DISCUSSION PAPER SERIES

No. 10636

PROJECTION EQUILIBRIUM: DEFINITION AND APPLICATIONS TO SOCIAL INVESTMENT AND PERSUASION

Kristóf Madarász

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

ISSN 0265-8003

PROJECTION EQUILIBRIUM: DEFINITION AND APPLICATIONS TO SOCIAL INVESTMENT AND PERSUASION

Kristóf Madarász

Discussion Paper No. 10636
May 2015
Submitted 20 May 2015

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801
www.cepr.org

PROJECTION EQUILIBRIUM: DEFINITION AND APPLICATIONS TO SOCIAL INVESTMENT AND PERSUASION†

Abstract

People exaggerate the extent to which their information is shared with others. I introduce such information projections into a large class of Bayesian games where people wrongly think that if they can condition their strategy on an event others can as well. I apply the model to a variety of settings. In the context of social investment, people misattribute the uncertainty others face about their preferences into others having antagonistic preferences. Even if all parties prefer mutual investment, none invests, but comes to believe through interacting with others that she is alone preferring mutual investment. In the context of communication, the model predicts credulity: persuasion by an advisor with a known incentive to lie will nevertheless induce uniformly inflated average posteriors. Complexity of an asset, but greater financial education as well, can enhance such credulity. I extend the model to incorporate ignorance projection and re- late the predictions of projection equilibrium to evidence on common-value trade. Here, consistent with the evidence in Samuelson and Bazerman (1985), the model predicts non-altruistic truth-telling by sellers. For buyers it predicts the winner's curse and provides a better fit of the data than BNE or cursed equilibrium. Further applications to zero-sum games and auctions are explored.

JEL Classification: C7 and D03
Keywords: persuasion belief-bubbles, pluralistic ignorance, projection and social investment

Kristóf Madarász K.P.Madarasz@lse.ac.uk
London School of Economics and CEPR

---

† An earlier version, the first online version, Madarasz (2014a), contains additional material and explores implications to auctions as well. I would like to thank Pedro Bordalo, Colin Camerer, Je¤ Ely, Erik Eyster, Ignacio Esponda, Philippe Jehiel, Marina Halac, Paul Heidhues, Navin Kartik, 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, Harvard, LSE, Stockholm, UCL, UC Los Angeles, UC San Diego, Royal Holloway, CEU, ITAM, Southampton, Berlin Behavioral Seminar, Gerzensee 2013, EBEAM 2013 Berlin. All errors are mine.

1 Introduction

Since information is typically dispersed across agents who then respond to such dispersion strategically; the way that people perceive informational differences is key in many settings. While the typical assumption in economics is that people are well-calibrated about such differences, evidence shows that they systematically underappreciate them. In particular, the typical person too often acts as if others shared her perspective and have access to her private information. Thus such information projection - empathy gaps in informational persepective-taking - has potential implications for strategic behavior in many contexts. For example, it may shape people's perceptions of the nature of strategic conflict that exists between them and others.

The goal of this paper is to incorporate informational projection into a large class of Bayesian games. I offer a simple but fully portable model that corresponds to a single-parameter extension of the standard solution concept. The paper, thus, allows one to explore a wide range of economic consequences of this phenomenon to strategic behavior both theoretically and empirically. I illustrate such consequences by developing applications to problems of social investment, persuasion, bilateral trade, and auctions.

Evidence for information projection comes from a variety of domains and dates back to classic work in psychology on the theory of mind following Piaget and Inhelder (1948).[1] Such robust and widely documented phenomena as the curse of knowledge (Camerer, Loewenstein, and Weber 1989; Newton 1990), the hindsight bias, (Fischhoff 1974; Biais and Weber 2010), the outcome bias (Baron and Hershey 1988), or the illusion of transparency (Gilovich, Medvec, and Savitsky 1998, 2000) are all consistent with the idea that people exaggerate the probability with which others can act upon their private information. Madarász (2012) offers a brief review of the evidence and explores implications for non-strategic inference problems.

