Projection Equilibrium: Definition and Applications to Social Investment and Persuasion[*]

Kristóf Madarász
London School of Economics[†]

First Version: May 2013. This Version: January 2015 (minor revision February 2015).

Abstract

People exaggerate the extent to which their private information is shared with others. This paper introduces this phenomenon portably into Bayesian games where people wrongly think that if they can condition their strategy on an event others can do as well. I apply the model to a variety of settings. In the context of social investment people misattribute the uncertainty they face about others preferences to others having antagonistic motives. Even if all parties prefer mutual investment, none invests, yet all come to believe that others prefer not to invest. In the context of communication with costly state verification, the model predicts credulity: persuasion by advisors, who are known to have an incentive to exaggerate the quality of an asset, will nevertheless induce uniformly exaggerated average posteriors for receivers. When endogenizing the conflict of interest between senders and receivers, I show that such credulous belief-bubbles rise discontinuously as the size of the market or the complexity of the asset increases. Further implications to auctions, common value trade and zero-sum games are explored.

Keywords: Projection, Social Investment, Pluralistic Ignorance, Persuasion Belief-Bubbles.

1 Introduction

The fact that information is dispersed and different people have access to different pieces of information is key for economics. Hence when interacting with others understanding the extent to which one's information differs from the information available to others is a crucial ingredient of strategic behavior. While it is typically assumed that people fully appreciate such differences, evidence shows that people systematically misperceive informational differences. In particular, the typical person projects information, that is, she exaggerates the extent to which her private information is shared with others, and too often acts as if others could condition their choices on her private information. The goal of this paper is

[*] I would like to thank Pedro Bordalo, Colin Camerer, Jeff Ely, Erik Eyster, Philippe Jehiel, Marina Halac, Paul Heidhues, Marek Pycia, Drazen Prelec, Matthew Rabin, Rani Spiegler, Adam Szeidl, Balazs Szentes, Andrei Shleifer, Tomasz Strzalecki, Jörgen Weibull, Georg Weizsacker as well as audiences at Arizona, Bonn, Columbia, CEU, Harvard, LSE, Stockholm, UCL, UC Los Angeles, UC San Diego, Royal Holloway, Southampton, Berlin Behavioral Seminar, Gerzensee 2013, EBEAM 2013 Berlin and ITAM. All errors are mine.

[†] contact: k.p.madarasz@lse.ac.uk, or London School of Economics, Houghton Street, London, WC2A 2AE UK.

thus to offer a simple portable model incorporating such information projection into strategic settings. By incorporating this robust phenomenon into Bayesian games, the paper allows one to explore a wide range of economic consequences both theoretically and empirically.

Evidence for such information projection comes from a variety of domains. In their classic work on the theory of mind, Piaget and Inhelder (1956) were among the firsts to argue that children too often acted as if lesser informed others had access to their superior information. Brich and Bloom (2007) showed that the same structure of mistake exists amongst undergraduates at Yale in slightly more complex settings. Robust and widely documented evidence on the *curse-of-knowledge* (Camerer, Loewenstein and Weber 1989, Newton 1990, Loewenstein, Moore and Weber 2006)*, hindsight bias* and *outcome bias* (Fischhoff 1974, Baron and Hershey 1988), *illusion of transparency* (Gilovich, Medvec and Savitsky 1998, 2000) show that people exaggerate the probability that if they know a piece of information others should know it as well. Madarász (2012) offers a partial review of the existing evidence, and introduces this phenomenon into dynamic - non-strategic - inference problems.

In a strategic context, Samuelson and Bazerman (1985) provide evidence studying common-value bilateral trade. They argue that privately informed sellers act as if uninformed buyers were symmetrically informed. A companion paper, Danz, Madarász and Wang (2014), provides evidence, not only on informatinal projection, but also that people anticipate a great deal of projection by others and best-respond to it and contains an empirical test of the model presented here.

Model In Section 2 I present the model. A person who exhibits information projection has an exaggerated belief that her opponent uses a strategy that is conditioned on her private information as well. In particular, such a player exaggerates the probability that if she can condition her strategy on an event, then her opponent can condition his strategy on this event as well. The extent of this false belief is characterized by the parameter $\rho \in [0, 1]$.

While there is substantial direct evidence on information projection, there is less evidence from strategic settings where higher-order perceptions matter. I consider two alternative ways to complete the model and consider both private and public information projection. The main focus of this paper is public information projection. A learning rationale and initial evidence on higher-order perceptions, Danz, Madarász and Wang (2014), supports the limited sophistication assumption in the specification, but in the absence of more evidence I present both models.

Private Projection Given private information projection, people do not anticipate the biases of others. Given an initially common BNE of the game, each player believes that her opponent expects her to play according to her strategy in that equilibrium. A biased player i then comes to believe that her private information has been unexpectedly leaked to her opponent with probability ρ. At the same time, she maintains the belief that her opponent never thinks that she attaches positive probability to such leakage. Judith thus thinks that with probability ρ Paul best responds to her original equilibrium strategy using their joint information and that with probability $1 - \rho$ Paul acts as he was initially supposed to. Judith's best response to this perception constitutes her private ρ information projection equilibrium strategy.

Public Projection In a public information projection equilibrium instead, a person assigns probability $(1-\rho)$ to her opponent playing the strategy this opponent actually plays, and probability ρ to a fictional event whereby her opponent best-responds to her actual strategy conditioning such a response on the players' joint information. A public information projection equilibrium - IPE (henceforth) - differs from a BNE by this belief in a fictional opponent on whom one's information is projected. This notion is consistent with feed-back in the following limited sense: a player expects her opponent to behave with positive probability in a way that this opponent always behaves. With probability ρ she also expects the opponent to behave in a way that he may never do.

After illustrating these concepts through some simple examples, I show their existence. I also show that all ex-post equilibria are projection proof. In particular, if a BNE is an ex-post equilibrium, then it is also an IPE- both private and public for any ρ. I further show that in case of private IPE a converse is also true. Namely, if all BNE are ex-post equilibria, then information projection does not alter predictions. In the case of public IPE the same does not hold. I show this, by demonstrating that public information projection can increase the set of Bayesian predictions in games with uncertainty but with perfectly aligned incentives such as the stage game of the classic co-ordinated attack problem, e.g., Rubinstein (1989). Crurcially, I also show that the predictions of IPE cannot be reconstructed as a BNE of an alternative game obtained from the original game by perturbing the information partitions. In particular, the behavior predicted relies - both for private and public projection - on a player believing that her opponent has the *wrong* view of her strategy. Finally, I compare the model to alternative models, such as analogy-based expectations equilibria of Jehiel (2005) and cursed-equilibrium of Eyster and Rabin (2005).

Social Investment In Section 3 I apply the model to the problem of investment into social assets. Partnerships in trade, friendships, cooperation in large organizations, the formation of social and political associations all require people pooling individually owned resources together. Such investment is risky because people face uncertainty regarding the motives of others. Investing with someone who is reciprocal or has matching goals is a source of potential gain. Investing with someone who is opportunistic or has opposing goals is a source of loss. Trust in this settings is associated with a player's belief that her opponent is of a former as opposed to of a latter type.

The model here implies that by projecting information about their own preferences people will misattribute the uncertainty that others face about these preferences into others having antagonistic preferences. By under-appreciating the extent to which others face the same uncertainty about their preferences as they do about the preferences of others, in equilibrium people will come to believe in a form of false antagonism: they will come to believe that the probability that the interest of others is less aligned with theirs than it truly is both conditional on any choice of their opponent and on average as well. This leads to the under-valuation of social assets.

I show that when continued interaction leads to fully efficient investment under Bayesian assumptions, it may still lead to no investment under any positive degree of repeated information projection. Even if all types value social investment, and all behave identically, they will all come to believe that they are alone with such preferences. The mechanism leading to this prediction implies the classic phenomenon described in sociology and psychology as *pluralistic ignorance* whereby "people erroneously infer that they feel differently from their peers, even though they are behaving similarly" (Prentice 2007; see also e.g., Katz and Allport 1931, Miller and McFarland 1987, Prentice and Miller, 1993).

I conclude the Section by exploring comparative static consequences of this mechanism. In the context of trust in trade, preference uncertainty concerns whether others are reciprocal or opportunistic. Here the model predicts the misattribution of the risk of hold-up to genuine mistrust. A longer history of weaker legal institutions will lead to sustained belief that others are untrustworthy even as enforceability of contracts become better.

In the context of political or organizational dissent, preference uncertainty concerns whether others are for or against the prevailing social norm or organizational practice. Here the mechanism predicts a misattribution of silence to loyalty. The costlier it is to express dissent, the more people who oppose the status-quo will come to exaggerate the extent to which others support the status-quo. Furthermore, after the cost from expressing dissent drops to a sufficiently low level, dissenters are well surprised to see how prevalent their attitude is in the population.

In the context of friendships or dating, uncertainty concerns whether one's partner is interested in forming a serious bond. Here the model predicts false segregation. The greater is the uncertainty a person faces about the preferences of an opponent, the more she will develop a sense of false antagonism. The less two people initially know about the intentions of each other, the more they will infer that the potential partner is not interested if they are interested, and interested if they are not interested.

Persuasion In Section 4 I apply the model to communication with costly state verification. I show how the model predicts a strong form of credulity: too much optimism when receiving good news from an advisor with commonly known incentives to exaggerate the truth. I also show how embedding this problem into a setting with endogenous conflicts of interests, the models predicts the rise and fall of belief bubbles as the size of the market changes.

A privately informed advisor sends a message to an investor whether a statement is true or false, e.g., an asset has high or low expected returns. The advisor's preference are misaligned towards claiming that returns are high. Investors have private information about the cost at which they can verify the advisor's recommendation. Differences in such cost might reflect differences in financial expertise or access to additional information sources.

Bayesian persuasion has two identifying properties: it improves the welfare of receivers on average, and it is neutral in that the average posterior of the receivers is equal to their prior. Under projection both of these properties are violated. The model identifies settings where persuasion will (i) strictly lower the welfare of receivers and (ii) move average beliefs systematically above the truth. In particular, the model predicts uniform credulity: it specifies conditions such that all receiver types will be too optimistic when hearing the advisor's positive recommendation and hence will over-invest in the asset. Such credulity is consistent with the evidence, e.g., Malmendier and Shanthikumar (2007), Della Vigna and Kaplan (2007).

Specifically, biased receivers will exaggerate the extent to which the advisor's incentive to lie is tailored to their own expertise. As a consequence, when hearing the advisor's messages, receivers with low cost of checking will be too credulous, while receivers with high cost of checking will be in disbelief. In a public IPE a sophisticated advisor, if the asset is sufficiently complex or the misalignment of his preference is sufficiently strong, he will always lie more often than any receiver type - except those for whom it is dominant strategy not to check - expects. Uniform credulity is more likely the greater is the conflict or the more common are receiver types who are neither clueless nor experts.

Finally, I endogenise the misalignment between the advisor and the investors, by considering the seller of the asset who observably pays the advisor for high recommendations. I show that while in the Bayesian case, the seller would never want to pay the advisor because persuasion is neutral, in the biased case, paying the advisor might be beneficial. In particular, the model predicts that if the size of the market is sufficiently large, the seller always pays the advisor and implements uniform credulity and exaggerated average beliefs, whereas if the size of the market is below this critical threshold, the advisor does not pay the advisor, and average beliefs are truthful. Thus the model predicts the discontinuous rise and burst of 'belief-bubbles' as the size of the market rises or shrinks. Policy implications on capping the payment from the seller to the advisor are explored.

Projection equilibrium Section 5 combines information projection with ignorance projection and presents the resulting notion of (public) *projection equilibrium*. In short, here a player projects both what she knows and also what she knows exaggerating the probability that her opponent has the same information she does. I derive implications of projection equilibrium to common value trade and compare the predictions with existing experimental data, and show how it compares to predictions of cursed equilibrium. Finally, I present the multi-player extensions of the models considered.

2 Model

Consider a Bayesian game Γ. For ease of exposition I restrict attention to games with two-players.[1] Let A_i be player i's finite action set and $u_i(a,\omega) : A \times \Omega \to \mathbb{R}$ be her bounded payoff function which depends on the action profile $a \in A$ and the state ω. The state is an element of a finite set Ω. Let π be the common prior over this state-space which I assume to be strictly positive. Information in this environment is given by partitional information correspondences $P_i(\omega) : \Omega \to 2^{\Omega}$ for each i. The game can then be summarized by the tuple $\Gamma = \{A_i, u_i, \Omega, \pi, P_i\}$.

To introduce information projection, consider the joint information of player i and player j as expressed by the following information correspondence:

[1] I will discuss later in this Section how to extend the definition to games with N players.

$$P^{+}(\omega)=\{\ \widehat{\omega}\ \mid\ \ \widehat{\omega}\ \in P_i(\omega)\ \cap\ P_j(\omega)\ \}\text{ for all }\omega\in\Omega \tag{1}$$

This correspondence P^{+} is also partitional and is equivalent to the coarsest common refinement of the original partitions of players i and j. It expresses the information distributed amongst the players. In particular, if an event, $E\subseteq\Omega$, is known at a state ω under either of the original partitions, i.e., if at least one of the players knows that something is true, then it is known at the state under P^{+} as well. Let me now distinguish between two sets of strategies for each player i.

Real Player The first set consists of the possible strategies of the real player i. These are the strategies over player i's action space that are conditioned on player i's *true* information P_i. The real player i always chooses a strategy from:

$$S_i=\{\sigma_i(a_i\mid\omega)\in\Delta A_i\ \text{ measurable with respect to } P_i\} \tag{2}$$

Fictional Player The second set consists of the strategies that would be available to player i if he could condition his strategy on the joint information. These are the strategies over player i's action space that are conditions on the joint information P^{+}.

$$S_i^{+}=\{\sigma_i^{+}(a_i\mid\omega)\in\Delta A_i \text{ measurable w.r. to } P^{+}\} \tag{3}$$

Information Projection I now turn to the first part of the definition. Information projection by player j corresponds to a mistaken belief that player i is choosing a strategy from S_i^{+} as opposed to S_i. The extent of this mistaken belief will be parametrized by $\rho\in[0,1]$. If player j projects information to degree ρ, she assigns probability ρ to the event that her opponent chooses his strategy from S_i^{+} and probability $(1-\rho)$ that he chooses it from S_i . This fictional type who chooses from S_i^{+} only exist in player j's mind.

I introduce two related ways to complete the model. They both incorporate the above belief, but differ in how a person thinks about (i) her opponent's view of her behavior and (ii) her opponent's views of her view of her opponent's behavior. The paper will mainly focus on the second specification, but presents the first alternative as well.

Private A *private information projection equilibrium* will describe a setting where people initially share a common view of what happens in the game that accords with a true BNE of the game. By projecting information, player i then comes to privately believe that her - and only her - information has been shared with her opponent with probability ρ. At the same time, she believes that his opponent does not recognize the fact that she believes this now.

Intuitively, Judith thinks that although her cards have leaked to Paul with probability ρ, Paul does not realize that she thinks this way. Judith thinks that Paul believes that she assigns probability 0 to such information sharing (leakage). Hence projection here is private.

Public A *public information projection equilibrium*, will describe a setting where players understand their real opponent's behavior. Specifically, each player will assign probability $1-\rho$ to her opponent behaving the way he always behaves. By projecting information, however, each player i, assigns probability ρ to her opponent being a fictional type who best responds on the basis of the joint information and the knowledge of player i's strategy.

Intuitively, Judith now thinks that with probability $(1-\rho)$ Paul has not seen her cards, in which case Paul believes that she puts positive probability on him seeing her cards. In addition, Judith thinks that with probability ρ Paul did see her cards, in which case Paul again believes that she puts positive probability on him seeing her cards, but also knows that she has not seen his cards. Hence people do anticipate each others projection, which in this sense is public

Crucially, while there is a lot of evidence supporting the basic notion of information projection, there is less evidence from strategic settings where higher-order perceptions are key as well.[2] The disciplined difference between private and public projection leads to different predictions and observable comparative statics; exploring their validity given the model offered in this paper is the task of future empirical work. In addition, some strategic domains might naturally lend themselves to predictions of private projection while settings with more feed-back or more anticipation of projection by others lend themselves to the predictions of public projection. I will discuss these issues in more detail after presenting the definition.

