its Applications. Princeton University Press.

[7] Cremer, Javques. (1995) The Armslength Relationship. Quarterly Journal of Economics, 110: 275-295.

[8] Danz, D. N. (2013) The curse of knowledge increases self-selection into competition. In: Three essays on irrational beliefs, 61-128. Dissertation, Technische Universität Berlin, Germany.

[9] Davis, M. H. (1983) Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of Personality and Social Psychology, 44(1):113.

[10] Fischbacher, U. (2007) z-Tree: Zurich Toolbox for Ready-made Economic Experiments, Experimental Economics 10(2): 171-178.

[11] Fischhoff, B. (1975) Hindsight / foresight: The Effect of Outcome Knowledge On Judgement Under Uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1:288-299.

[12] Gilovich, T., Savitsky, K., and Medvec, V. (1998) The Illusion of Transparency: Biased Assessment of Other's Ability to Read our Emotional States. Journal of Personality and Social Psychology, 76:743-753.

[13] Gilovich T., Medvec, V., and Savitsky, K. (2000) The Spotlight Effect in Social Judgment: An Egocentric Bias in Estimates of the Salience of One's Own Actions and Appearance. Journal of Personality and Social Psychology, 78:211-222.

[14] Greiner, B. (2004) An Online Recruitment System for Economic Experiments. In: Kurt Kremer, Volker Macho (eds.), Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63, Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, 79-93.

[15] Hölmstrom, B. (1979) Moral Hazard and Observability. The Bell Journal of Economics 10(1):74–91.

[16] Jackson, R. and Righi, A. (2006). Death of Mammography: How Our Best Defense Against Breast Cancer Is Being Driven to Extinction. Caveat Press.

[17] Madarász, K. (2012) Information projection: Model and applications. The Review of Economic Studies, 79(3):961-985.

[18] Madarász, K. (2014) Projection Equilibrium: Definition and Applications to Social Investment and Persuasion. Working paper, London School of Economics and Political Science.

[19] Madarász, K. (2014) Bargaining under the Illusion of Transparency. Working paper, London School of Economics and Political Science.

[20] Malmendier, U. and G. Tate (2008) Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction. Journal of Financial Economics, 89(1):20-43.

[21] Myerson, R. B. (1981) Optimal Auction Design. Mathematics of Operations Research 6(1):58–73.

[22] Newton, E. (1990) Overcondence in the Communication of Intent: Heard and Un- heard melodies. Ph.D. Thesis. Stanford University, Stanford, CA.

[23] Prat, A. (2005) The Wrong Kind of Transparency. American Economic Review, 95(3): 862-877.

[24] Pronin, E., Gilovich, T., and Ross, L. (2004) Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others. Psychological Review, 111(3):781-799.

[25] Rachlinski, J. (1998) A Positive Psychological Theory of Judging in Hindsight. The University of Chicago Law Review, 65:571-625.

[26] Rammstedt, B. and John, O. P. (2007) Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality, 41(1):203–212.

[27] Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997) To see or not to see: The need for attention to perceive changes in scenes. Psychological Science, 8(5):368–373.

[28] Scharfstein, D. and J. Stein. (1990). Herd behavior and investment. American Economic Review, 80, 465- 479.

[29] Simons, D. J. and Levin, D. T. (1997) Change blindness. Trends in Cognitive Sciences, 1(7):261–267.

[30] Spiegler, R. (2011) Bounded Rationality and Industrial Organization. Oxford: Oxford University Press.

[31] Studdert, D., M. Mello, W. Sage, C. DesRoches, J. Peugh, K. Zapert and T. Brennan. (2005) Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment. Journal of the American Medical Association, 293:2609-2617.

[32] Wang, Stephanie W., Michelle Filiba, and Colin F. Camerer. (2010) Dynamically optimized sequential experimentation (DOSE) for estimating economic preference parameters. Working paper, California Institute of Technology.

[33] Wu, D.-A., Shimojo, S., Wang, S. W., and Camerer, C. F. (2012) Shared Visual Attention Reduces Hindsight Bias. Psychological Science, 23(12): 1524-1533.

Appendix

We now briefly describe the definition of an information projection equilibrium (IPE, Madarász 2014) for our setting. Consider a two-player static game with private information $\Gamma = \{\Omega, \phi, \{A_j\}_{j=1}^2, \{u_j\}_{j=1}^2, \{I_j\}_{j=1}^2\}$ where Ω is the state-space, ϕ the prior over the state-space, u_i, A_i and I_j are the payoff function, the action set, and the information set of player i. To incorporate information projection, we have to distinguish between two sets of strategies for each player. Consider player j. Let S_j be the set of (mixed) strategies given actions A_j that are measurable with respect to the information I_j - the set from which the regular variant of player j chooses. Let S_j^+ be the set of (mixed) strategies given actions A_j that are measurable with respect to $I_1 \cup I_2$ - the set from which the super variant of player j chooses.