In a strategic context, Samuelson and Bazerman (1985) provide evidence from bilateral trade. They show that privately informed sellers act as if uninformed buyers were symmetrically informed naming prices that equal buyers' conditional valuations, and that uninformed buyers act as if sellers were symmetrically uninformed naming prices that equal sellers' unconditional valuations. A companion paper, Danz, Madarász, and Wang (2014), to be described later, contains a direct test of the model presented here including some of its implications for higher-order perceptions, such as the extent to which people anticipate the biases of others.

Model In Section 2 I present the model of information projection equilibrium (IPE). A person who projects information has an exaggerated belief that her opponent uses a strategy

[1] Classic *false beliefs* task demonstrated that children too often acted as if others shared their superior informational perspectives, e.g., Wimmer and Perner (1983). Recently, Brich and Bloom (2007) showed that the same structure of mistakes exists among Yale undergraduates in slightly more complex tasks.

that is conditioned also on her private information. The extent of this false belief is characterized by the parameter $\rho \in [0, 1)$. After briefly presenting a model with a private version of projection, I turn to the main model.[2] To incorporate this phenomenon parsimoniously into games, where higher-order perceptions potentially matter, I distinguish between the regular and the projected (super) versions of a player. In reality, all players are regular, but fictional projected versions enter into players' beliefs about each other. Specifically, regular player i, Paul, conditions his strategy on his true information, while super Paul, who is real only in her opponent Judith's imagination, conditions his strategy on Judith's and Paul's joint information. If Judith projects to degree ρ, she believes that with probability ρ, Paul is a super as opposed to a regular version.

Two properties characterize IPE. First, projection is *all-encompassing*; that is, Judith believes that super Paul knows everything she does. Second, a player's expectation about others is consistent with how others actually behave. If the true game is poker, in which people privately observe only their own cards, Judith believes that, with probability ρ, Paul knows the value of her cards and also that Judith does not know the value of Paul's cards. At the same time, Judith believes that, with probability $1 - \rho$, Paul is regular and does not know her cards but wrongly believes that she knows his cards with probability ρ. In short, each player expects her opponent, with positive probability, to behave the way this opponent always behaves, but each players assigns positive probability to her opponent behaving as though he knew everything she did, thus expecting behavior that might never occur or might occur with a different probability.

The above properties imply that, in equilibrium, people exhibit partial sophistication about the biases of others. While each player anticipates that her opponent projects onto her, proportional to the extent that she projects, she underestimates the extent of this. As a result, and in contrast to the logic of BNE, a person has not only wrong views about her opponent's strategy on average, but also wrong beliefs about her opponent's belief of her strategy, and she believes that her opponent has wrong beliefs thereof and so on. All such higher-order perceptions, however, are solely a function of the degree of projection ρ. After presenting the model, I also illustrate some basic consequences of the model through some simple examples and establish existence. I also show that in all games with symmetric information, IPE is equivalent to BNE and that all BNE that are ex-post equilibria are also projection-proof.

Social Investment In Section 3, I present the main application of the model to the problem of social investment. Partnerships in trade, friendships, cooperation in large organizations, and the formation of social and political associations all require people pooling individually owned resources together. Such investment is risky because people face uncertainty regarding others' motives. Investing with someone who is reciprocal or has matching

[2] Madarász (2014a) contains a more extensive discussion of private projection.

goals is a source of potential gain. Investing with someone who is opportunistic or has opposing goals is a source of loss. Social investment is risky because people face uncertainty regarding the motives or preferences of others. Trust in these settings is associated with a player's belief that her opponent is the former as opposed to the latter type.

Suppose that two players decide independently and simultaneously whether or not to invest. If neither invests, each gets the outside option. If both invest, each realizes his or her private value of mutual investment. If only one party invests, the investing party achieves a benefit relative to the outside option if both players are positive, and a loss otherwise. The model here implies that biased people come to believe that others have preferences that are antagonistic to their own. By projecting information, a person underestimates the uncertainty that her opponent faces about her preferences. In equilibrium, she exaggerates the probability with which her opponent's choice should be aligned with her preferences and comes to wrongly believe that his preferences are antagonistic to hers. A person who values mutual investment positively will do so both conditional on her opponent investing or not investing and on average as well, and thus come to underestimate the return on investing in social assets.