2.1 Private Information Projection

To state the definition of private projection equilibrium, I need to distinguish between strategies that are played in equilibrium, and strategies that describe players' beliefs about how how their opponents behave. As mentioned, in contrast to the Bayesain equilibrium where these always coincide, these need not correspond to each other.

I denote the strategy profile which describes how people truly behave in the game by the profile $\sigma^{\rho}\in S_i\times S_j$. Since people can only condition their true strategies on the information they truly have, this strategy profile is an element of the true strategy space.

I denote the strategy profile which describes people's view of their opponent's behavior by a probabilistic mixture of two strategy profiles. In particular, each player i will believe that with probability $(1-\rho)$ her opponent picks a strategy σ^0_{-i} from S_{-i} and with probability ρ he picks a strategy σ^+_{-i} from S^+_{-i}. The compound lottery that assigns probability $1-\rho$ to strategy σ^0_{-i} and ρ to the strategy σ^+_{-i} will be denoted by $(1-\rho)\sigma^0_{-i}\circ\rho\sigma^+_{-i}$.[3]

Definition 1 *A strategy profile $\sigma^{\rho}\in S_i\times S_j$ is a private ρ information projection equilibrium (IPE) of*

[2] In a one-sided private information setting Danz, Madarasz and Wang (2014) find strong evidence not only for information projection but also that people - even without much feed-back - people correctly anticipate information projection by others, lending some initial support for public projection.

[3] The notation BR_{S_i} stands for the best-response operator when player i picks a strategy from S_i and takes expectations given partition P_i to maximize her expected utility. Similarly, $BR_{S^+_i}$ stands for the best-response operator when player i picks a strategy from S^+_i and takes expectations given partition P^+ to maximize her expected utility.

Γ *if there exists strategy profiles* $\sigma^0 \in S_i \times S_j$ *and* $\sigma^+ \in S_i^+ \times S_j^+$ *such that for all* i*,*

1.

$$\sigma_i^\rho \in BR_{S_i}((1-\rho)\sigma_{-i}^0 \circ \rho\sigma_{-i}^+)$$

2.

$$\sigma_{-i}^0 \in BR_{S_{-i}}(\sigma_i^0) \text{ and } \sigma_{-i}^+ \in BR_{S_{-i}^+}(\sigma_i^0)$$

■ Above the strategy profile σ^0 always describes a Bayesian equilibrium of the game. This profile describes people's initial shared view of behavior in the game. The definition thus corresponds to a parametric extension of BNE. If $\rho = 0$, there is no projection and there is no change in behavior relative to the suggested Bayesian equilibrium. Here for any given σ^0, the strategy profile $\sigma^\rho = \sigma^0$ satisfies the definition. If $\rho > 0$, the model deviates from that of BNE. Specifically, given σ^0, describing how players initially expect each other to behave, a biased player i mistakenly assigns probability ρ to the event that her opponent best responds to her Bayesian equilibrium strategy, σ_i^0, by conditioning her action on the joint information in the game, σ_{-i}^+. She assigns the remaining probability to her opponent using only his true information, and this acting as before, σ_{-i}^0. Player i's private IPE strategy σ_i^ρ is then a best-response to such wrong beliefs

■ For simplicity, I assumed that players were equally biased and hence their degree of projection was the same. The definition extends immediately to the case where ρ is a vector and players can be differentially biased. Furthermore, note that since players' deviations from the suggested Bayesian equilibrium are governed solely by their own degree of information projection, a change in their opponent's degree of information projection does not affect their behavior.

■ A private IPE consists of a minimal deviation from a BNE of the game in the following sense. Consider again the suggested Bayesian equilibrium σ^0. This equilibrium defines people's beliefs about their opponent's strategies which then corresponds also to higher order beliefs about these strategies. In a private ρ-IPE, it is *only* a player's first-order belief about her opponent's strategy that is changed. All higher-order beliefs about strategies remain the same. In particular, player i thinks that player j plays σ_{-i}^0 for sure and thinks that player j thinks that player i plays σ_i^0 for sure and so on. This means that the belief that opponent picks a strategy from S_{-i}^+ as opposed to S_{-i} enters only into first-order beliefs about strategies.

■ Note that in this definition players best respond to a wrong theory of their opponent's behavior - one that does not contain in its support the truth - but that these wrong theories are derived from a common heuristic about play in the game. This feature of a private IPE links this model to cognitive hierarchy models of strategic behavior, e.g., Stahl and Wilson (1995), Camerer, Ho and Chong (2004). In those models a player's theory of how her opponent plays might not contain in its support how that opponent actually plays for a similar reason as here. In both cases a person's theory of her opponent's theory of how she behaves need not cover the why she actually behaves. An important difference between my model and these models, especially when applied to incomplete information games, that these models leave open the specification of what the underlying level 0 heuristic is. In contrast, in the above model expectations are anchored to a BNE of the underlying true environment. Players deviate from this purely as a function of the single parameter ρ through the logic of information projection. This renders the model both portable and empirically easy to test, and allows for clear comparative statics with respect to the distribution of information without needing to re-define the model.[4]

■ Finally, an important feature of the above model is that finding a private IPE is very simple given a BNE of the game. It simply involves calculating individual - as opposed to mutual - best-responses. This feature of a private IPE makes it particularly easy to apply it to settings where the set of BNE is well understood.

2.1.1 Example 1: Zero-sum Games

To illustrate the model, consider a hide-and-seek game. Each player picks one of two locations: A or B. If the defender is strong, $\omega = 0$, she wins iff the players pick the same location. If she is weak, $\omega = \omega_w > 0$, then even if they both pick A she wins only with probability $1 - \omega_w$. When the defender is weak A is her Achilles heel. Formally,

$$\begin{array}{ccc} \text{attacker/defender} & \text{A} & \text{B} \\ \text{a} & \omega, 1-\omega & 1, 0 \\ \text{b} & 1, 0 & 0, 1 \end{array} \tag{4}$$

D-Day (Calais-paradox) To illustrate, consider one of history's most noted zero-sum games: the landing of the Allies on the shores of Normandy on June 6th of 1944, (D-day). Here the Allies had the choice to land at Calais (A) or Normandy (B). The Axes had to decide to concentrate troops at one of these two locations. There was good reason to believe that Calais would be the easier terrain for an attack. German forces occupying both locations also had some private information whether this was true or not. The historic success of D-day is typically attributed to the Axes' firm expectations that an attack would take place at Calais and thereby defending it.

Suppose the state is the defender's private information and the ex-ante each state is equally likely. The table below summarizes the defender's strategy in the unbiased and the fully biased case. Since

[4] If in a *level k* model, level 0 types play according to the equilibrium, then the predictions of the level k model are equivalent to equilibrium predictions. Furthermore, note that it is easy to see that independent of what level 0 heuristic is assumed or what the shape of the cognitive hierarchy is, these models differ from IPE, because here players always get their opponent's strategy space right. In contrast the key point of this model is that people have wrong models about the *strategy space* of their opponents.

the attacker has no private information he mixes symmetrically, in both settings.

defender	weak	strong	EU_D^ρ
$\rho = 0$	B	A	$\frac{1}{2}$
$\rho = 1$	A	B	$\frac{1}{2} - \frac{\omega_w}{4}$

Under BNE the defender *hides* optimally behind her private information: she defends A when strong, and B when weak. Hence she never defends her Achilles heel, and wins half of the game irrespective of ω_w. In contrast, a fully biased defender always plays her Achilled heel. Thinking that her type has leaked to the attacker, but that he does not realize that she recognizes this, when strong, she expects him to attack at A and when weak, she expects him to attack at B. Her best response is then to defend A when weak and B when strong. The next observation implies that even as it becomes ex-ante virtually certain that the defender is weak, $p \to 1$, while the BNE converges to the defender mixing symmetrically, any ρ-IPE converges to defending her Achilles heel for sure. Specifically,

Claim 1 *Note that for any p, $\sigma_2^\rho(A\,|weak) = 1$, iff $\rho > 0$.*

Finally, let me present a reversal of the key informational comparative static result. For any given prior over the states π, I compare the defender's ex-ante equilibrium winning probability in two cases: (i) the defender is privately informed about ω as above, (ii) she only knows the true prior. I present a 'chocking' effect. While in the Bayesian case, private information has positive value for the defender. In the fully biased case, it has negative value to her.

Claim 2 (Negative Value of Private Information) *For all π, if $\rho = 0$, the defender wins more often in (i) than in (ii), if $\rho = 1$, the reverse is true.*[5]

2.1.2 Example 2: IPV Auctions

As a second-example, consider a symmetric independent private-value auction problem. Suppose each player's valuation is distributed according to some π over a finite set of valuations, $v_1 < v_2 < .. < v_N$. The classic Bayesian result in this setting is r*evenue equivalence*: the seller's expected equilibrium revenue is independent of whether a first- or a second-price auction is adopted, e.g., Riley (1989). The result below shows that revenue equivalence is systematically violated for any degree of information projection. Consistent with much of the existing evidence - for a survey see Kagel (1995) - there is always an equilibrium where players over-bid in the first-price auction relative to the second-price auction. Furthermore, here the increase in revenue is discontinuous as one moves from no-bias to any positive bias.

Claim 3 *If $\rho = 0$, revenue equivalence holds. If $\rho > 0$, there always exist a private IPE such that the first-price auction generates discretely higher revenue than the second-price auction.*

[5] It is easy to see that the payoff predictions delivered above cannot be derived using a level k model with any common level-0 heuristic. Furthermore, the same will hold if one explores the comparative static predictions.

Note first that a second-price auction has an ex-post equilibrium. As Proposition 3 will show this implies that the Bayesian predictions are unchanged. Consider now the first-price auction. The key feature of the BNE is that players shield their bids below their valuations and collect information rents. By projecting information a player comes to believe that if her opponent has a lower valuation than hers, he now has an incentive to bid higher. In contrast, if he has a higher valuation, he now has an incentive to bid lower. Both of these (fictional) effects imply that a biased player has an incentive to increase her bid. If valuations are discrete, then each type bids on an interval that has positive measure. Hence by projecting information the increase in revenue is discrete when moving from the case of $\rho = 0$, to the case where $\rho > 0$.[6][7]

2.2 Public Information Projection

Let's now turn to the main model of the paper: public information projection equilibrium. Let again σ^{ρ} describe the strategy profile that players actually play in equilibrium. People's theory of their opponent's strategy corresponds to a probabilistic mixture of two strategy profiles. Each player i puts probability $(1-\rho)$ on her opponent playing the strategy that he actually plays, σ^{ρ}_{-i}, and probability ρ to a strategy that her fictional opponent would play if he best-responded to her strategy knowing the joint information in the game as well as her strategy, σ^{+}_{-i}.

Definition 2 *A strategy profile $\sigma^{\rho} \in S_i \times S_j$ is a public ρ information projection equilibrium of Γ if there exists $\sigma^{+} \in S^{+}_i \times S^{+}_j$ such that for all i,*

1.

$$\sigma^{\rho}_i \in BR_{S_i}((1-\rho)\sigma^{\rho}_{-i} \circ \rho\sigma^{+}_{-i})$$

2.

$$\sigma^{+}_{-i} \in BR_{S^{+}_{-i}}(\sigma^{\rho}_i)$$

■ A public ρ IPE is a directional extension of the BNE of the game. The model departs from that of BNE only in people assigning probability ρ to a fictional strategy of their opponent. If $\rho = 0$, each players puts full probability on how her opponent actually behaves and thus the definition corresponds to the definition of a BNE. If $\rho > 0$, a biased player i assigns probability ρ to a fictional event that her opponent plays a strategy that is a best-response to her strategy and is conditioned on the joint information in the game. Judith's public $\rho - IPE$ strategy is a best-response to such a wrong theory of Paul's behavior.

[6] As noted by many authors, a failure of quasi-linearity of the payoff functions, can result in over-bidding in the first price auctions. The experimental evidence does not support the idea that the observed over-bidding in first-price auctions is consistent with it being due to simply risk-aversion, see Kagel (1995, pp525).

[7] Note that because this is a private value environment, cursed equilibrium or behavioral equilibrium here makes the same predictions as BNE.

■ As before, the definition extends immediately to the case where $\rho = (\rho_i, \rho_{-i})$ is a vector and players can be differentially biased. Players who are differentially biased differ in the extent to which they expect their opponent to be omniscient. Specifically, if for example $\rho_i = 0$, then player i has correct expectations about the strategy of her opponent while this opponent only attaches probability $(1 - \rho_{-i})$ to player i's true strategy. Here if a player i is unbiased, she has a $\rho_i = 0$, this person is also sophisticated: she puts full weight on how her opponent actually behaves in the game and thus her action is optimal given her information and the truth.

■ The definition here allows people to exhibit *partial sophistication* about the information projection of others. Specifically, the definition ties together two realistic features in a parsimonious manner: people anticipate their opponent's information projection, and do so exactly inversely to the extent to which they themselves project information. Judith puts weight $(1 - \rho_i)$ on facing her real opponent and ρ_i on facing the projected opponent. Since real Paul's strategy is based on him projecting information to degree ρ_{-i}, thus with probability $1 - \rho_i$ Judith effectively expects Paul to project information of degree ρ_{-i}. At the same time, the projected Paul knows what information Judith has, and thus with probabilityρ_i expects Paul to have correct beliefs about her information. I refer to this feature of information projection as *all-encompassing*. Such all-encompassing projection has a number of consequences that I highlight now.

■ First, real and fictional versions of a player have different beliefs about her opponent's strategy. The real player i puts probability weight $(1 - \rho_i)$ on her opponent being the real version, playing strategy σ^{ρ}_{-i}, and probability weight ρ_i on her opponent being the fictional version playing strategy σ^{+}_{-i}. In contrast, the fictional version of player i is believed to put probability 1 on her opponent being the real version and playing σ^{ρ}_{-i}.

■ Second, by the above token, player i *underestimates* the probability with which her opponent thinks that she is a super (projected) version. A biased Judith thinks that Paul thinks that Judith is super with probability $(1 - \rho_i)\rho_{-i}$, while in reality Paul thinks Judith is super with probability ρ_{-i}. Such underestimation stems from Judith's own projection. Particularly, if Judith is unbiased, $\rho_i = 0$, she correctly estimates this probability to be ρ_{-i}. In contrast, if she is fully biased, $\rho_i = 1$, she perceives this probability to be 0: Judith thinks that her opponent knows for sure that what information she has, e.g., that she does not know his private cards. Danz, Madarasz and Wang (2014) testing the model find evidence for this distinct implication of the model.

■ Third, the above implies that what real Judith thinks about Paul's strategy, does *not* agree with what real Paul thinks Judith's perception of his strategy is on average. This fact differentiates the model from a case where people might disagree with their opponent's view of their strategy, but where people's perception of their opponent's strategy can be understood as commonly known.

■ Fourth, even in games with one-sided private information, the degree to which the lesser-informed player is biased will thus matter. A lesser-informed player's degree of projection affects her perception with which her opponent knows that she is uninformed. By projecting information, the lesser informed party exaggerates the probability that her opponent attaches to the event that she does not have his privately known cards. In the extreme, if the lesser informed party is fully biased, she thinks that her opponent realizes that she does not have his private information for sure.

■ Fifth, since Judith puts positive probability on her opponent playing the way he actually plays, the model is consistent - in a *limited* way - with full feed-back in the environment. A biased player assigns probability $(1-\rho)$ to the strategy that her opponent actually plays. Hence what she expects can happen in equilibrium is consistent with what will happen in equilibrium. In this regard, the way (public) information projection equilibrium differs from BNE is that Judith expects something to happen with probability ρ that might never happen or might happen with a different probability. Nevertheless her theory is never explicitly contradicted either.

2.2.1 Example 1: Continued

Let's return to Example 1. Consider now public IPE. It remains true that a biased defender *over-protects* the weak location as before. In addition, the defender also *over-mixes* relative to the BNE. The example will help to highlight two additional facts. First that the lesser informed player's bias does affect the predictions. Second, that players might use strategies that would never be equilibrium strategies even in games that are derived from the original game by perturbing information partitions.