Below the subscript of the best-response operator reflects whether it is taking with respect to strategies in S_j or S_j^+. Finally, the notation $\cdot$ refers to a compound lottery consisting of two (mixed) strategies.

Definition 2 *A strategy profile $\sigma^\rho = \{\sigma_1^\rho, \sigma_2^\rho\} \in S_1 \times S_2$ is a ρ information projection equilibrium if there exists $\sigma^+ = \{\sigma_1^+, \sigma_2^+\} \in S_1^+ \times S_2^+$ such that for any j*

$$\begin{aligned} \sigma_j^\rho &\in BR_{S_j}(\rho\sigma_{-j}^+ \cdot (1-\rho)\sigma_{-j}^\rho) \\ \sigma_j^+ &\in BR_{S_j^+}(\sigma_{-j}^\rho) \end{aligned}$$

Figure 6: CDFs of principals' average estimate of the agents' success rate (reference treatment) by treatment.

Figure 7: Insurance decisions over time per session.

Figure 8: CDFs of individual insurance rates by treatment.

Table 5: Time trends of insurance decisions (probit regressions).

	Uninformed	Informed	All
1_{Informed}			0.729^{***}
			(0.205)
Period	0.010	0.018^{*}	0.014^{*}
	(0.012)	(0.011)	(0.008)
Constant	-0.586^{***}	0.042	-0.639^{***}
	(0.214)	(0.222)	(0.181)
N	690	720	1410
$\log L$	-432.107	-481.530	-913.861
χ^2_{k-1}	0.650	2.718	17.049

Note: Values in parentheses represent standard errors. Stars represent p-values: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6: Insurance decision conditional on treatment and successful task completion.

Dependent variable	Insurance decision (1-insurance, 0-no insurance)		
(Probit)	(1)	(2)	(3)
Treatment	0.727^{***}	0.754^{***}	0.738^{***}
(1-informed)	(0.205)	(0.211)	(0.221)
Success		-0.429^{***}	-0.451^{***}
(1-task solved)		(0.096)	(0.126)
Treatment×Success			0.040
			(0.190)
Constant	-0.467^{***}	-0.299^{**}	-0.291^{**}
	(0.146)	(0.149)	(0.147)
N	1410	1410	1410
R^2	−916.436	−897.552	−897.512
F	12.575	29.023	29.581

Note: Values in parentheses represent standard errors corrected for clusters on the individual level. Stars represent p-values: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 7: Individual differences between second-order beliefs and first-order beliefs conditional on treatment and successful task completion.

Dependent variable	$(b^A_{2,t,i} - b^A_{1,t,i})$		
(OLS)	(1)	(2)	(3)
Treatment	0.068^{***}	0.068^{***}	0.067^{***}
(1-informed)	(0.019)	(0.019)	(0.024)
Success		-0.039^{***}	-0.041^{**}
(1-task solved)		(0.011)	(0.017)
Treatment×Success			0.003
			(0.022)
Constant	0.044^{***}	0.058^{***}	0.058^{***}
	(0.015)	(0.017)	(0.019)
N	470	470	470
R^2	0.090	0.117	0.117
F	12.828	17.767	13.296

Note: Values in parentheses represent standard errors corrected for clusters on the individual level. Stars represent p-values: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 8: Estimated aggregated parameter of overestimation in both treatments.

	Uninformed principals	Informed principals
$\widehat{\alpha}$	0.0298	0.3113
	(0.0188)	(0.0201)
p-value	0.417	0.000
R^2	0.7033	0.8179
N	480	480

Note: Values in parentheses represent standard errors corrected for clusters on the individual level. A regression based on all principals including a fixed effect for the parameter of information projection in the informed condition yields $p < 0.001$ for a test of the treatment difference in the estimated parameter of information projection against zero.

Table 9: Estimated aggregated parameter of anticipated overestimation in both treatments.

	Agents matched to uninformed principals	Agents matched to informed principals
$\widehat{\beta}$	0.0936	0.2101
	(0.0439)	(0.0405)
p-value	0.044	0.000
R^2	0.8157	0.8698
N	230	240

Note: Values in parentheses represent standard errors corrected for clusters on the individual level. A regression based on all principals including a fixed effect for the parameter of information projection in the informed condition yields $p = 0.0275$ for a one-sided test of the treatment difference in the estimated parameter of information projection against zero.