As a corollary, in a setting with repeated myopic encounters, I show that even 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 players value social investment, none invests, but all come to believe that the reason that others do not invest is that they do not value social investment at all. Thus the presence of uncertainty about others' preferences lead them to believe that they are alone in having such preferences. The joint prediction on the resulting behavior and inference is consistent with the classic phenomenon discussed in the psychological and sociological literature under the rubric 'pluralistic ignorance'. It describes a systematic discrepancy between private attitudes and assessed public attitudes and behavior generated by a process whereby "people erroneously infer that they feel differently from their peers, even though they are behaving similarly" (Prentice 2007).[3]

I conclude the Section by exploring comparative static consequences to three different applications: (i) expressing dissent against a prevailing social norm or organizational practice; (ii) initiating or investing into a friendships; (iii) trust in trade.

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

[3]The term originates from Katz and Allport (1931), who describe that students at Syracuse University who generally didn't object to allowing minorities into then-segregated frat houses, but were convinced that contrary to their preferences their peers wouldn't accept such a multicultural move. Krech and Crutchfield (1948) describe pluralistic ignorance as a situation where 'no one believes, but everyone thinks that everyone [else] believes'.See also, e.g., Miller and McFarland (1987), Prentice and Miller (1993).

truth. Embedding this problem into a setting with endogenous conflicts of interests, the model predicts the rise and fall of a belief bubble as the size of the market changes.

A privately informed advisor sends a message to an investor whether a statement is true or false – an asset has high or low expected returns. The advisor's preferences 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 that help verify the advisor's claim.

Bayesian communication has two identifying properties: on average it improves the welfare of receivers, and it is neutral in that the ex-ante expected posterior of the receivers' is equal to their prior. Under projection by receivers when talking to an unbiased (hence sophisticated) sender both of these properties are violated. The model identifies settings in which 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 overinvest in the asset.

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 an IPE, if the asset is sufficiently complex or the misalignment of his preference is sufficiently strong, a sophisticated advisor will always lie more often than any receiver type - except those for whom it is dominant strategy not to check - expects. The greater is the conflict or the more common receiver types who have neither very high costs (clueless) nor very low costs (experts), the more likely uniform credulity will be.

Finally, I endogenize 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 this is no longer true. In particular, if the size of the market is sufficiently large, the seller always pays the advisor and implements credulity, whereas if the size of the market is below this critical threshold, the advisor does not pay the advisor, and average beliefs are truthful. Capping payments can always restore truthful beliefs.

Projection equilibrium Section 5 combines information projection - the positive side of the information gap - with ignorance projection - the negative side of the information gap. In short, here, a player projects both what she knows and what she does not know, exaggerating the probability that her opponent has the information that she has and only that information. I derive implications of the resulting projection equilibrium (PE) to bilateral trade with common values, Akerlof (1970). Consistent with the evidence – Samuelson and Bazerman (1985) – when a seller has all the bargaining power, projection equilibrium predicts (non-altruistic) truth-telling and underbidding relative to the buyer's acceptance behavior. In contrast, when the uninformed buyer has all the bargaining power, projection equilibrium predicts overbidding in situations of the winner's curse, and underbidding in the situation of the loser's curse. I also compare the predictions of the model to evidence from Holt and Sherman (1994) and show that projection equilibrium provides a better fit of the data than BNE or cursed equilibrium in all treatments. Finally, I present the multi-player extensions of the models considered.

2 Model

This section introduces the model and describes some of its key properties. I first briefly present a more naive version (private projection) and then turn to the main definition (public projection). I then illustrate some of the model's implications. For ease of exposition, I restrict attention to two-player games and present the extension to N players in Section 5.