Consider the case where the players are differentially biased. Suppose the defender is sufficiently biased and the attacker is unbiased. I maintain the assumption that ex-ante the states are equally likely. The table below describes the$\rho = (\rho_D, \rho_A)$ public information projection equilibrium:

		weak	strong
$\rho_D \geq 1/(2-\omega_w)$	defender	$\frac{1}{2-\omega_w}\text{A} \circ \frac{1-\omega_w}{2-\omega_w}\text{B}$	$\frac{1}{2}\text{A} \circ \frac{1}{2}\text{B}$
$\rho_A = 0$	attacker	b	b

Can public IPE be re-constructed as BNE with Perturbed Partitions? Consider the set of BNE of any perturbed game Γ' that is derived from Γ by considering some alternative set of information partitions P_1' and P_2'. If the attacker's strategy is constant across states, as it is here, then the defender always best responds by playing a constant strategy over states, independent of the information she has. This renders the attacker's strategy inconsistent with any BNE of a perturbed game Γ'.

This example demonstrates that IPE has novel empirical content relative to BNE even when players have different perceptions about the Bayesian game they are playing. The reason for this is as follows: the fact that the attacker plays *b* for sure (lands at Normandy) above rests on the fact that the attacker believes in equilibrium that the defender has false beliefs about his strategy. If the attacker (even if mistakenly) believed that the defender has correct beliefs about his strategy, it would never be an equilibrium to play *b* for sure. The novel empirical content of IPE relies on the fact that players here believe that their opponents systematically mispredict their choices. The Allies in equilibrium believe that the Germans exaggerate the probability with which they will land in Calais and such belief is key for their strategy choice.

2.3 Discussion and Related Literature

Having presented the model, let me turn to some of its basic properties. I first show that a ρ projection equilibrium - both private and public - always exists. Furthermore, both models differ from BNE only when players are differentially informed. I then turn to a key difference between public and private projection. I show that under both models if a BNE is an is an ex-post equilibria, then it is always an IPE for any ρ. Hence ex-post equilibria are projection proof. I then show a converse, namely that if all BNE are ex-post equilibria, then IPE is equivalent to BNE only holds for private projection. Under public IPE the model implies illusory co-ordination on ex-post equilibria that are not ex-ante BNE. The first result claims existence.

Proposition 1 *For any Γ and ρ both a public and a private ρ information projection equilibrium exists.*

In case of the private projection equilibrium this follows immediately since expectations are anchored to an underlying BNE of the game and best-responses are well-defined. In the case of public projection equilibrium, it follows directly from Kakutani's theorem since mis-perceptions are continuous in the strategy space. The next result shows the corollary that the model delivers novel predictions only to the extent that the players are differentially informed.

Corollary 1 *If $P_i(\omega) = P_j(\omega)$ for all ω, then all IPE - private or public - are BNE.*

Let me now turn to the relation of the model to notion of an ex-post equilibrium.[8] Here players' strategies satisfy an ex-post no-regret condition: a BNE strategy profile is an ex-post equilibrium if no player has an incentive to deviate even conditional on the realization of the state. Hence it is robust to ex-post deviations once the state is learned. Formally, σ is a strict ex-post equilibrium if the following is true. For every i and ω

$$u_i(\sigma \mid \omega) > u_i(\sigma^0_{-i}, a_i \mid \omega) \text{ for all } a_i \in A_i$$

In contrast to BNE and IPE, an ex-post equilibrium often does not exist given the players information and hence in many games this concept provides no predictions. The following proposition states that when it does exist an ex-post equilibrium is also robust to information projection.

Proposition 2 *If a BNE is a strict ex-post equilibrium in Γ, then it is also a ρ-IPE - private or public - for all $\rho \in [0, 1]$.*

Intuitively, since such an equilibrium satisfies a no ex-post deviation condition, it follows that even if a player knows the true state, she has no incentive to deviate. As a result, when Judith believes that

[8] For the common use of this concept in the context of auction and mechanism design see e.g., Cremer and McLean (1985) or Dasgupta and Maskin (2002).

Paul knows her private information, if they are playing an ex-post equilibrium, Judith has no reason to believe that Paul has any incentive to deviate.

A converse is also true for private IPE but not for public IPE. Let me first state the result for private IPE. If all BNE satisfy the strict no-regret condition, then the model of BNE and the model or private IPE coincides.

Proposition 3 *Suppose all BNE are strict ex-post equilibria in* Γ*. Then all private IPE are BNE.*

The logic of this proposition follows simply from the fact that a private IPE is linked to a BNE of the underlying game. If this Bayesian equilibrium is an ex-post equilibrium, then by the above proposition it is projection proof. Hence the predictions of BNE and that of private IPE are identical here.

To show why this need not hold for the case of public IPE consider a simple example. Let there be two states $\omega \in \{\omega_1, \omega_2\}$. Suppose one player is privately informed about the state ω and suppose ω_R is ex-ante less likely. The payoffs and actions are given as follows, where $\varepsilon > 0$:

ω_L	L	R	ω_R	L	R
l	$1,1$	$0,-\varepsilon$	l	$0,0$	$0,-\varepsilon$
r	$-\varepsilon,0$	$0,0$	r	$-\varepsilon,0$	$1,1$

In this game all BNE requires the players to play a constant action across states. The predictions of private IPE coincides with this. Consider now public IPE. If ρ is sufficiently high, then there is a ρ IPE with $\{(L,l,\omega_1);(R,l,\omega_2)\}$ and hence the privately informed player tunes his action to the state: expecting illusionary co-ordination. To see this, note that if $\rho = 1$ the informed party can believe that if the state is ω_R, she should play R because her opponent who knows her strategy and her information will play R in response and then she has no reason to deviate. By continuity, the same will hold for lower ρ.

Related Literature A growing literature in game theory considers approaches where people fail to correctly understand the underlying environment in which they act. Jehiel (2005) studies analogy-based expectations. Eyster and Rabin (2005) study behavior under the assumption that a person correctly understand informational differences, but believe that with some probability her opponent plays the identical strategy over all of his information sets.[9] The identifying assumption in both Jehiel (2005) and in Eyster and Rabin (2005) is that people on average have correct expectations about their opponents' strategy. In contrast, in an IPE players have wrong beliefs about the strategies of their opponent on average as well. Furthermore, the logic of iiformation projection differs significantly from cursedness. For example, in all private value environments, such as the social investment problem considered below, cursedness makes the same predictions as BNE, while information projection makes different predictions. In Section 5 of thos paper

[9] Esponda (2008) considers a related model with the same feature where players' cursed perceptions are constrained by their observations of payoffs.

3 Social investment

Efficient outcomes often require people to pool individually-owned resources together and make social investments. Partnerships in trade, friendships, production in large organizations or the formation of social and political associations all require the pooling of material, temporal, or informational resources. More often than not, interaction in such settings takes place under some risk whereby people face uncertainty regarding the motives and goals of others. Investing with someone who shares the same goals is a source of potential gain, investing with someone who does not, is typically a source of loss. Hence a key component of such interactions is trust: the belief that one's opponent is of the former type.

For example in the context of trade, preference uncertainty concerns whether one's opponent is reciprocal or not. Such uncertainty is potentially key, because as Arrow (1972) argues, "virtually every commercial transaction has within itself an element of trust" and "much of the economic backwardness in the world can be explained by the lack of mutual confidence". When contracts are incomplete or badly enforced, such trust is a key ingredient of economic exchange. Social investments in large organizations, where people interact less frequently, have been widely considered and found to be a key determinant of economic success, e.g., La Porta et al. (1997), Algan and Cahuc (2010). Let me now turn to the application of the model of public information projection equilibrium to such a problem of social investment.

The main result of this Section shows that information projection causes people to misattribute the risk inherent in social investment into antagonistic preferences. By under-appreciating the extent to which others face the same uncertainty about their preferences as they themselves do about the preferences of others, people come to believe that the interest of others is less aligned with theirs than it truly is. Continued interaction under preference uncertainty breeds mistrust and leads to the under-appreciation of social assets. Furthermore, the model predicts that people will come to develop a sense of *false uniqueness.* Interacting in a perfectly symmetric situations, and acting identically, people will nevertheless come to infer that all others have the opposite preference as they do. The resulting phenomenon has been classically described in sociology and psychology under the rubric of *pluralistic ignorance* (e.g., Prentice 2007).

4 Setup

Consider a social investment problem. Each player i has a privately observed type θ_i describing her valuation from investment. If this type is positive, player i gains from mutual investment, if it is negative, she suffers a loss from investment. Upon observing their own valuations, parties decide independently whether to enter or stay out. If both enter, each realizes his or her own valuation. If both stay out, each gets an outside option of zero. The game is described as follows:

$$\begin{array}{ccc} & \text{In} & \text{Out} \\ \text{In} & \theta_1 \ , \ \theta_2 & g(\theta_1,\theta_2) \ , \ f(\theta_2) \\ \text{Out} & f(\theta_1) \ , \ g(\theta_2,\theta_1) & 0 \ , \ 0 \end{array} \tag{5}$$

where each θ_i is distributed i.i.d. according to a uniform density on $[\theta_{\min}, \theta_{\max}]$ with $\theta_{\min} < 0 < \theta_{\max}$.

A key distinction will be between positive and negative types. Positive types who value investment are assumed to be reciprocal: if their opponent invests they prefer to invest as well. In contrast, negative types who do not value investment are assumed to be 'opportunistic': if their opponent invests they prefer not to invest. This is captured by the Sorting condition below. Furthermore, one-sided investment is risky because investing with a non-entering negative type leads a payoff lower than that of no-investment. This is captured by the Investment Risk condition below. Formally, I assume that:

1. **Sorting** $f(0) = 0$ and $f' \in (0,1)$.

2. **Investment Risk** If $\min\{\theta_i, \theta_{-i}\} < 0$, then $g(\theta) < 0$,

3. **Monotonicty** $g_1, g_2 \geq 0$, and $g(0, \theta_{-i}) = 0$ for all $\theta_{-i} \geq 0$.

In the above specification a negative type prefers the outside option to mutual investment. The analysis generalizes to the case where negative types may also prefer mutual in to mutual out, but will still not want to reciprocate a uni-lateral investment by their opponent. In particular, subtracting any positive constant b from the outside option for *each* type, leaves the analysis unaffected, as long as if $\min\{\theta_i, \theta_{-i}\} < 0$, then $g(\theta) < -b$, that is negative types would want to stay out even if their opponent enters. In some of the interpretation, I will invoke the case where all types prefer mutual investment to mutual no-investment. Here, if all types were negative, the game would correspond to a prisoner's dilemma. If all types were positive, to a coordination game. Below, I illustrate the main results in an example where investments are almost perfect substitutes. I then state the results for the more general setup above.

4.1 Main Example

Consider a simple specification where uni-lateral investment and mutual initial investments are (almost) perfect substitutes. This example will highlight the main results that hold more generally and makes it easy to illustrate some of the intuitive applications. Let the game be:

$$\begin{array}{ccc} \theta > 0 & \text{In} & \text{Out} \\ \text{In} & \theta_1, \theta_2 & \gamma\theta_1, \gamma\theta_2 \\ \text{Out} & \gamma\theta_1, \gamma\theta_2 & 0, 0 \end{array} \qquad \begin{array}{ccc} \text{else} & \text{In} & \text{Out} \\ \text{In} & \theta_1, \theta_2 & -c, \ f(\theta_2) \\ \text{Out} & f(\theta_1), -c & 0, 0 \end{array} \tag{6}$$

where $\gamma \to 1$, and for simplicity π_i is given by the uniform distribution on $[1, -1]$. Here mutual investment is an (almost) perfect substitute of one-sided investment if both types are positive. This captures the implicit sequential idea, that if types are positive they reciprocate investment. A one-sided investment with a negative type, however, leads to a loss of $c > 0$ for the investing player and a payment of $f(\theta_i)$ for the non-investing negative player, which itself may be positive or boundedly negative. The following examples describe some applications.

♢ **At the Bar**. Two people are sitting at a bar. Each has a private value of how much she herself would enjoy a match with the other party. Each player can decide to make a move (in) or not (out). If both make a move, a match is formed. If both stay out, each gets zero. If only player i makes a move, player $-i$ accepts if she has a positive value for the match, and rejects if she has a negative value for the match. If she accepts, a match is formed. If she rejects, no match is formed, but the proposer incurs a cost. This cost c can be thought of as the cost associated with shame, embarrassment or simply the cost associated with making a futile move. Although the setup describes a sequential-move game, since in the second stage there is a dominant strategy, it is equivalent to the simultaneous game presented above.

♠ **Partnership in Trade.** Partners need to invest into relationship specific assets to maximize benefits from trade, Williamson (1979). While each party can benefit from mutual investment, the return on one-sided investment depends on the type of one's partner.[10] If one's partner is opportunistic (negative), he does not reciprocate investment, and unilateral investment leads to a cost c to the investor. This leads to the classic hold-up problem: if parties are opportunistic, even though investment is valuable, it is never reciprocated; anticipating this, partners do not invest - e.g., Grossman and Hart (1988). In contrast, if one's partner is reciprocal (positive), one-sided investment is always reciprocated leading to a benefit for both players. Note that since an opportunistic partner may benefit from her opponent's investment, i.e., $f > 0$, parties cannot reduce the risk of social investment by pre-pay communication. A positive type cannot credibly communicate her type because a negative type always benefits from pretending to be positive.

♣**Dissent** A member of an organization (commercial or political) does or does not agree with a norm or an organizational practice. She can decide to *voice* her concern and deviate from the norm (in), or stay silent and act loyal to the norm (out).[11] Neighbors or workers might express their disapproval of the 'ruler' in front of each other, or not conform to social norms such as homophobia or segregation. Similarly, they might decide to openly disaprove of an existing business practice within the firm. Dissent in front of a member of the organization who also dislikes the existing practice or norm leads to a coalition or a friendship, i.e. a benefit relative to staying silent. At the same time, if one's critical attitude is expressed to a member who supports the prevailing norm, this leads to a loss relative to staying silent.

[10] As mentioned, by assuming that the benefit of mutual investment over the outside option is $\theta_i + b > 0$ for all θ_i, leaves the analysis unchanged.

[11] Such a choice is often argued to be key in maintaininig efficiency in economic and political organizations as well as markets, e.g., Hirschman (1970), Banerjee and Somanathan (2000).

The speaker might be punished, ostracized, persecuted, leading to a loss of c. If f is positive for the negative type, the negative player may receive a reward for reporting the positive opponent.

4.2 Projection Equilibrium

Let's turn to the predictions of a public IPE. The next proposition shows that the unique equilibrium has a cut-off structure and that information projection decreases people's willingness to invest. Furthermore, it implies that types come to underestimate the probability that their opponent has matching preferences. Below, the operator E^{ρ} refers to the expected inference of a ρ biased player given the true joint distribution of type-dependent actions given in equilibrium σ^{ρ}.

Proposition 4 *For any ρ, there is a unique public $\rho - IPE$. This is given by symmetric cutoffs,*

$$\theta_i^{*,\rho} = \sqrt{c/(1-\rho)} \text{ for all } i$$

Furthermore, in any $\rho - IPE$,

I. $E^{\rho}[\theta_{-i} \mid a_{-i}, \theta_i > 0]$ is decreasing in ρ.

II. $E_{\sigma^{\rho}}[E^{\rho}[\theta_{-i} \mid \theta_i > 0]]$ is decreasing in c iff $\rho > 0$.

III. $E_{\sigma^{\rho}}[E^{\rho}[\theta_{-i} \mid \theta_i]]$ is decreasing in θ_i, iff $\rho > 0$.[12]

Let me first describe the intuition for the result on actions. Consider the dating example. By projecting information, an interested Judith exaggerates the extent to which her opponent Paul knows that she is interested. As a consequence, Judith under-estimates the perceived risk that Paul faces when contemplating a move, and exaggerates the probability that Paul will enter if interested. Since Judith still does not know Paul's type, and because a reciprocated entry is almost as beneficial as mutual initial entry, it becomes relatively more important for Judith to stay out. This way Judith reduces the risk of being shamed if Paul were to reject her. Since the game is perfectly symmetric, the same argument holds for Paul. In the limit, both player stay out even if interested, but both expect the other player to move if interested.

Consider now the implications to inference, that is the way players update their beliefs about their opponent. Such inference determines trust and hence it is key in guiding future interactions.