Consider a Bayesian game Γ. Let there be a finite set of states Ω and an associated strictly positive prior π. Each state $\omega \in \Omega$ describes all payoff-relevant facts. Player i's information about ω is given by a standard information partition $P_i : \Omega \to 2^{\Omega}$, her finite action set by A_i, and her bounded payoff by $u_i(a, \omega) : A \times \Omega \to \mathbb{R}$, where $a \in A = \times_i A_i$. In short, the game is summarized by $\Gamma = \{\Omega, \pi, P_i, A_i, u_i\}$.

To introduce information projection, let me express the joint information of the two players i and j. Formally, consider the following correspondence:

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

This correspondence is also partitional and describes the coarsest common refinement of the two players' partitions – i.e., the information distributed between these two players. This means that if an event, $E \subseteq \Omega$, is known at a state ω by either of the players in the game, this event is also known at that state under P^{+}. Conversely, any event that is known under P^{+} is known given the combined information of the two players. The joint information, thus, corresponds to the natural object to capture the idea of information projection here.[4]

A key fact of strategic interactions is that they might depend on higher-order perceptions which must then be specified. To introduce information projection in a parsimonious manner, I distinguish between two versions of each player i. The *regular* version of i conditions her strategy on her true information. Formally, she chooses a strategy from the

[4]Note that for an information partition there is a unique knowledge operator $K : 2^{\Omega} \to 2^{\Omega}$ associated with it. Given a partition $P_i(\omega)$, the corresponding knowledge operator is $K_i(E) = \{\omega \mid P(\omega) \subseteq E\}$, denoting the set of states where E is known. The knowledge operator corresponding to P^{+} is uniquely defined by $K^{+}(E) = \{\omega \mid \cap_i P_i(\omega) \subseteq E\}$.

set

$$S_i = \{\sigma_i(\omega) \mid \sigma_i(\omega) : \Omega \to \Delta A_i \text{ measurable with respect to } P_i\}.$$

The *projected* (super) version conditions her strategy on the joint information of i and j.[5] Formally, she chooses a strategy from the set

$$S_i^+ = \{\sigma_i(\omega) \mid \sigma_i(\omega) : \Omega \to \Delta A_i \text{ measurable with respect to } P^+\}.$$

In reality, all players are regular for sure. Fictional projected versions only enter into people's beliefs about each other.

Information projection by a player j corresponds to a mistaken belief that, with probability ρ, player i is a super as opposed to a regular version. For ease of notation, below, I assume that this degree of projection is common across players but then immediately extend the definition to heterogeneous projection.[6]

Private versus Public I now turn to the definition. The main model and applications are developed under a public notion of projection. This will be introduced in Section 2.1 and adopted throughout the paper thereafter. Before turning to this main model, however, let me digress and briefly introduce a private notion of projection. This private notion will violate both the *consistency* and the *all-encompassing* properties of the main model to be described. For a more extensive discussion of this naive but easy-to-use model, and a comparison with public projection, see Madarász (2014a). The reader interested in the main model can also skip this part and continue from Section 2.1.

Below, σ^ρ describes the actual strategy profile of the players - the strategy profile of the regular versions. The operator $\circ$ denotes the mixture of the two probability-weighted lotteries. The operator BR denotes the standard best-response operator; its subscript refers to the set of strategies and the player who maximizes her expected utility; its argument refers to this player's belief about her opponent's strategy to which she wishes to best-respond.

2.0.1 Private Projection

A private information projection equilibrium (play*)* describes a setting where people initially share a common view about behavior in the game which accords with a true BNE of the

[5] I use the notion of a projected and a super version interchangeably.

[6] Note, that since all payoff relevant information can be encoded ω, the extent that player i has information about her own taste, or the taste of her opponent, she projects that information as well. See Section 5 for evidence on this in the context of common value trade.

In a basic common value environment with positively correlated valuations, information projection may imply an exaggeration of how close the realized valuations are. Given negatively correlated valuations, it may imply an exaggeration of how far these realized valuations are. For example, suppose that player i's valuation is $x \in \mathbb{R}$ and player j's is (i) x or (ii) $-x$. Let the common prior on x be $N(0, 1)$. Suppose that only player i gets a signal, which is given by $s_i = x + \varepsilon_i$ where ε_i is a Gaussian noise. A biased player will then exaggerate the similarity of the players' expected perception of their valuations in case (i), and exaggerate the difference in players' expected perception of valuations in case (ii).

game. Player i then comes to privately believe that, with probability ρ, her opponent can now also use her private information in choosing his strategy. She believes, however, that her opponent does not recognize the fact that she believes this now. Intuitively, Judith thinks that her cards have leaked to Paul with probability ρ, but that super Paul does not realize that she has figured this out. Judith thinks that Paul wrongly believes that she assigns probability 0 to such information sharing (leakage). Hence, projection here is private.

To state the definition, I need to distinguish between strategies played in equilibrium, and strategies that describe players' beliefs about their opponents' behavior. In contrast to BNE, these need not correspond to each other. In particular, each player i believes 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}.

Definition 1 *A strategy profile $\sigma^\rho \in S_i \times S_j$ is a private ρ information projection equilibrium (play) of Γ if there exist strategy profiles $\sigma^0 \in BNE[\Gamma]$ an4d $\sigma^+ \in S^+_i \times S^+_j$ such that for all i,*

1.

$$\sigma^\rho_i \in BR_{S_i}\{(1-\rho)\sigma^0_{-i} \circ \rho\sigma^+_{-i}\}$$

2.

$$\sigma^+_{-i} \in BR_{S^+_{-i}}\{\sigma^0_i\}$$

■ The definition corresponds to a parametric deviation from BNE. If $\rho = 0$, there is no deviation from the suggested Bayesian equilibrium σ^0. If $\rho > 0$, the model deviates from that of BNE. Specifically, given σ^0, describing the ways 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 supposed equilibrium strategy, σ^0_i, by conditioning her action on the joint information in the game – i.e., playing σ^+_{-i}. She assigns the remaining probability to her opponent being regular and acting as before, σ^0_{-i}. Player i's private IPE strategy σ^ρ_i is, then, a best response to such a wrong belief.

■ 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 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^0_{-i} for sure and thinks that player j thinks that player i plays σ^0_i for sure, and so on. This means that the belief that an opponent picks a strategy from S^+_{-i} as opposed to S_{-i} enters only into first-order beliefs about strategies.[7]

■ Finding a private IPE is very simple given a BNE of the game. It involves calculating individual - as opposed to mutual - best-responses. This feature makes this notion particularly easy to apply in settings in which BNE is well understood. It also immediately implies existence whenever BNE exists and the fact that all ex-post equilibria are projection-proof.

2.0.2 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, $\omega = \omega_w \in (0,1)$, 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,

$$\begin{array}{ccc} \text{attacker/defender} & \text{A} & \text{B} \\ \text{a} & \omega, 1-\omega & 1,0 \\ \text{b} & 1,0 & 0,1 \end{array} \tag{2}$$

D-Day To illustrate, consider the landing of the Allies on the shores of Normandy in June 1944. Suppose Allies had a choice between two potential locations Calais (A) and Normandy (B). There was good reason to believe that Calais might be the easier terrain for an attack. Suppose German forces occupying both locations also had some private information whether this was true or not. The historic success of D-day is often attributed to the occupiers firm expectations that an attack would take place at Calais. Suppose that the state is the defender's private information, and that ex-ante, each state is equally likely. The table below summarizes the defender's strategy in the unbiased and the fully biased cases. The attacker 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 Achilles

[7] Here players best respond to a potentially fully wrong theory of their opponent's behavior. This feature, which does not hold in the case of public projection, links private projection to cognitive hierarchy (level-k) models, e.g., 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. These models leave open the specification of what the underlying (level 0) heuristic to the specific application at hand. Here the relevant heuristic is always BNE.

Note also that cognitive hierarchy models differ from private projection play, 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.

heel. Thinking that her type has leaked to the attacker, but that he does not realize that she recognizes this, when she is strong, she expects him to attack at A, when she is weak, she expects him to attack at B. Her best response is, then, to defend A when weak and B when strong. 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, in any ρ-IPE, she defends her Achilles heel for sure.