I. Underestimation If Judith is interested she underestimates Paul's interest both if Paul enters and if Paul stays out. This is true because by projecting information Judith believes that Paul's action should reflect Paul's preferences more than it actually does. In particular, Judith believes that Paul will enter using a lower average cut-off than what he actually uses. Hence when seeing Paul enter, Judith

[12] All qualitative results continue to hold under private IPE. In the unique private $\rho-$ IPE the cutoff is given by:

$$\theta_i^{*,\rho} = \sqrt{c}/(1-\rho) \text{ for all } i$$

and all of the three inferential results described hold as well.

too often thinks that Paul entered only because he knew Judith would not reject it. When seeing Paul stay out, Judith is too convinced that Paul is not interested. Both of these effects are increasing in the degree of information projection ρ.

II. Misattribution of Risk The second inference result claims that positive types misattribute the payoff risk associated with social investment to antagonistic preferences leading to underestimate not only conditionally, but on average as well. By under-estimating how much risk her opponent faces, a positive type exaggerates the extent to which her opponent should invest if she is positive and not invest if she is negative. As a result, an interested party over-infers from the event that her opponent does not invest and under-infers from the event that her opponent does invest. Since the decision to invest is positively correlated with valuation, this leads to average underestimation.

The fact that biased players make differential attributions about their own and their partner's reason of staying out leads to an observable non-Bayesian comparative static result. In the Bayesian case, given the martingale property of beliefs, a change in c should have no effect on how a positive type perceives others on average. In contrast, for any $\rho > 0$ an increase in the payoff risk associated with investing with a negative type leads to a decrease in the expected ex-post beliefs about the interest, trustworthiness or matching objectives of others.

III. False Antagonism Finally, the same way as positive types underestimate their opponents on average, negative types overestimate their opponents on average. Intuitively, if Judith is not interested, she will exaggerate the probability that Paul will adjust her actions to Judith's preferences and stay out even if he is interested. Hence Judith over-infers from the event that Paul enters, and under-infers from the event that Paul stays out. Learning under information projection thus induces a false *negative correlation* between one's own type and the perceived type of the opponent. Hence through interacting with others, each player will exaggerate the probability that others have the opposite preference compared to theirs.[13]

4.3 Dynamics

The probability that investment happens is decreasing in the payoff risk c. In many applications, it is then natural to consider a setting where the above interaction repeats over time with a changing c; either until a match is formed, or until some exogenous deadline is reached.

A change in c over time could correspond to (i) a change in how formal the setting is where players interact, a wrong move is very costly in a formal environment; (ii) the extent to which legal institutions and enforceability contracts can substitute for trust decreasing the cost associated with being held up; (iii) a weakening of the disciplinary or societal sanctions against dissent.

Let the dynamic interaction be characterized by a strictly decreasing sequence of risks $\underline{c} = \{c_t\}_{t=1}^T$ such that $c_1 < 1$, and T being finite. As will be clear, the fact that the sequence is decreasing is without

[13] The above results hold with even greater force in the case of private projection. This is true because here a biased type believes that with probability $1 - \rho$ her opponent enters whenever $\theta_{-i} \geq \sqrt{c}$ while she herself only enters if her type $\theta_i \geq \sqrt{c}/(1-\rho)$.

loss of generality. For simplicity, I focus on myopic interaction where changes in c_t is unanticipated at any given time s.[14] Nevertheless, at each t players fully recall the history of their past interactions. This means that in each round t the players' beliefs are inherited from round $t-1$ and play a $\rho-$IPE of the stage game given these beliefs.

In this context, the psychologically natural assumption is that a player publicly projects some information at the beginning of each new encounter. That is to say, at the beginning of each period t, each party will believe that there is probability ρ that her valuation has leaked to her opponent even if it did not in the previous rounds. To describe the dynamic implications of the model, let $\Pr^{\rho}(M \mid \underline{c})$ be the true ex-ante probability that entry happens - a match is formed - by the end of the sequence, provided that both players are positive and play according to a $\rho-$IPE in each round.

Corollary 2 *Suppose $\rho = 0$. For any $\underline{c}$, entry happens with positive probability in each round and matching is efficient:* $\Pr^0(M \mid \underline{c}) = 1 - c_T$.

In the Bayesian case, matching is efficient, and as the payoff risk of social investment vanishes, all positive matches are formed. In contrast, the next proposition shows that given *any* positive degree of information projection, the reverse might be true. Even as c vanishes, no matches are formed. Furthermore, an extreme form of false uniqueness follows: even if all types are positive, they all come to conclude that everyone else is negative. Let $q^{\rho}_{\underline{c}}$ be the limiting probability that any positive type attaches to her opponent being positive by the end of the sequence $\underline{c}$.

Corollary 3 *For any $\rho > 0$ and $\tau > 0$ there exists $\underline{c}$ such that $c_T = \tau$, but* $\Pr^{\rho}(M \mid \underline{c}) < \tau$ *and* $q^{\rho}_{\underline{c}} \leq \tau$.

The logic of the above result rests on finding a cost-sequence that decreases the payoff risk associated with investment sufficiently slowly. While in the Bayesian case more and more matches form, the underestimation and under-entry properties of IPE imply that nobody enters, because they become more and more sceptical about the type of their opponents. The logic again follows from the differential attribution of non-entry to self and to others.[15]

Note, using Corollary 2, it also follows that if for a given τ the result holds for a cost sequence $\underline{c}$, then it will hold a fortiori for any cost sequence that dominates this cost sequence but has the same final element. This equilibrium with such extreme false belief is also confirming in the sense that players beliefs about their opponent is never explicitly contradicted. Let me now describe some implications of the above results and link it to findings and mechanisms described in the psychological

[14] Hence players do not withhold entry for the reasons of utilising an option value, nor do they enter for reasons of experimentation. I conjecture that one can show, however, that the qualitative result below holds also when people take into the dynamic option value of entry.

[15] Note that the predictions explore here differ from the predictions of rational herding models in common value environments in many ways. Most importantly perhaps, while here people might act identically they always correctly understand the extent to which this is consistent with receiving different signal realizations. Here a norm change is perfectly predictable on average. Furthermore, acting identically here is a signal of similar preferences as opposed to contrarian preferences as in this model.

and sociological literature. In doing so I describe the comparative static predictions with respect to the cost of mis-coordination and the true preference uncertainty faced by the players.

4.4 ♣ Dissent and Norms

Let me first describe the consequences of the above results to social or organizational dissent: the practice of voice whereby members of a community or organization can engage in the costly process of communicating their preferences against an existing practice or norm. Although the interactions above describe bilateral situations, it can be applied to such bilateral interactions taking place between all members of a community. Below, I first describe further potential examples, then some evidence that supports the results.

Norm Falsification The above mechanism makes the following predictions. If there is a cost risk associated with expressing dissent with existing norms - such as homophobia or segregation - or an organizational practice - such as a product innovation - people will come to exaggerate the public support for that norm or practice because they misattribute the *lack of voice* by others to their *loyalty.* This will be true even if the cost of expressing dissent vanishes over time, implying that even if none supports the norm, everyone comes to believe that everyone else supports it and none deviates and none sees even the hope that others would ever want to deviate.[16]

Disciplinary Organizations The results may also matter in understanding organizations that use disciplinary methods against dissent, i.e., organizations where expressing a preference for a change is costly for the individual unless it is met with approval. Proposition 4 implies that the greater is the payoff risk - due to institutionalized terror, humiliation, risk of dismissal - the more members will come to believe that there is genuine support for the practice. While everyone will be surprised that none speaks up they will conclude that the reason others do not speak up is because they support the status-quo.

Proposition 6 also makes predictions on how to effectively maintain obedience. A sophisticated political ruler or managerial leader who understands the above mechanism and attempts to induce discipline, can do so effectively by making sure that the cost of expressing dissent is eliminated sufficiently gradually. Initially high terror is necessary to suppress potential resistance. Due to mistaken inference, however, this strong terror can then gradually be replaced by weak terror, which requires much fewer resources, and still no resistance will occur, even if all are against the ruler.

Silent Revolutions The results of Proposition 6 allows for a comparative static with respect to the sequence $\{c\}$. This result implies a predictable discontinuity in beliefs: a sufficiently large drop of c will lead to an unexpected increase in the fraction of people who enter. To see this consider the case where

[16] For example, graduating students face a choice between early family or early career. If the norm is to start with career after graduation, even if most people disagree with this norm, they would not express this if they fear that everyone else has the reverse preference. If the desire to conform to the majority preference is a sufficiently strong force, then this mechanism can greatly distort group identities. Similarly, in public expressions of attitudes of homophobia, racial segregation, or political correctness. Opponents of these norms, will over-conform exaggerating genuine support for the status-quo.

c drops to zero from one round to the next. While over the course of their interactions people become very skeptical whether anyone else is against the norm, q^t is arbitrarily close to 0, as the loss associated with investment with the wrong type actually becomes zero, all players express a preference against the norm to the group's greatest surprise. Generalizing this argument, since positive types underestimate their opponents following non-entry, in a dynamic context they under-estimate the likelihood of entry that will follow given either a *sufficiently large* drop in c from period t to $t+1$, or a drop of c to a *sufficiently low* level.[17]

Kuran (1995) argues that revolutions and major social changes are unpredicted and come as a surprise. For example, a year after the collapse of the Berlin Wall, former citizens of the German Democratic Republic were surveyed if they expected such a change, and, despite the benefit of hindsight, 76% of respondents indicated that they were totally surprised.[18] Tocqueville (1856) argued that the French revolution came as a major surprise to the monarchy which might explain why the monarch, Louis XVI, was slow to respond with economic reforms. Furthermore, consistent with the predictions of the model, what sealed the fate of the monarchy was the impetus that anti-monarchy forces got from districts where peasants enjoyed newly increased freedoms.[19]

Evidence for **Pluralistic Ignorance** The predictions of the model match an phenomenon discussed in social psychology under the rubric *pluralistic ignorance.* Prentice (2007) describes pluralistic ignorance as "the phenomenon that occurs when people erroneously infer that they feel differently from their peers, even though they are behaving similarly." The results described help understand the strategic forces that might lead to this effect.[20]

In an illustrative study, Miller and McFarland (1987) asked students to evaluate their understanding of a very difficult text. In one condition (the unconstrained condition), students were explicitly given the opportunity that in case they needed help, they could choose to stand up in front of their group-mates, leave the room and ask for clarification from the experimenter in a nearby office. In the language of the model they could choose to 'enter', where such entry could be presumably very embarrassing in the eye of someone who understood the text, but not so in the eye of someone who did not. In the other condition (the constrained condition), students were explicitly told they could not seek clarification, i.e., there was no option to enter. It was confirmed that asking for clarification in the first condition was embarrassing in case others did understand the text.

[17] The players' prediction at a beginning of round t after they have observed that $c_t < c_{t-1}$ is affected by two forces. First they exaggerate the likelihood of entry given their beliefs about the opponent's type - due to static information projection. Second they underestimate the likelihood of entry due to their accumulated pessimism due to dynamic information projection. An increase in the size of the drop from c_{t-1} to c_t leaves the force of the first effect unchanged but increases the force of the second effect.

[18] Kuran (1995) offers similar stylized facts in the context of (i) the spread of affirmative action, (ii) the perseverance of the caste system in India, (iii) the end of slavery in the US.

[19] There are alternative accounts that claim that the French revolution was the most predictable event in history ever.

[20] The description of this phenomenon dates back at least to Hans Christian Andersen's famous tale of the "Emperor's Cloth". Early accounts include Tocqueville (1856). See also Elster (2007).

Note that given the fact that the task was very difficult, on the border of being incomprehensible, most types were positive, i.e., would have benefitted from clarification. The results of the experiment, consistent with the above prediction of under-estimation, showed that while no one left the room in the unconstrained condition, students' evaluation of their own understanding relative to that of the others was significantly *lower* in the unconstrained than in the constrained condition. The same was true for their predicted future relative performance involving comprehension of the text. Furthermore, subjects rated themselves *lower* than they rated the average other group member, both on comprehension and on predicted future performance, in the unconstrained condition, but not in the constrained condition.

Related evidence on how people exaggerate the public support for norms comes inter alia from Prentice and Miller (1993) who showed using Princeton undergraduates as a sample, that people greatly exaggerated the extent to which others were comfortable with the existing drinking norms on campus. Subjects rated the average comfort of others, including the average comfort of their friends as much higher than their own and hence that of reality. Furthermore, not only did students overestimate how much others were comfortable, but how universal such a support was. Similar effects were in the context of racial segregation where white males greatly exaggerated how much other white males supported racial segregation, O'Gorman (1979). In the context of complying with norms of political correctness van Boven (2003) provides suggestive evidence.

4.5 ♠ Trade: Mistrust

The above results imply that people misattribute uncertainty regarding whether their opponents are trustworthy or not, into antagonistic preferences. For example, even if all sellers and buyers are reciprocal or interested in a long relationship, information projection causes them to come to believe that others are opportunistic leading to no investment even if the payoff risk associated with investment becomes small.

The mechanism leading to this result may help explain why not only currently existing legal institutions, but the *legacy* of old institutions matter for trust and the quality of economic activity - an effect that might be harder to explain using repeated game comparative statics. To see this, consider a comparative static on the path through which c converges to some fixed c_T. Here, as mentioned before, c can be interpreted as the degree to which parties remain vulnerable to defecting opponents, due to a lack of contractual security in the environment. An improvement in legal institutions that improve the quality and the enforceability of contracts, decreases the risk of investing with a "wrong" type.

Corollary 4 *Consider two converging sequences c and c' s.t. $c_T = c'_T$, but $c_t \geq c'_t$ for all t. It follows that $E_{\sigma^\rho}[E^\rho_T[\theta_{-i} \mid \theta_i > 0, c]] \leq \; E_{\sigma^\rho}[E^\rho_T[\theta_{-i} \mid \theta_i > 0, c']]$.*

The above corollary implies that even if institutions related to the enforceability of contracts improve over time, when comparing two otherwise identical communities, one that got here from an initially safer environment to one that got here from an initially riskier environment, members of the former community will exhibit more distrust towards each other. In the former community people will genuinely more often believe that others are just opportunistic types.

In the economic literature the question of long-term legacy costs determining the performance of institutions has been a crucial one, Acemoglu, Johnson and Robinson (2001). In particular, AJR (2001) argue that the initial institutions set up by Europeans in their Colonies had a long-lasting impact on economic performance. The above result then implies that a source of such variation might be based on the fact that in areas where enforceability of contracts was initially high, people acquired genuinely more optimistic opinions about the trustworthiness or reciprocal nature of others compared to areas where Europeans set up extractive institutions without much legal safeguards or a concern for a rule of law.

Algan and Cahuc (2010), using data on first- and second-generation migrants to the US, provide empirical support on how trust affects economic performance consistent with the mechanism described here. They first show strong evidence for intergenerational transmission of distrust, providing support for an inherent belief-based account of trust, they then document that such trust has a strong explanatory power over time-varying differences in the economic performance of a country.

4.6 $\diamondsuit$ Matching: Segregation

Finally, one can perform comparative statics with respect to the uncertainty that players face about each other's type. Specifically, consider the probability with which a person can identify ex-ante whether her opponent has a positive or a negative valuation. Note that iff this probability is 1, then there is no asymmetric information and information projection has no bite. Here beliefs are fully correct on average and no false antagonism arises. In contrast, false antagonism is maximal when this probability is 0 and smoothly decreases as this probability increases. Formally,

Corollary 5 *Let α be the commonly known ex-ante probability that player i knows ex-ante the type of player $-i$. For any $\rho > 0$, $E_{\sigma^\rho}[E^\rho[\theta_{-i} \mid \theta_i > 0, \alpha]]$ is increasing in α.*

The mechanism described implies that greater uncertainty about the preferences leads to an increase in false antagonism. To illustrate, consider two groups B and W. Suppose members of the same group can read each other's type ex-ante with a higher probability. It follows that people will come to believe that members of their own group are more likely to have matching objectives than members of the other group. Trustworthy members of group B will find more trustworthy people in group B than in group W. Similarly, trustworthy members of group W will find more trustworthy people in group W than in group B.

Importantly, such involuntary segregation has not only different welfare implications, but also different observable implications than standard theories of segregation. In particular,

1. The extent of segregation decreases when initial payoff risk associated with matching with the wrong type is decreased - in contrast to genuine preference-based explanation.

2. The extent of segregation potentially increases the more people interact with members of the other group - in contrast to statistical accounts where segregation is the result of ex-ante uncertainty as opposed to ex-post inference.

Evidence In the context of inter-racial friendships, Shelton and Richeson (2005) showed that both White and Black students at Princeton and the U. Mass desired having more inter-racial friendships - the same was not true for same race friendships. Yet students attributed their lack of initiative to the fear of rejection and the lack of initiative by members of the other racial group to lack of interest. In addition, such differential attribution was true for non-prejudiced Whites, who wanted more interracial friendships, but not to prejudiced Whites, who did not want interracial friendships.

4.7 Linear Investment Games

Let me conclude by returning to the more general case. Below I characterize the equilibrium for a class of monotone linear games. To present the results, a formal distinction between complement and substitute investments is needed. I call investments *complements* if given positive types, than the return on investing with another investing type is increasing in type. In contrast, investments are *substitutes* if this return is decreasing in type.

Definition 3 *Investments are complements if $\theta_i - f(\theta_i) - g(\theta_i, \theta_{-i})$ is increasing in θ_i for all $\theta > 0$, and substitute if $\theta_i - f(\theta_i) - g(\theta_i, \theta_{-i})$ is decreasing in θ_i for all $\theta > 0$.*

For clarity the analysis below focuses on a linear specification, where the marginal utility of one-sided investment, not only of mutual investment, is independent of type. This is a natural specification give the description of type. Hence I impose that for all $\theta_i > 0$, it follows that $g_{11} = 0$, $f_{11} = 0$ and $g_2 = 0$. The next proposition summarizes the results on behavior.

Proposition 5 *For any $\rho > 0$, all equilibria are given by cut-off strategies,*

(i) If investments are substitutes, there is a unique symmetric equilibrium with cut-off $\theta^{*,\rho}$ increasing in ρ.

(ii) If investments are complements, there are at most two equilibria and both cut-offs $\theta_l^{*,\rho}$ and $\theta_h^{*,\rho}$ are decreasing in ρ.

(iii) (false antagonism) in all equilibria $E[E^{\rho}[\theta_{-i} \mid \theta_i]]$ is decreasing in θ_i iff $\rho > 0$.

(iv) (underestimation) in all equiibria $E^{\rho}[\theta_{-i} \mid a_{-i}, \theta_i > 0]$ is decreasing in ρ for all $a_{-i} \in A_{-i}$

The logic of the above result follows from the earlier examples. Since the imaginary "informed" player knows her opponent's type, she will enter if and only if both players value social investment. If actions are complements, this implies that the perceived return on entry is exaggerated, because a positive type exaggerates the probability that her opponent enters. As a result, information projection leads to over-entry relative to the Bayesian case in supermodular games. If investments are substitutes, then the perceived return on entry is under-estimated for the same reason, and information projection leads to under-entry in submodular games.

False Antagonism All projection equilibria will exhibit *false antagonism*, and the source of such false antagonism is that people will exaggerate the extent to which others will act in accordance with their preferences. A positive type exaggerates the ex-ante probability of investment by her opponent and over-infers from the event that her opponent stays out. A negative type exaggerates the ex-ante probability her opponent staying out, and over-infers from entry.

Undervaluation of Social Assets. Finally, note that since all positive types - types that would ever potentially invest - come to under-estimate the type of their opponent both if the opponent enters or if the opponent stays out, it follows that as a consequence of information projection people will come to underestimate the return on investing on social assets. If a match is formed with their opponent they will be too skeptical how much their opponent values the match. If no match is formed, they will be too skeptical about the expected value of future investment opportunities.

5 Persuasion

In this Section, I study a simple sender-receiver problem with costly state-verification. A financial adviser makes a recommendation to an investor whether to buy or sell an asset. I show that in a setting with commonly known conflict of interests between the parties, public information projection predicts credulity: receivers believe good news too much and sophisticated senders take advantage of such credulity leading to exaggerated average posteriors. While Bayesian persuasion (i) cannot shift average beliefs and (ii) improves welfare, persuasion under information projection can (i) inflate average beliefs and (ii) reduce the welfare.

Simple comparative static predictions show how small changes in market size can push communication from inducing correct average beliefs and investment from to inducing highly exaggerated average beliefs and investments. Finally, while in the Bayesian case communication is always weakly beneficial, under information projection, financial advisers can uniformly lower welfare. The model implies novel normative conclusions on eliminating belief-bubbles and welfare-reducing communication.

Before turning to the model, note that the problem of credulity in financial advice and persuasion in general is a widely recognized issue in economics. For example, Malmendier and Shanthikumar (2007, 2009) provide evidence that small investors take positive recommendations too literally and fail to sufficiently discount the extent to which these are inflated. In the context of political persuasion, DellaVigna & Kaplan (2007) provide evidence that sheer access to Fox News made people support conservative statements more strongly. For a review of the evidence see Della Vigna and Gentzkow (2010).

5.1 Setup

Timing A privately informed sender (rater) provides advice to a receiver (investor) whether a statement (asset) is true (good) $\{\theta = 1\}$, or false (bad) $\{\theta = 0\}$. Only the sender knows θ. Upon receiving the advice, the receiver can verify the message at some cost c. If she verifies the message, she learns θ. If she does not, she learns nothing. Finally, the receiver takes an action y. For simplicity I assume that the prior on this state to be symmetric.[21]

Heterogeneity The cost of verification c is the receiver's private information. It is distributed ex-ante according to a positive density $f(c)$ over $[0, \infty)$ leading to a cdf of $F(c)$. Heterogeneity in costs may reflect differences in different receivers' financial expertise, or their differential access to additional sources of information or background information affecting the cost at which they can process the information provided.

Investment Upon hearing the sender's recommendation, the receiver takes an action $y \in [0, 1]$. This action could correspond to an amount of investment made, or the maximum willingness to pay. To keep the analysis transparent, I assume that in optimum this action is equal to the receiver's posterior that the asset has a high return, $\theta = 1$. This is captured by the standard assumption that the receiver's utility function is given by

$$u_r(y, \theta) = -(y - \theta)^2 \tag{7}$$

This means that absent advice the optimal investment equals the prior, i.e., 1/2.

Conflicts of Interest The sender's interest differs from that of the receiver. Conflicts of interest are such that the sender gets a kickback of B - potentially from the issuer of the asset to be introduced later - whenever he issues a good report claiming that the state is $\theta = 1$. At the same time, if following the issuance of a positive report if the receiver decides to check and finds out the the rater lied, the rater incurs a reputational loss of S. In the analysis below, without loss of generality I normalize $S = 1$. This means that B is always interpreted in proportional terms relative to S. I first analyze the case where the rater's incentives are fixed. In Section 4.3, I then endogenize the incentives that the seller of the asset provides to the rating agency. One application of this setting is then one where a financial advisor provides advice to investors or mutual fund managers in confidence and the advisor is paid by the seller of the asset.

5.2 Bayesian Persuasion

Consider first the BNE. This equilibrium has a simple structure and is described as follows: the sender tells the truth if the state is good, and lies with probability p if the state is bad. The receiver checks a 'good' message if her cost is below a threshold, and does not check if it is above this threshold. All receivers believe a negative report claiming that the asset is bad. Equilibrium is maintained by the fact

[21] No result depends on this symmetry assumption, and all extend to all priors.

that the overall probability with which a receiver checks has to make the sender indifferent between lying and telling the truth when the state is bad.

Proposition 6 *Suppose $\rho = 0$. The receiver checks iff $c \leq c^*$ and the sender lies with probability p^*. Furthermore, $c^*(F, B)$ and $p^*(F, B)$ are increasing in F - in the sense of fosd - and in B . Communication is always neutral, $E[y_c^*] = \frac{1}{2}$.*

Neutrality. The above result makes the following straightforward claims. First, equilibrium is given by a cut-off structure. Second, an upward shift in the cost distribution or in the conflict leads to a greater probability of lying, and thus to less information transmitted. Importantly, communication on average is neutral: each receiver's type equilibrium belief follows a martingale. This means that the ex-ante expected investment (belief) is the same as the expected ex-post distribution of investment (beliefs). Persuasion, as is *always* the case under Bayesian assumptions given the martingale property of Bayesian beliefs, does not shift *average* beliefs.

5.3 Biased Persuasion

Consider now communication between an unbiased sender ($\rho_S = 0$) and a biased receiver ($\rho_R = \rho$). The key feature of communication here is that persuasion is no longer neutral on average. Rather information projection leads to two kinds of mistakes: *credulity* by some types and *disbelief* by some other types. The former means that persuasion causes expected posteriors to exceed the prior. Here the receiver over-infers from a high recommendation and is too optimistic about the state on average relative to the truth. The latter means that persuasion causes expected posteriors to be lower than the prior. Here the receiver under-infers from a high recommendation and is too pessimistic about the state on average relative to the truth.

Under credulity, belief updating forms a sub-martingale process and expected ex-post investment is higher than the ex-ante expected investment, i.e., $E[y_c^*] > \frac{1}{2}$. Under disbelief, belief updating forms a super-martingale process, and the expected ex-post investment is lower than the ex-ante expected investment, i.e., $E[y_c^*] < \frac{1}{2}$. Both contradict the defining feature of Bayesian updating: it always being a martingale process.

The next proposition describes the unique (public) $\rho-$IPE of the game. It is characterized by a structure where low types always check, medium types mix between checking and not checking, high types never check. As in the unbiased case, checking follows only a high recommendation. Given that the rater has an incentive to exaggerate the truth, low recommendations are always believed.

Proposition 7 *There exist $c_1^\rho < c_2^\rho < c_3^\rho$ such that*

(i) if $c < c_1^\rho$, the receiver always checks and has correct average beliefs.

(ii) if $c \in [c_1^\rho, c_2^\rho]$, the receiver mixes, is credulous and overinvests

(iii) if $c \in [c_2^\rho, c_3^\rho]$, the receiver mixes, is in disbelief and underinvests

(iv) if $c > c_3^\rho$, the receiver does not check and is in disbelief.

(v) finally, c_1^ρ is decreasing and c_3^ρ is increasing in ρ.

Note that in an IPE each receiver type exaggerates the extent to which the sender knows her type, i.e., her actual cost of verification c. This means that each type believes that with probability ρ the seller's incentive to lie is personalized to the receiver; high cost types think that the sender lies more often than he actually does and low cost types think that the sender lies less often than he actually does. Intuitively, a high cost type believes that the sender knows that she is unable to check and will thus lie to her more often, while a low cost type believes that the sender knows that she could check cheaply if she wanted to and hence the sender will lie to her much less. In short the projected fully informed sender will use a receiver cost specific strategy where the probability of lying is increasing in the receiver's cost.

A key feature of the above unique equilibrium that some receiver types now must mix. While the sender lies to each type with the exact same probability, different types will thus have different beliefs about this probability. As a consequence, they will also have different posteriors upon hearing 'good' news. The above characterization implies that type c_1^ρ underestimates the probability with which the sender lies, and c_3^ρ overestimates this probability. By continuity of the equilibrium construction, it then follows that there exists c_2^ρ such that types below who mix will be credulous and all types above will be in disbelief. [22]

In particular, the lowest types, still always check so they always learn the truth irrespective of projection. Medium types think that with probability ρ the sender's lying probability is such as to keep them indifferent between checking and not checking and assign probability $1 - \rho$ to the sender's true lying frequency. Since the higher is the type, the greater is the lying probability needed to keep this type indifferent between checking and not checking, here the posterior after good news will decrease in type. Medium low types, however, now mix and upon not checking a high recommendation they are too optimistic because they underestimate the probability with which the sender lies to them. As a consequence they overinvest. Finally, high types think that with probability ρ the sender always lie to them, knowing that they never check, and also assign probability $1 - \rho$ to the true lying frequency. Hence they under-invest. [23]

[22] The basic results on credulity and disbelief immediately carry over to the case of private IPE. Checking behavior here, however, is non-monotone and is always given by a pure strategy. Lowes types always check, medium low types, those below c^*, do not check. Medium high types above c^* check, high types do not check. Hence medium low types are credulous, and high types are in disbelief.

[23] The basic results on credulity and disbelief immediately carry over to the case of private IPE. Checking behavior here, however, is non-monotone and is always given by a pure strategy. Lowes types always check, medium low types, those below c^*, do not check. Medium high types above c^* check, high types do not check. Hence medium low types are credulous, and high types are in disbelief.

5.4 Credulity

Part (v) of the above proposition implies some key comparative statics. These are best illustrated by considering first what happens as the degree of projection increases. As a result of such a change, the set of types who always check decreases, i.e., c_1^ρ decreases. At the same time, the set of types who never check decreases, i.e., c_3^ρ increases.

To see the logic, note first that these are functions of the equilibrium probability that the sender lies. Suppose in contrast that after an increase in the degree of projection c_3^ρ decreased. This must mean that the probability of lying by the real sender must have also decreased since type c_3^ρ always thinks that the projected sender always lies. The analogous logic implies that c_1^ρ must also decrease. Now, however, the total incentive for the real sender to lie have increased, which leads to a contradiction. Hence c_1^ρ must decrease and c_3^ρ increase in ρ.

The above fact has a key consequence: if the bias is sufficiently high all types become at least weakly credulous and a positive measure of types become strictly credulous. Such types believe high recommendations too much and hence overinvest in the asset. I refer to this case as *uniform credulity.* In the unique $\rho - IPE$ all receiver types will have exaggerated average posteriors. Such uniform credulity is not a limit result: it holds anytime the degree of projection is greater than the minimally necessary to induce strict lying by the sender.

Proposition 8 1. *There exists $\rho^*(B, F) < 1$ such that if $\rho > \rho^*(B, F)$, then $E_\theta[y_c^{*,\rho}] \geq \frac{1}{2}$ for all c, with strict inequality for a positive measure of types (uniform credulity). Furthermore, $\rho^*(F, B)$ is decreasing in B and decreasing in F in the sense of first-order stochastic dominance.*

2. *For any $\rho > 0$, there exists $\overline{B}(\rho)$, decreasing in ρ with $\lim_{\rho \to 1} \overline{B}(\rho) = 0$, such that if $B \geq \overline{B}(\rho)$, uniform credulity follows.*

The logic of the above result relies on the assumption that there are always receiver types for whom it is never rationalizable to check. It then follows from part (v) of Proposition (7) that there always exists a $\rho < 1$ such that c_3^ρ is as high as the highest type for whom it is ever rationalizable to check. For this type to be indifferent between checking or not, it must thus be the case that the real sender always lies. This implies that all types below c_3^ρ believe a positive message too much while all types above c_3^ρ have correct beliefs. Hence all types in $c \in (c_1^\rho, c_3^\rho)$ are strictly credulous.

The above result has some important observable comparative static consequences. An increase in the complexity of the asset, F, or an increase in the degree of the conflict B, substitutes for a higher degree of projection. A sufficient increase in any of these observables brings about uniform credulity. Let me now turn to these predictions in more detail.

- Consider first the comparative static on the conflict of interests B. In the Bayesian case a change in B should have no effect on average beliefs. Under projection, however, there always exists $\overline{B}(\rho)$

such that for any $B \geq \overline{B}(\rho)$ uniform credulity follows for any ρ. Clearly, if $B > 1$, such that it is a dominant strategy for the sender to lie, then projection has no implications. Similarly, when $B = 0$, it is a dominant strategy for the seller to always tell the truth and again projection does not matter. For conflicts of interests in between these extremes, however, some types will always be credulous and if the conflict is greater than $\overline{B}(\rho)$ then all types will be at least weakly too optimistic.

- Let me now turn to a comparative static on the distribution of checking costs F. Again in the Bayesian case, the cost of processing information does not affect the effectiveness of persuasion on the average. Under projection, an increase in F has a potentially non-monotonic effect on average expected beliefs. If no expertise is required to evaluate the recommended product, F is concentrated on 0, then again information projection has no implications. As the product becomes more complex, in that it is more costly to check the sender's recommendation, the probability of such overly optimistic average beliefs increases, however. If F is sufficiently high, however, such that virtually no types would ever check, then the ρ IPE induces correct beliefs: the sender always lies and almost all receivers expect this. Here a downward shift in F can *lower* investors welfare. Under a lower F, investor types now have a biased reason to be credulous. They can now believe that the advisor, if she knew their type, would not want to lie to them. A further shift in F where all types become full experts then again eliminates credulity because it eliminates much of the need for financial advice. Hence it is the presence of a mass of types with some, but not full financial education (MBA types) who drive the market for financial advice towards credulity in this model.