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

A further non-Bayesian 'chocking' result is easy to demonstrate. For any given prior π, I compare the defender's ex-ante equilibrium winning probability in two cases: (i) the defender is privately informed about ω as above; or (ii) she has no private information, and as the attacker, only knows the true prior. In the Bayesian case, the defender's expected payoff is always greater in case (i). In the fully biased case, the opposite holds.

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.*

2.0.3 Overbidding in IPV Auctions

As a second example, consider a symmetric independent private-value auction problem. Suppose that each player's valuation is distributed according to some π over any finite set of valuations, $v_1 < v_2 < .. < v_N$. The classic Bayesian result in this setting is *revenue 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 in which players overbid 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 exists a private IPE such that the first-price auction generates discretely higher revenue than the second-price auction.*

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, given information projection, the increase in revenue is discrete when moving from the case of $\rho = 0$, to the case where $\rho > 0$.[8]

2.1 Definition

I now turn to the paper's main model in which projection is public. In the remainder of this paper I focus on this model including when I extend it in Section 5. As before, σ^{ρ} describes the actual strategy profile of the players - the strategy profile of the regular versions. It is supported by a profile σ^{+} describing the conjectured behavior of the projected versions.

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

1.
$$\sigma_i^{\rho} \in BR_{S_i}\left\{(1-\rho)\sigma_{-i}^{\rho} \circ \rho\sigma_{-i}^{+}\right\} \tag{3}$$

2.
$$\sigma_{-i}^{+} \in BR_{S_{-i}^{+}}\{\sigma_i^{\rho}\} \tag{4}$$

The solution $\rho-$IPE is a single-parameter extension of Bayesian Nash equilibrium. If $\rho = 0$, each player puts full probability on her opponent's true strategy; hence, the set of IPE coincides with that of BNE for any given Γ. If $\rho > 0$, each player mistakenly assigns probability ρ to her opponent playing a strategy which is also conditioned on her private information *and* is also a best response to her true strategy. I now describe two defining features of the model.

All-encompassing Projection. First, projection is all-encompassing in the sense that the imagined super opponent of player i understands that player i is regular for sure. This is reflected by Eq. (4). Regular Judith believes that super Paul not only knows her private information, but also correctly understands that Judith is regular for sure. If the true game is poker, each player sees only his/her own cards, Judith believes that with probability ρ Paul knows the value of her cards, and, at the same time, also that she does not know Paul's cards. This implies that in equilibrium, Judith thinks that super Paul also knows her exact strategy at each state.

Consistency. Second, each player's expectation of how her opponent plays is consistent - in a limited way - with how her opponent actually plays. This is reflected by Eq.(3). Each regular player assigns probability $1-\rho$ to her opponent behaving in the way this opponent

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

always behaves. The deviation from BNE is simply that Judith expects something to happen with probability ρ that might never happen or happen with a different probability. Thus in equilibrium nothing happens that explicitly contradicts a player's theory of how her opponent may behave.

Partial Sophistication The above two properties imply that the equilibrium is consistent with a player exhibiting *partial* sophistication about her opponent's projection: Judith expects Paul to wrongly think that she knows his cards, but underestimates the likelihood of this. Specifically, Judith assigns probability $1-\rho$ to Paul thinking that she knows his cards with probability ρ, and probability ρ to Paul correctly thinking that she does not know them. In sum, she thinks that he assigns probability $\rho-\rho^2$ to her knowing his cards, *on average.* Hence, in the model, a player not only has wrong beliefs about her opponent's strategy, on average, but also wrong beliefs about her opponent's beliefs about her strategy.

Higher-order Perceptions In a similar fashion to considering player i's belief of player j, Judith, being super (first-order belief) and player j's belief of player i's belief of player j being super (second-order belief), one can construct player i's k^{th} order belief and player j's $k+1^{th}$ order belief thereof. [9] In doing so, the same qualitative features will be present. Judith's $k+1^{th}$ order belief of her being super is smaller than Paul's k^{th} order belief of Judith being super. The discrepancy, however, decreases as k increases and always corresponds to ρ^k. Paul's higher-order beliefs are decreasing and Judith's are increasing, and both converge to $\rho/(1+\rho)$ at infinity.