The above result still does not fully describe the welfare consequences of financial advice, as it takes the conflict of interest between the advisor and the investors to be exogenous. This then only allows for a partial welfare analysis holding such conflicts constant. Let me now turn to an analysis that endogenizes the conflict of interest between the parties and allows for a cleaner and more general welfare analysis.

5.5 Endogenous Conflict and Belief Bubbles

In the analysis above, the conflict of interest B was exogenous. Given the results above, let me now endogenize this benefit by invoking the owner of the asset. As it is typically the case, owners of the asset pay the rater (sender) or continue business with a financial advisor conditional on positive recommendations. As before, I maintain the assumption that this link between the owner of the asset and the rater is fully transparent. In other words, the owner of the asset - the seller - ex-ante offers to pay B to the sender (rater) for sending a high message and this is common knowledge amongst all players. The question I ask then concerns the optimal level of conflict B that the seller of the asset wants to offer to the sender. In other words, what is the optimal transfer that the seller, who is unbiased and hence understands the behavior of the sender and the receiver in equilibrium, would want to offer.

Suppose the seller's profit is simply given by the sale of the asset minus the bonus it pays

$$R(\rho, B) - B = \gamma E_{c,\theta}[y^*] - B \tag{8}$$

where γ is the extent of the market or the marginal benefit - profitability/markup - of demand for the asset.[24] Recall that $E_{c,\theta}[y^*]$ simply reflects the expected aggregate demand for the asset in the population, which is none else here than the ex-ante expected posterior of investors.

Consider first the Bayesian case. Here by virtue of the martingale property of Bayesian communication, the seller's expected revenue is *independent* of the size of the bonus. This is true since receivers correctly account for the bias in incentives when evaluating the rater's advice and hence persuasion is neutral on average. This then implies that if offering a bonus is costly, then in optimum the bonus will be zero.

Lemma 1 *Suppose $\rho = 0$, then $R(0, B)$ is constant in B. The optimal B is 0 and ratings are fully revealing.*

The above fact then implies that under fully Bayesian information processing, persuasion is neutral on average and leads to full revelation of the state. Hence equilibrium is fully efficient.

Let's now turn to the biased case. The proposition below shows that if the extent of the market or the markup is sufficiently high, then the seller always wants to bribe the rater because such bribing always induces excess revenue that is greater than the cost of the bribe. Furthermore, the optimal bribe is always bounded from above by $\overline{B}(\rho)$, implementing uniform credulity, and the seller's optimal profit is always bounded from below by the profit that can be achieved by adopting $\overline{B}(\rho)$.

Proposition 9 *For any $\rho > 0$, if $\gamma \geq \overline{\gamma}(\rho)$, then $\overline{B}(\rho) \geq B^*(\rho) > 0$ and $E_{c,\theta}[y^*] > 0.5$. If $\gamma \leq \underline{\gamma}(\rho)$ then $B^*(\rho) = 0$ and $E_{c,\theta}[y^*] = 0.5$.*

Under the assumption of a constant marginal benefit of additional demand, if the marginal benefit is sufficiently high, the seller's optimal choice is to set a bribe that is strictly positive and is at most as high as is needed to implement uniform credulity. Any bribe higher than that will not generate any excess demand. Furthermore, a key feature of the optimum here is that it always involves overinvestment in the asset on average.

The above result allows for further comparative statics on how aggregate beliefs about the quality of an asset change in equilibrium as the market becomes bigger - greater participation in financial markets -or more complex - financial innovation.

Belief bubbles and market size The Proposition above allows for a limited comparative static result on γ. As the market size reaches a critical value, $\overline{\gamma}(\rho)$, the seller surely pays the rater and there is a potentially *discontinuous* increase in the confidence about the quality of and the demand for

[24] This parameter γ is measured in relative terms given the sender's cost of a loss of reputation S. Thus the lower is the reputational loss, the higher is the absolute value of γ.

the assets. Similarly, after a decrease in the profitability parameter below $\underline{\gamma}(\rho)$ the seller withdraws payment from the rater, and these optimistic beliefs will burst, and investors' average expectations will become realistic again. Importantly, such bubbles are accompanied by observable changes in the transfer between sellers of the asset and the raters of the asset. This in principle allows the external unbiased observer to detect whether average expectations are too optimistic or not.

Belief bubbles and complexity A similar comparative static prediction holds when there are changes in the complexity of the underlying asset An increase in the complexity of the asset makes it cheaper to induce uniform credulity. Hence as expertise about the asset is accumulated, the seller might no longer find this optimal, leading to a *discontinuous* decrease in the aggregate beliefs about the asset's quality.

5.5.1 Welfare: Caps versus Disclosure

The analysis above has implications on how contracting between the asset owner and the asset rater may affect aggregate beliefs and investors' welfare. If the extent of the market is big enough and/or the evaluation of the asset is complex enough, the seller will choose a contract that induces systematic average credulity. This link between contracts and average beliefs about the environment is a unique prediction of this model.

In a simple model, Inderst and Ottaviani (2012) consider the role of naïveté in evaluating financial advice - where naive people act as if there were no conflicts of interest.[25] The solution in such models, and also the one proposed by the authors is mandatory disclosure: highlighting the conflict and thereby eliminate naïveté. In contrast, in my model the conflict is always common knowledge, but credulity persists despite this fact. Disclosure of the conflict is ineffective.

While disclosure is ineffective, capping the bonus has non-trivial effects. It can improve the quality of communication, but more surprisingly it can reduce the average optimism about the quality of the asset and push aggregate beliefs more towards realism. While a complete cap will always restore full efficiency, a more limited cap will have mixed effects. While it will reduce credulity in certain parts of the population - those with medium financial expertise - it will increase disbelief in the part of the population with low level of financial sophistication. Under the conditions of Section 4.3 a more stringent cap on he bonus will, however, always reduce the aggregate demand for the asset.

6 Projection Equilibrium

The model above focused on information projection. A logical counter-part of information projection is *ignorance projection:* the wrong belief that if one cannot condition her strategy on an event than others cannot condition their strategy on that event either. Direct evidence for such ignorance projection is much more sparse than that for information projection and is likely to be a weaker force than

[25] For related analysis for exogenously invoked naive types who always take recommendations at face value, see Kartik, Ottaviani and Squintani (2007).

that of information projection. Nevertheless, the model introduced allows one to incorporate not only information projection, but also the joint presence of information and ignorance projection. I refer to the resulting solution as (public) *projection equilibrium.*

Specifically, one can consider general informational projection whereby a person projects both her information and her ignorance. This can simply be done, by assuming that player i believes that the projected version of player $-i$ conditions his strategy on P_i as opposed to P_+. In words, here a player who exhibits general projection of degree ρ believes that her opponent conditions his strategy on exactly the same information as she does. By making such a substitution all other aspects of the definition are maintained. To state this formally, let the strategy set of the fictional version of player $-i$ - who exists as real only in the imagination of player $i-$ be given by:

$$S^i_{-i} = \{\sigma^i_{-i}(a_{-i} \mid \omega) \in \Delta A_{-i} \text{ measurable w.r. to } P_i\} \tag{9}$$

We can then state the analogous definition of a projection equilibrium given the same logic as before:

Definition 4 *A strategy profile $\sigma^\rho \in S_i \times S_{-i}$ is a ρ projection equilibrium of Γ if there exists $\sigma^\pm \in S_i^{-i} \times S^i_{-i}$ such that for all i,*

1.

$$\sigma_i^\rho \in BR_{S_i}((1-\rho)\sigma^\rho_{-i} \circ \rho\sigma^i_{-i})$$

2.

$$\sigma^i_{-i} \in BR_{S^i_{-i}}(\sigma^\rho_i)$$

Note that projection is again all-encompassing: the fictional opponent of Judith knows Judith's strategy. In other words, Judith believes that with probability ρ Paul knows what she knows and also understands that Judith is real (regular). The existence of this solution follows from the same logic as that of information projection equilibrium and the equivalent versions of Proposition 2 and Corollary 1 continue to hold.

6.1 Common -Value Trade

Let me briefly apply projection equilibrium to the classic common value trade problems, Akerlof (1970), as classically studied experimentally by Samuleson and Bazerman (1985). Here a seller wants to sell an item of quality $q \in \mathbb{R}$ to the buyer. The seller's valuation is q, the buyer's is $w(q) > q$. Quality q is distributed ex-ante according to a continuous density π. The realization of q is the seller's private information. The experimental evidence comes predominantly from the bargaining protocol where the buyer makes a take-it-or-leave-it offer (TIOLI) that the seller can accept or reject. Let me thus turn to the predictions of projection equilibrium to this setting.

Projection by the uninformed buyer implies that the buyer exaggerates the probability with which the seller is also uninformed. I denote the uninformed buyer's equilibrium pricing strategy by p^ρ_b. Since the seller always has a dominant strategy, his strategy is projection-proof. Note that this is true not only for the real seller who is informed, but also for the fictional version of the seller who is as uninformed as the buyer. It then follows that a ρ-biased buyer's perceived expected utility when making an offer of price p_b depends whether it exceeds the estimated value of the object to the seller or not. This is true because in equilibrium the fictional version of the seller will accept the buyer's offer if it exceeds this value and rejects it otherwise. The buyer's perceived expected utility given the seller's strategy and an offer of price p_b is given by:

$$E^{\rho}U(p_b) = \begin{array}{c} (1-\rho)\Pr(q \leq p_b)(E[w(q) \mid q \leq p_b] - p_b) \text{ if } p_b < E_{\pi}[q] \\ (1-\rho)\Pr(p_b \geq q)(E[w(q) \mid q \leq p_b] - p_b) + \rho(E_{\pi}[w(q)]) - p_b) \text{ if } p_b > E_{\pi}[q] \end{array} \quad (10)$$

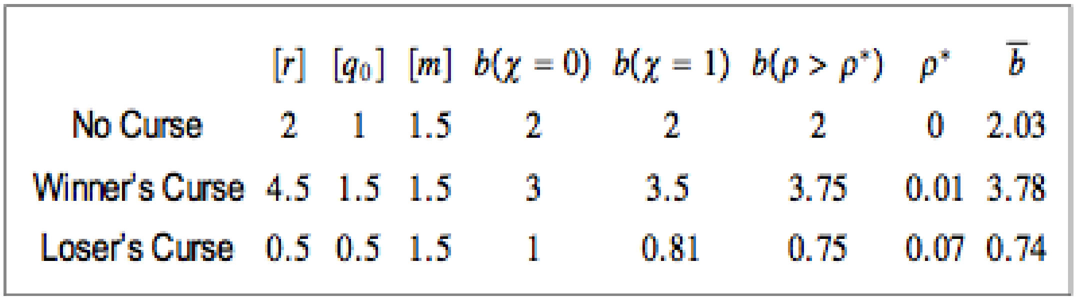
where $E_{\pi}[q]$ is the ex-ante expected quality of the object. It is easy to see, that the solution to this problem is generically unique and always such that if the bid is below $E_{\pi}[q]$, than it is the same as the BNE.

6.1.1 Multiplicative Lemons Problem - Holt and Sherman (1994)

Holt and Sherman (1994) consider a multiplicative lemons problem where $w(q) = mq$ with $m > 1$ with π uniform on $[q_0, q_0 + r]$. The projection equilibrium is discontinuous in ρ. If $\rho < \rho^*$, it is the same as the BNE. If $\rho > \rho^*$, it is instead $p_b^{\rho} = E_{\pi}[q]$. Hence, relative to the Bayesian prediction, it is sufficient to represent projection equilibrium by the calculation of ρ^* - the minimal degree of ignorance projection such that the buyer bids differentially from the Bayesian prediction.

Table 1 below calculates the projection equilibrium in the three conditions studied experimentally by Holt and Sherman (1994) and studied by Eyster and Rabin (2005). In the table below $\overline{b}$ is the average empirical value reported; the BNE corresponds to $b(\chi = 0)$. Cursed equilibrium (CE) spans the interval between this BNE and the fully cursed prediction $b(\chi = 1)$.

In all conditions examined by ER (2005) projection equilibrium provides a closer fit of the data than BNE or CE. Furthermore, the minimal degree of projection required to achieve an almost perfect fit is



	$[r]$	$[q_0]$	$[m]$	$b(\chi = 0)$	$b(\chi = 1)$	$b(\rho > \rho^*)$	ρ^*	$\overline{b}$
No Curse	2	1	1.5	2	2	2	0	2.03
Winner's Curse	4.5	1.5	1.5	3	3.5	3.75	0.01	3.78
Loser's Curse	0.5	0.5	1.5	1	0.81	0.75	0.07	0.74

Holt and Sherman (1994), Eyster and Rabin (2005).

very low. In the winner's curse condition - where people overbid relative to the Bayesian prediction - it is $\rho^* = 1.6\%$. In the loser's curse condition - where people underbid relative to the Bayesian prediction - it is $\rho^* = 7\%$. The fact that in the data players bid concentrate on the point predictions of projection equilibrium provides further support.

A variant of the Holt and Sherman (1994) specification is where $q_0 = 0$ and $r = 1$. This bargaining problem is studied experimentally by Ball, Bazerman, and Carroll (1991), who allow for multiple rounds of learning. Here following the above analysis projection equilibrium predicts that if $\rho > \rho^*$ then $b^* = \frac{1}{2}$ and 0 otherwise. The Table below summarizes the results:

r	q_0	m	$b(\chi = 0)$	$b(\chi = 1)$	$b(\rho^*)$	$\bar{b}$
1	0	1.5	0	$\frac{3}{8}$	0.5	0.55

Besides the point predictions and the perceived discontinuity versus smoothness there is a further difference between projection equilibrium and cursed-equilibrium in these settings. This concerns the buyer's equilibrium belief about the probability that her offer will be accepted. A fully cursed buyer - the prediction that in the in this parametric class is closest to the data - will believe that her offer should always be accepted. In contrast, a buyer who projects her ignorance, since she assigns probability $(1 - \rho)$ to the true behavior of the seller, might attach a significant probability to the fact that her offer is rejected. This is particularly so given the low cutoff value for the bias. Furthermore, given biased bidding the probability of rejection relates negatively to the value of the bias ρ. Finally, note while cursed-equilibrum would predict a positive bid even if $m < 1$, projection always predicts a bid of 0 in this case.

6.1.2 Samuelson and Bazerman (1985)

Samuleson and Bazerman (1985) consider an additive problem where $w(q) = q+30$ and π is the uniform on $[0, 100]$. The unique prediction of the model here is given by

$$p_b^\rho = 30 \text{ if } \rho \leq 1/41 \text{ and } p_b^\rho = 50 \text{ if } \rho > 1/41.$$

Their data is described in a more aggregate form. Samuelson and Bazerman find that offers bunch on 50 ,with 70% of bidders bidding between 50 and 80 and significant fraction bidding above 60. The sellers follow their dominant strategy to accept prices above their reservation value. Note that under Bayesian assumptions bidding above 60 leads to a negative earning and hence it is a dominated strategy. In contrast under projection bidding below 80 can be rationalized by some degree of projection as these all lead to positive perceived expected earnings. Fudenberg and Peysakhovich

Samuelson and Bazerman also study the bargaining protocol where the seller makes a single TIOLI offer that the buyer can accept or reject. They show two facts. First, half of the sellers bid a value that is equal to the conditional reservation value of the object to the buyer, that is $w(q)$. Second, the overwhelming majority of the remaining bids are in the interval $[q, w(q)]$. Furthermore, buyers accept such bids almost uniformly. Relative to the buyer's acceptance choice sellers *underbid.* It is easy to see that for a sufficiently high ρ, there exists a fully separating pure ρ projection equilibrium with $p_s(q) = q+30$. No such pure equilibrium exists if $\rho = 0$. A sufficient condition for this is that $\rho \geq 10/13$. Here $(1-\rho)130 \geq 30$, and then by monotonicity it follows that no seller type has the incentive to deviate for $p_s(q)$.

6.2 Multi-Player Extension

The model so far considered only two-player games. The logic introduced can readily be extended to N player games. Let me here focus only on the main model of public projection.[26] A player i again attaches probability $(1-\rho)$ to the fact that her opponent player j plays the strategy that player j truly plays, and probability ρ of player j being the projected version. For simplicity, I present the definition of projection equilibrium. The extension considering only information projection is exactly analogous. The only difference being that the projected version of player j - who is real only in the mind of player i - conditions her strategy on the joint information of players i and j as opposed to the information of player i.