Heterogeneous Projection As mentioned, the definition extends to differentially biased players. Heterogenous projection is described by a vector $\rho=(\rho_i,\rho_j)$ with ρ_i replacing ρ in Eq. (3) for each i. If $\rho_i=0$, then player i is unbiased and assigns full weight to her opponent's actual strategy, i.e., she is sophisticated and understands that her opponent projects to degree ρ_j. If $\rho_j>0$, she expects her opponent to know her cards with probability ρ_i and expects her opponent to wrongly think that she is super - and knows his cards - with $(1-\rho_i)\rho_j$ on average. Hence, since projection is all-encompassing, the extent to which a player projects is also the extent to which she underestimates the projection of her opponent.

[9] Let me define the higher-order iterative perceptions of the real players about player j being a super version inductively as follows: the first-order iterative belief is the probability that player i assigns to j being super, the second-order is the overall probability that player j assigns to player i assigning to player j being super, etc. Every odd element in this sequence refers to a belief by player i. Every even element refers to a belief by player j. The k^{th} element of this sequence is given by $\sum_{s=1}^{k}(-1)^{s-1}\rho^s$. In this sequence, (i) the sub-sequence of odd elements is decreasing in k, (ii) the sub-sequence of even elements is increasing in k, and (iii) each odd element is larger than the subsequent even element, and each even element is smaller than the subsequent odd element, (iv) both the even and the odd subsequence converges to $\rho/(1+\rho)$. Hence for all higher orders k even, player j underestimates the strength of player i's belief at order $k-1$, but the discrepancy, which is always ρ^k, vanishes as k increases.

Evidence As mentioned, Danz, Madarász and Wang (2014) test the partial sophistication aspect of the model, and find evidence consistent with it.[10] They find that better-informed players project their information onto lesser-informed players, and that on average lesser-informed players anticipate but underestimate the degree of projection of the better-informed players. Furthermore, the extent of the former, ρ, is remarkably consistent with the extent of this underestimation, ρ^2, as predicted by the model.

2.2 Discussion

Having presented the model, let me turn to some of its basic properties. I first show that a ρ information projection equilibrium always exists.

Proposition 1 *For any* Γ *and* ρ *an information projection equilibrium exists.*

This fact 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 IPE=BNE for all* ρ.

Let me now turn to the relation of the model to the notion of ex-post equilibrium where players' strategies satisfy an ex-post no-regret condition. A BNE strategy profile is a strict ex-post equilibrium if no player has a weak incentive to deviate even conditional on the realization of the state. Such an outcome is robust to ex-post deviations once the state is learned. In contrast to IPE, an ex-post equilibrium often does not exist and hence provides no predictions. The following proposition states, however, that when it does exist an ex-post equilibrium is projection-proof.

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

The above proposition implies that in games, such as second-price auctions, where BNE is also an ex-post equilibrium, information projection does not reduce the set of Bayesian predictions.

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.[11] Eyster and Rabin (2005) study behavior under the assumption that a person correctly understands informational differences, but believes

[10] Technically, Danz et al. (2014) estimate projection equilibrium - incorporating ignorance and information projection - as described in Section 5. The structure of higher-order perceptions, and hence partial sophistication there, however, is the same as here.

[11] See also Jehiel and Koessler (2008).

that, with some probability, her opponent plays the identical strategy over all of his information sets. An identifying assumption in both Jehiel (2005) and Eyster and Rabin (2005) is that people, on average, have correct expectations about their opponents' strategy. They have a coarse understanding of other player's strategy, but are correct on average. An information projection equilibrium, and the latter-introduced projection equilibrium, differs. In an IPE players have *wrong* beliefs on average about their opponents' strategies, and, thus violate the identifying restriction of an ABEE or a CE. Furthermore, the economic and psychological logic of information projection differs markedly from cursedness. For example, in all private value environments, such as the social investment problem considered below, cursedness makes the same predictions as BNE. In Section 5, when incorporating ignorance projection into the model, as well, I will return to the link between projection and cursedness.