To formally introduce the multi-player extension, let again

$$S^i_j = \{\sigma^i_j(a_j \mid \omega) \in \Delta A_j \text{ measurable w.r. to } P_i\} \tag{11}$$

denote the set of strategies from which the projected version of player j - who is real only in the imagination of player i - chooses from. Note that this fictional version of j is now specific to a given player i. Since the information of player $k \neq i$ may well differ from that of player i, this set depends on the information of the player who projects, i.e., generically S^i_j differs from S^k_j. In other words, the information and the strategy set of the projected version of player j who is real in the mind of player i differs from the information and the strategy set of the projected version of the same player j who is real in the mind of player k. This follows form the fact that player i and k have different information and hence also project different information. Let $S^i_{-i} = \times\prod_{j\neq i} S^i_j$ the strategy set of the fictional opponents of player i. I denote the generic element of this set by σ^i_{-i}.

To complete the description, one needs to describe the beliefs that a projected version of player j - who is real in the imagination of player i - has about his opponents. As before, I assume that projection is all-encompassing. In words, this means that such a player variant has exactly the *same* belief about the strategy of players other than herself as player i does. The above description then corresponds to the following formal statement:

Definition 5 *Consider an N-person game. A strategy profile $\sigma^\rho \in S$ is a projection equilibrium if for*

[26] In the case of private projection such extension is straightforward since deviations from a BNE σ^0 are unco-ordinated. A player i then best-responds to the belief that given any opponent j this opponent plays a strategy with probability ρ that is a best response given the perturbed information of player j to the belief that the opponents of player j play σ^0_{-j}.

all i there exists $\sigma^i_{-i} \in S^i_{-i}$ such that

$$\sigma^\rho_i \in BR_{S_i}(\rho\sigma^i_{-i} \cdot (1-\rho)\sigma^\rho_{-i})$$

where σ^i_{-i} is such that for each $j \neq i$

$$\sigma^i_j \in BR_{S^i_j}(\rho\{\sigma^\rho_i, \{\sigma^i_k\}_{k\neq i,j}\} \cdot (1-\rho)\sigma^\rho_{-j})$$

As before, the case where $\rho = 0$, corresponds to the definition of BNE. The key additional feature of the N-player extension concerns the beliefs that projected version of a player j belonging to player i. Such a version of player j - who is real only in the imagination of player i - *shares* the beliefs of player i about the strategies of all other players. This means that in equilibrium such a player variant not only shares the information of player i, that is, conditions his choice on the same information that the real player i does, but also knows the strategy that player i is truly playing and for each player $k \neq i$, assigns probability $(1-\rho)$ to this player k's actual strategy and probability ρ to player k being the projected version belonging to player i. In other words, projection is again all encompassing for each player in the game.

7 Conclusion

The goal of this paper is to introduce the informational projections into a class of Bayesian games. The paper considered applications of the phenomenon of information projection to strategic settings. The list of course is not exhaustive. Future work can apply the model to a variety of other problems. Strategic settings describing communication, signalling, deception, trade may utilize the insights developed in this paper. This paper considered only static games. Future research can also incorporate information projection into dynamic games. As an example Madarasz (2014b) applies a dynamic version of this model to classic sequential bargaining.

8 Appendix

Proof of Proposition 1. The existence of private information projection or projection equilibrium follows directly from the existence of a BNE. The existence of a public IPE or projection equilibrium follows from Kakutani's theorem. Consider information projection equilibrium. Note that the mapping from real to a perceived strategy profile $\sigma^\rho \to (\sigma^\rho, \sigma^+)$ is always upper hemicontinuous convex and well-defined. Furthermore, all the introduced best-response mappings are upper hemicontinous and convex. It then follows from Kakutani (1941) that a fixed point of the mapping $\sigma^\rho \to \{\sigma^\rho, \sigma^+\} \to \{BR(\sigma^\rho), BR(\sigma^+)\}$ exists. The proof for projection equilibrium is analogous .

Proof of Corollary 1. Note if $P_i = P_j$, then $P^+ = P_i = P_j$. Hence for any σ^0 that is BNE of Γ there exists a σ^+ such that it satisfies the conditions of the definition of private IPE. Hence $\sigma^+ = \sigma^0$, and $\sigma^\rho = \sigma^0$ is a private ρ IPE for any ρ. Similarly, for any σ^0, there exists a $\sigma^\rho = \sigma^+ = \sigma^0$ that satisfies the definition of a public IPE since $BR_{S_j^+} = BR_{S_j}$. By the same toke the reverse direction is also true. The logic immediately extends to projection equilibrium .

Proof of Corollary 2&3. Suppose that σ^0 is a BNE and it is also a strict ex-post equilibrium in Γ. It follows from the no-regret condition, that $\sigma_i^+ = \sigma_i^0 \in BR_{S_i^+}(\sigma_{-i}^0)$ for all i. Hence $\sigma_{-i}^0 \in BR_{S_i}(\rho\sigma_i^+ \circ (1-\rho)\rho\sigma_i^0)$. Suppose that σ^ρ is a private IPE. It then follows that there exists σ^0 that is a BNE of Γ on which σ^ρ is based. If σ^0 is also a strict ex-post equilibrium, then it must be true that for any $\sigma_i^+ \in BR_{S_i^+}(\sigma_{-i}^0)$, $\sigma_i^+ = \sigma_i^0$ hence $\sigma^\rho = \sigma^0$ and this it is also a strict ex-post equilibrium .

Proof of Auction Result. As shown by Riley (1989), the BNE of the first-price auction is given by an efficient mixed-strategy equilibrium where each payoff type mixes over an interval of positive measure such that different payoff types mix over intervals that are non-overlapping. Consider now a σ^+ with the following properties. If the fictional super player $-i$ has a lower valuation than his opponent, $\theta_{-i} < \theta_i$, then he will bid higher than the regular player $-i$ with the same valuation θ_{-i}, if the lowest value of the support over which type θ_i mixes is lower than θ_{-i}. If the fictional super player has a weakly higher payoff type than his opponent, $\theta_{-i} \geq \theta_i$, then he will under-bid, and bid the highest value of the support over which θ_i mixes under $\sigma_i^0(\theta_i)$. Consider now the biased player's best response.

$$b^*(\theta_i) \in \arg\max E_\pi[\rho \Pr(win \mid b, \sigma_{-i}^+(\theta_i)) + (1-\rho)\Pr(win \mid b, \sigma_{-i}^0)](\theta_i - b)$$

Note first that since the auction was efficient under σ^0, the equilibrium probability of winning was zero conditional on the opponent having a higher valuation. In contrast under σ^+ it is positive if the bidder bids above the relevant part of σ^0. At the same time, bidding lower than under σ^0, the probability of winning is lower than in the case where $\rho = 0$. It is thus easy to see that bidding below the lowest value of the support over which this payoff type was mixing under the BNE cannot be an equilibrium. Finally, note that if one's opponent has the same valuation as she, an event that happens with positive probability given the finite support, then given the indifference condition under BNE and the deviation of such an informed type, it is now strictly beneficial to bid above the original bid. The discontinuity in the revenue result arises from the fact that for any $\rho > 0$, a biased player will not bid below the highest point of the interval on which she was supposed to mix under the BNE combined with the fact that all such intervals have positive measure. .

Proof of Zero-Sum Games. The derivation of the private $\rho - IPE$ follows directly from algebra.

Specifically, it is given by:

$p<\frac{1}{2}$	weak	strong	EU_D	$p>\frac{1}{2}$	weak	strong	EU_D
$\rho=0$	B	$\frac{1}{2-2p}\text{A}\circ\frac{1-2p}{2-2p}\text{B}$	$\frac{1}{2}$		$\frac{2p-1}{p(2-\omega_w)}\text{A}\circ\frac{1-\omega_w p}{p(2-\omega_w)}\text{B}$	A	$\frac{1-\omega_w p}{2-\omega_w}$
$\rho=1$	A	B	$\frac{1-\omega_w p}{2}$		A	B	$\frac{1-\omega_w}{2-\omega_w}$
	$\frac{1}{2}a\circ\frac{1}{2}b$	$\frac{1}{2}a\circ\frac{1}{2}b$			$\frac{1}{1+\omega_w}a\circ\frac{\omega_w}{1+\omega_w}b$	$\frac{1}{1+\omega_w}a\circ\frac{\omega_w}{1+\omega_w}b$	

To see the revenue result, note that in case the defender does not have private information, her expected utility (winning probability) is $\frac{1-p\omega_w}{2-p\omega_w}$. To show the result, note that $\frac{1-p\omega_w}{2-p\omega_w}\leq\frac{1}{2},\frac{1-p\omega_w}{2-\omega_w}$. At the same time, $\frac{1-\omega_w p}{2-\omega_w}>\frac{1-\omega_w}{2-\omega_w},\frac{1-\omega_w p}{2}$ where the latter follows from the fact that $1>\omega_w p$.

Consider now the case of public IPE under the symmetric prior. We need to show that there exists a mixed strategy for the fictional attacker that makes a ρ_D biased defender indifferent between using either of his actions. In the case where the defender is strong, this is always true since $\rho_D+(1-\rho_D)q=(1-\rho_D)(1-q)$ implies that $q=\frac{2p_D-1}{2\rho_D-2}\in[0,\frac{1}{2}]$. In the case where the defender is weak, the indifference condition is given by $(1-\rho_D)+\rho_D q=\rho_D(1-q)(1-w)$, which implies that $q=-\frac{1}{2\rho_D-w\rho_D}(-2\rho_D+w\rho_D+1)$ which is feasible if $\rho_D\geq 1/(2-\omega_w)$.

Proof of Proposition 2. Note first that the projected opponent has a dominant strategy entering iff $\min(\theta_i,\theta_{-i})\geq 0$. I first show that the equilibrium is in cut-off strategies. Let p_{-i} be the probability that real player $-i$ enters given strategy σ^{ρ}_{-i} and the prior distribution of θ_{-i}. For any given type θ_i, the utility differentials from entering versus staying out is

$$E^{\rho}[u_i(in)]-E^{\rho}[u_i(out)]= \tag{12}$$
$$=\rho[(\textstyle\int_0^{\theta_{\max}}d\theta_{-i})(\theta_i-f(\theta_i))+(\int_{\theta_{\min}}^0 g(\theta_i,\theta_{-i})d\theta_{-i})]+(1-\rho)[p_{-i}(\theta_i-f(\theta_i))+(1-p_{-i})E_{\pi}[g$$

Since only positive types can be indifferent, from $f'<1$ and $g_1>0$, it follows that the difference is strictly increasing in θ_i for any given σ_{-i}. Hence the equilibrium must be a cut-off strategy given by some $(\theta_i^{*,\rho},\theta_{-i}^{*,\rho})$.

1. We can thus consider the best-response function $BR(\theta_{-i})$ which determines the solution $E^{\rho}[u_i(in)]-E^{\rho}[u_i(out)]=0$ for a fixed θ_{-i} describing the cut-off of player $-i$. Note that the best-response functions for the two-players are perfectly symmetric. Using the implicit function theorem we obtain that the slope of this best-response function is given by

   $$\frac{d\theta_i}{d\theta_{-i}}\mid_{(\theta_i^*,\theta_{-i}^*)}=\frac{(1-\rho)(\theta_i-f(\theta_i)-g(\theta_i,\theta_{-i}))}{\rho[(\Pr(\theta_{-i}>0)(1-f'(\theta_i))+\int_{\theta_{\min}}^0 g_1(\theta_i,\theta_{-i})d\theta_{-i}]+(1-\rho)[p^*_{-i}(1-f'(\theta_i))+\int_{\theta_{\min}}^{\theta_{-i}^{*,\rho}}g_1(\theta_i,\theta_{-i})d\theta_{-}}$$

   It follows from the assumptions, that the denominator is always positive. If investments are substitutes, the numerator is negative, hence $BR(\theta_{-i})$ is downward sloping. If investments are complements, the numerator is positive, and hence $BR(\theta_{-i})$ is upward sloping. Furthermore, given a linear investment problem, for all $\rho < 1$

$$sign\{\frac{d^2\theta_i}{d^2\theta_{-i}}\} = sign\{\overbrace{(1 - f'(\theta_i) - g_1(\theta_i, \theta_{-i}))}^{I}\overbrace{(N(d\theta_i/d\theta_{-i}) + (\theta_i - f(\theta_i) - g(\theta_i, \theta_{-i})))}^{II}\} > 0.$$

where N is a positive number. Since the sign of term I is always the same as the sign of term II, both are positive if investments are complements, and negative if they are substitutes, as long as both players enter with positive probability, this second-derivative is always positive in the relevant domain.

2. Consider substitute investments. A symmetric equilibrium must exists due to the fact that $BR(\theta_{-i})$ and $BR(\theta_i)$ are downward-sloping and they are each-other's mirror image on the 45 degree line. Furthermore, given that $BR(\theta_{-i})$ is strictly decreasing and $sign\{\frac{d^2\theta_i}{d^2\theta_{-i}}\} > 0$, it follows that the symmetric equilibrium is unique.

3. Consider complement investments. It is easy to see that all equilibria must be symmetric since $BR(\theta_{-i})$ is a strictly increasing, hence it cannot be the case that for a given $(\theta_i^{*,\rho}, \theta_{-i}^{*,\rho})$, $BR(\theta_{-i}^{*,\rho}) > BR(\theta_{-i}^{*,\rho})$ and $\theta_{-i}^{*,\rho} < \theta_i^{*,\rho}$. Since $BR(\theta_{-i})$ is increasing and convex, it follows that $BR(\theta_{-i})$ and $BR(\theta_i)$ can cross at most two points in the relevant domain of where $\theta \geq 0$.

4. Let's consider the comparative static with respect to ρ. We can re-write the equilibrium cut-off condition for θ_i as

$$\rho[\overbrace{\int_0^{\theta_{-i}^{*,\rho}} (\theta_i - f(\theta_i) - g(\theta_i, \theta_{-i}))d\theta_{-i}]}^{III} + [\left(\int_{\theta_{-i}^{*,\rho}}^{\theta_{\max}} d\theta_i\right)(\theta_i - f(\theta_i)) + \int_{\theta_{\min}}^{\theta_{-i}^{*,\rho}} g(\theta_i, \theta_{-i})d\theta_{-i}] = 0$$

   Consider substitute investments. Here term III is negative. Hence the LHS is decreasing in ρ. Furthermore, given that $g_1 > 0$, the LHS of the equation is increasing in θ_i, since $\theta_{\min} < 0$ and $\rho < 1$. This means that holding $\theta_{-i}^{*,\rho}$ constant, the solution in terms of θ_i is increasing in θ_i. Hence an increase in ρ shifts the best-response functions upwards, and the unique and symmetric equilibrium cutoff increases. Consider complement investments. Here term III is positive. Hence the LHS is increasing in ρ. Furthermore, as before, the LHS of the equation is increasing in θ_i .This means that the cut-offs are decreasing in ρ.

5. **Underestimation**. Any real positive type of player i perceives the cut-off of the real opponent to be higher than the cut-off of the fictional informed opponent in equilibrium. Hence player i's posterior condition on entry of her opponent is a probability weighted average of the type distribution truncated on $[\theta_{-i}^{*,\rho}, \theta_{\max}]$ and on $[0, \theta_{\max}]$. Since $\theta_{-i}^{*,\rho} \geq 0$, this implies underestimation of θ_{-i} relative to the truth following entry of player $-i$. Similarly, player i's posterior conditional on non-entry by her opponent is a probability weighted average of the type distribution truncated

on $[\theta_{\min,}\theta^{*,\rho}_{-i}]$ and on $[\theta_{\min,},0]$, which for the same reason implies underestimation of θ_{-i} relative to the truth following exit of player $-i$.

6. **False antagonism**. Consider a positive type $\theta_i > 0$. This type always exaggerates the probability of entry by player $-i$, hence she under-infers from entry and over-infers from exit. Since the conditional posterior following entry is higher than following exit, this type underestimates her opponent's type on average. Formally, let $p^\rho(a_{-i} = I \mid \theta_i > 0, \sigma^\rho_{-i})$ be the perceived probability of type θ_i that player $-i$ will enter given σ^ρ_{-i}. Correspondingly let $p(a_{-i} = I \mid \theta_i > 0, \sigma^\rho_{-i})$ be the true probability. It must be true that in equilibrium that the the perceived expected posterior equals the prior:

$$p^\rho(a_{-i} = in \mid \theta_i > 0)[E^\rho[\theta_{-i} \mid a_{-i} = in, \theta_i > 0]] + (1 - p^\rho(a_{-i} = I \mid \theta_i > 0)[E^\rho[\theta_{-i} \mid a_{-i} = out, \theta_i > 0]] = E_\pi[\theta_{-i}].$$

Since $p^\rho(a_{-i} = in \mid \theta_i > 0) < p(a_{-i} = in \mid \theta_i > 0, \sigma^\rho_{-i})$, it follows that $E_\pi[E^\rho[(\theta_{-i} \mid \theta_i > 0)]] < E_\pi[\theta_{-i}]$. The case for $\theta_i < 0$ is analogous, since here $p^\rho(a_{-i} = in \mid \theta_i < 0) > p(a_{-i} = in \mid \theta_i < 0, \sigma^\rho_{-i})$.

.

Proof. Suppose $\rho = 0$. Note first that beliefs in round t are left-truncation of beliefs of period $t-1$, given the monotonicity of the equilibrium. Let q_t denote the probability that player i assigns to player $-i$ having a positive type, conditional on no entry up to period t. This will correspond to a cut-off type θ_t, such that the opponent's type is uniformly distributed $[\theta_t, -1]$. This means that symmetry is preserved in each round and by the above argument equilibrium remains symmetric and unique. Player i's entry decision, in period t is

$$(\frac{\theta^*_{t-1}}{\theta^*_{t-1} + 1})\theta^*_i - (1 - (\frac{\theta^*_{t-1}}{\theta^*_{t_{-1}} + 1}))c_t = \gamma z_{-i,t}\theta^*_i$$

where $z_{-i,t}$ is the probability that the opponent enters in period t. Solving these equations and taking $\gamma \to 1$, we get that

$$\theta^*_t = \frac{\sqrt{\frac{1}{1+\theta^*_{t-1}}}}{\sqrt{\frac{\theta^*_{t-1}}{1+\theta_{t-1}}}}\sqrt{\theta^*_{t-1}}\sqrt{c_t} = \sqrt{c_t}$$

and thus history is irrelevant for setting the strategic-cutoff in the Bayesian case. Furthermore it is easy to see that efficiency by period t is given by $1 - c_t$ again independent of the details of the history. It also follows that efficiency is bounded below by $1 - \varepsilon$, if $\lim c_s \leq \varepsilon$. .

Proof. Consider now the biased case. Note that in period t again there is some probability that player i assigns to player $-i$ having a positive type: $q_{t,-i}$. The indifference condition for player i is given by

$$q_{t,-i}\theta^*_{i,t} - (1 - q_{t,-i})c_t = \rho q_{t,-i}\theta^*_{i,t}\gamma + (1 - \rho)z^*_{-i,t}\gamma\theta^*_{i,t}$$

Solving for this condition and taking $\gamma \to 1$, we first get that

$$\theta^*_{i,t} = c_t \frac{(1-q_{t,-i})c_t}{q_{t,-i} - q_{t,-i}\rho - z^*_{-i,t}(1-\rho)}$$

and the estimate that $z^*_{-i,t} = q_{t,-i}\frac{\theta^*_{-i,t-1}-\theta^*_{-i,t}}{\theta^*_{-i,t-1}}$. Note that by symmetry of the equilibrium we have that $\theta^*_{i,t} = \theta^*_{-i,t}$ and $q_{t,-i} = q_{t,i} = q_t$

Information projection implies that $q_t < \frac{\theta^*_{t-1}}{1+\theta^*_{t-1}}$ because each player attaches some probability to her opponent having been informed in the previous round. Hence we have

$$\theta^*_{i,t} = \frac{\sqrt{(1-q_t)}}{\sqrt{q_t}}\sqrt{\theta^*_{t-1}}\sqrt{\frac{c_t}{1-\rho}} > \sqrt{\frac{c_t}{1-\rho}} \text{ for all } t > 1$$

and thus entry at round t following a history is lower than if c_t was characteristic of the first interaction.

Consider now a sequence $\{c_s\}$ such that $c_s \geq (1-\rho)$ for all $s < N(\rho)$. It follows from the above that in a $\rho - IPE$ there will be no entry up to period $N(\rho)$. Consider the inference process. Let q_t be the probability assigned to the opponent's type being positive. It follows that

$$q_{t+1} = \frac{q_t(1-\rho)}{q_t(1-\rho)+(1-q_t)} < q_t \text{ for all } t < N(\rho) - 1$$

As we have seen before to achieve a matching loss that is bounded by τ in the Bayesian case, we can set $\lim_{s\to\infty} c_s = \tau$. To show the proposition we need to show that there exists $N(\rho)$ such that

$$\theta^*_{i,t} = \frac{\sqrt{\left(1-q_{N(\rho)}\right)}}{\sqrt{q_{N(\rho)}}}\sqrt{1}\sqrt{\frac{\tau}{1-\rho}} = 1-\tau$$

solving this equation we get that $q_{N(\rho)} = \frac{\tau}{1-\tau(1-\tau)-\rho(1-\tau)^2} > \frac{\tau}{1-\tau(1-\tau)} > 0$ for all $\tau > 0$. Since the above q_t sequence converges to 0 as $N(\rho)$ grows, this proves the corollary. .

Proof. Note that for any p, the BNE is a cut-off strategy since the net benefit of checking is strictly decreasing in c. The point of indifference is always $c = \frac{p}{(1+p)^2}$. This c is then determined by the solution to

$$(1-F(c))B - F(c)(1-B) = 0$$

It follows that $c^{*,0}$ - and as a consequence $p^{*,0}$ - is increasing in B and also in F in the sense of fosd. .

Proof. Note first that if $\rho > 0$, a pure-strategy equilibrium need no longer exist. Suppose there was a cut-off equilibrium where types checked iff $c \leq c^*$, then for sufficiently small $\tau > 0$, $c^* - \tau$ would have a strictly weaker incentive to check than $c^* + \tau$, because $p^+(c^* - \tau) = 0$ and $p^+(c^* + \tau) = 1$. This leads to a contradiction. It follows that $p^+(c) : \mathbb{R}^+ \to [0,1]$ - the probability which which the fictional informed sender lies given c- must smoothly increase on some interval $[c^\rho_1, c^\rho_3]$ and be surjective. Hence for any p^ρ there exists $c \in [c^\rho_1, c^\rho_3]$ such that $c = p^\rho/(1+p^\rho)^2$. Let this be c^ρ_2. It follows that $p^+(c^\rho_2) = p^\rho$. Hence conditional on checking behavior, types in $[c_1^\rho, c_2^\rho]$ are credulous and types above c_2^ρ are in disbelief.

To show that c_3^ρ is increasing in ρ, suppose in contrast that after an initial increase in ρ, $c_3^\rho < c^{*,0}$. Now the sender has strictly more incentive to lie. Hence $p^\rho = 1$. If $p^0 < 1$, however, then this discontinuous increase in p^ρ implies that types above c^* would want to check, a contradiction. Hence c_3^ρ must increase and c_1^ρ must decrease initially. Consider now $\rho' > \rho$. Suppose that $c_3^{\rho'} < c_3^\rho$. This means that $p^{\rho'} < p^\rho \leq 1$ must hold, since $p^+(c_3^\rho) = 1$ for any ρ. Hence $c_1^{\rho'} > c_1^\rho$ must also be true. This leads to a contradiction, however, since $p^+(c_1^\rho) = 0$ for any ρ. Hence c_3^ρ must increase in ρ and hence c_1^ρ must decrease in ρ. Consider now an increase in B it follows that the total mass of types who check ove the types who do not check must increase, i.e. $c_1^\rho/(1-c_3^\rho)$ must increase. Note that if c_1^ρ increases (decreases) this must mean that p^ρ increases (decreases) which implies that c_3^ρ increases (decreases) as well. Hence an increase in B leads to an increase in both c_1^ρ and c_3^ρ. .

Proof. Note that for sufficiently high types it never pays to check. If $\rho = 1$ the real sender always lies for any $B > 0$ and F with full support. Since the set of types for whom it is not rationalizable to check is strictly bounded away from 0, and c_1^ρ is smoothly decreasing and c_3^ρ is smoothly increasing in ρ, there also exist $\rho^* < 1$ such that $p^{\rho *} = 1$ again for any $B > 0$ and F. Here $c_2^{\rho^*} = c_3^{\rho^*}$and thus uniform credulity follows. Furthermore, c_3^ρ is increasing in B and also in F in the sense of fosd for any ρ. Hence the comparative static results follow. Finally, since c_1^ρ and c_3^ρ is increasing in B for any ρ, it follows that there exists $\overline{B}(\rho) < 1$ such that $c_3^\rho = c_{\max}$, here uniform credulity holds. Furthermore, since c_3^ρ is increasing in ρ and c_1^ρ is decreasing in ρ it follows that $\overline{B}(\rho)$ must decrease in ρ. .

Proof. Note that if $B = \overline{B}(\rho)$ then $R(\rho, \overline{B}(\rho)) > 0.5$ hence there exists $\overline{\gamma}(\rho)$ such that $\overline{\gamma}(\rho)[R(\rho, \overline{B}(\rho)) - 0.5] > \overline{B}(\rho)$. Since $R(\rho, 0) = 0.5$ and since $R(\rho, \overline{B}(\rho))$ is non-increasing in B for all $B > \overline{B}(\rho)$, because c_3^ρ is constant and c_1^ρ is increasing in B, the result follows. Since $R'(\rho, 0)$ is bounded because change in c_1^ρ and c_3^ρ is smooth in B, the second part also follows. .

References

[1] Acemoglu, Daron, Simon Johnson and James Robinson. (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation." American Economic Review, 91: 1369-1401.

[2] Akerlof, George. (1970). “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism.” Quarterly Journal of Economics 84: 488-500.

[3] Algan, Yann, and Pierre Cahuc. (2010). "Inherited Trust and Growth." American Economic Review, 100(5): 2060-92.

[4] Arrow, Kenneth. (1972). "Gifts and Exchanges." Philosophy and Public Affairs, 1: 343-362.

[5] Ball, Sheryl, Max Bazerman, and John S. Carroll. (????). “An Evaluation of Learning in the Bilateral Winner’s Curse.” Organizational Behavior and Human Decision Processes, 48:1-22.

[6] Baron J. & Hershey J.C. (1988). "Outcome Bias in Decision Evaluation." Journal of Personality and Social Psychology, 54(4): 569-579.

[7] Biais, Bruno and Martin Weber. (2009). "Hindsight bias and Investment Performance." Management Science, 55: 1018-1029.

[8] Birch, Susan and Paul Bloom. (2007). "The Curse of Knowledge in Reasoning About False Beliefs." Psychological Science, 18(5): 382-386.

[9] Camerer, Colin, Teck-Hua Ho, and Juin Kuan Chong. (2004). “A Cognitive Hierarchy Model of Games.” Quarterly Journal of Economics, 119(3): 861-898.

[10] 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.

[11] Danz, David, Kristof Madarasz, Stephanie Wang (2014). "Do People Anticipate Information Projection: An Experimental Investigation." *mimeo LSE*

[12] Dawes, Robyn and Matthew Mulford (1996). "The False Consensus Effect and Overconfidence: Flaws in Judgment or Flaws in How We Study Judgment?" Organizational Behavior and Human Decision Processes, 65(3)*:* 201–211.

[13] Elster, Jon. (2007). *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences* Cambridge University Press.

[14] Esponda, Ignacio. (2008). "Behavioral Equilibrium in Economies with Adverse Selection." American Economic Review, 98(4): 1269-91.

[15] Eyster, Erik and Matthew Rabin. (2005). "Cursed Equilibrium." Econometrica, Vol. 73, No. 5., 1623-1672.

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

[17] Kagel, John. (1995). "Auctions: a Survey of Experimental Research." in *Handbook of Experimental Economics* ed. J. Kagel and S. Roth, Princeton University Press.

[18] Kakutani, Shizuo. (1941) "A Generalization of Brouwer’s Fixpoint Theorem." Duke Math J. 8, 457-59.

[19] Katz, Daniel and Floyd H. Allport. (1931). *Student Attitudes.* Syracuse, N.Y.: Craftsman.

[20] Kuran, Timur. (1995). *Public Lies and Private Truth*, Harvard University Press.

[21] Gilovich, Thomas, Victoria Medvec, Kenneth Savitsky. (1998). "The Illusion of Transparency: Biased Assessments of Others’ Ability to Read One’s Emotional States." Journal of Personality and Social Psychology.

[22] Grossman, Sanford and Oliver Hart. (1986). "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." Journal of Political Economy, 94(4): 691-719.

[23] Harsanyi, John. (1967- 1968). "Games with incomplete information played by Bayesian players." Management Science, 14: 159-182, 320-334, 486-502.

[24] Hirschman, Albert. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States.* Cambridge, MA: Harvard University Press.

[25] Holt, Charles and Roger Sherman. (1994). "The Loser's Curse." American Economic Review, 84(3): 642-652.

[26] Indherst Roman and Marco Ottaviani. (2012). "How Not to Pay for Financial Advice." Journal of Financial Economics (forthcoming).

[27] Jehiel, Philippe. (2005). "Analogy-based Expectations Equilibrium." Journal of Economic Theory, 123: 81–104.

[28] Jehiel, Philippe and Frederick Koessler. (2008). "Revisiting games of incomplete information with analogy-based expectations." Games and Economic Behavior, 62: 533-557.

[29] La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny. (1997). "Trust in Large Organizations." American Economic Review, 87: 333-38.

[30] 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.

[31] Madarasz, Kristof. (2012). "Information Projection: Model and Applications." Review of Economic Studies, 79: 961–985.

[32] Madarasz, Kristof. (2014): "Bargaining under the Illusion of Transparency." *mimeo LSE.*

[33] Malmendier Ulrike, and Devin Shanthikumar. (2007). "Are small investors naive about incentives?" Journal of Financial Economics: 85(2): 457–89.

[34] Malmendier Ulrike, and Devin Shanthikumar. (2009). "Do security analysts speak in two tongues?" Working Paper.

[35] Miller, Dale and McFarland Cathy. (1987). "Pluralistic ignorance: When similarity is interpreted as dissimilarity." Journal of Personality and Social Psychology, 53(2): 298-305..

[36] Nisbett, R. E., & Ross, L. (1980). "Human inference: Strategies and shortcomings of social judgment." Englewood Cliffs, NJ: Prentice-Hall.

[37] O'Gorman, Hubert. (1975). "Pluralistic Ignorance and White Estimates of White Support for Racial Segregation." Public Opinion Quarterly, 39 (3): 313–30.

[38] Piaget, Jean and Barber Inheldar. (1955). *The Child's Conception of Space.* New York.

[39] Prentice, Deborah and Dale Miller. (1993). "Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm." Journal of Personality and Social Psychology, 64: 243-256.

[40] Prentice, Deborah. (2007). "Pluralistic ignorance." *In Encyclopedia of Social Psychology.* Sage.

[41] Perry, Motty and Phil Reny. (2002). "An Efficient Auction," Econometrica, 70: 1199-1212.

[42] Riley, John. (1989). "Expected Revenue from Open and Sealed Bid Auctions." Journal of Economic Perspectives, 3(3):

[43] Ross, L., Greene, D., & House, P. .(1976). "The "false consensus effect": An egocentric bias in social perception and attribution processes." Journal of Experimental Social Psychology, 13: 279-301.

[44] Rubinstein, Ariel. (1989). "The Electronic Mail Game: Strategic Behavior Under Almost Common Knowledge." American Economic Review, Vol. 79, No. 3. (Jun., 1989), pp. 385-391.

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

[46] Shelton, J. Nicole and Richeson, Jennifer. (2005). "Intergroup Contact and Pluralistic Ignorance." Journal of Personality and Social Psychology, 88(1):

[47] Wimmer, H and Perner, J. (1983). "Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception". Cognition, 13(1): 103–128.

[48] Stahl, Dale and Paul Wilson.(1994). "On Players' Models of Other Players: Theory and Experimental Evidence." Games and Economic Behavior, 10(1): 218-254.