2.3 Two Examples

Illusory Coordination Let there be two states $\omega \in \{\omega_1, \omega_2\}$. The payoffs and actions are given as follows, where $\varepsilon \in (0,1)$:

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

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

Suppose only the column player knows the state. As a result, in all *pure*-strategy BNE, players must play a constant action across states. In contrast, IPE can exhibit illusory coordination where the informed party adjusts her action fully to the state. Specifically, if ρ is sufficiently high, and ω_R is the ex-ante less likely state, $\{(L,l,\omega_1);(R,l,\omega_2)\}$ is a $\rho - IPE$. To see this, note that if $\rho \to 1$ the informed party can believe that if the state is ω_R, her opponent who knows the state and her strategy will play r in response to her playing R.

Over- and Under-Mixing in Zero-Sum Games Let's return to the hide-and-seek game. The next result shows that IPE again predicts that the defender will play his Achilles heel too often. Given the partial sophistication feature of the model, a biased defender, however, now over-mixes in equilibrium. Suppose again that there is a symmetric prior on the state. Given a sufficiently high ρ, in a ρ-IPE, the privately informed defender mixes in each state as if the state was common knowledge:

defender	weak	strong	EU_D^ρ
$\rho \to 1$	$\frac{1}{2-\omega_w}\text{A} \circ \frac{1-\omega_w}{2-\omega_w}\text{B}$	$\frac{1}{2}\text{A} \circ \frac{1}{2}\text{B}$	$\frac{1}{2} - \frac{1}{4}\frac{\omega_w}{2-\omega_w}$

The defender's expected payoff here decreases in ω_w and converges to $\frac{1}{4}$ when ω_w converges to 1. It is, however, always higher than in the case of private projection. Consider now the case where only the defender is biased and the attacker is unbiased. The following table describes an IPE:

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

An unbiased attacker, can thus *under-mix* relative to the Bayesian case and play b (Normandy) for sure. This example also illustrates that the predictions of IPE are *inconsistent* with the predictions of BNE of any perturbed game Γ' which differs only from Γ in that people have some alternative information partitions P_1' and P_2'. The reason for this is as follows: the fact that the attacker plays b for sure is inconsistent with the belief that his opponent has \correct beliefs about his strategy and best-responds to it. If the defender had correct beliefs about his strategy, it would never be an equilibrium to play b for sure. Instead this relies on the fact that players here believe that their opponents systematically mispredict their choices or fail to best-respond to it.

3 Social investment

Efficient outcomes often require people to pool individual resources 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. Interaction in such settings often takes place under some uncertainty that people face regarding others' motives and goals. 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 rather than the latter 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).

This section applies information projection to the problem of social investment. I show that information projection causes people to misattribute the risk associated with social investment to antagonistic preferences. By underappreciating, the extent to which others face the same uncertainty about her preferences as she faces about others', a person comes to believe that the motivation of others is less aligned with hers than it truly is. Continued interaction under preference uncertainty breeds mistrust and lead people to develop a sense of *false uniqueness.* Interacting in a perfectly symmetric situation, and acting identically, people will, nevertheless, come to infer that all others have the opposite preference as they do. I link this to the classic phenomenon described in sociology and psychology under the rubric '*pluralistic ignorance*' (e.g., Prentice 2007).

3.1 Setup

Consider a social investment problem with two-sided private information. Each player i has a type θ_i describing her private valuation of mutual investment. Upon observing their types privately, each decides independently whether to invest (enter) or stay out. If both invest, each player realizes his or her own valuation. If both stay out, each gets an outside option normalized to 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 i.i.d. with a uniform density on $[\theta_{\min}, \theta_{\max}]$ with $\theta_{\min} < 0 < \theta_{\max}$. The key distinction is between positive and negative types. The former are reciprocal and prefer to invest if their opponent invests. The latter are opportunistic and prefer not to invest even if their opponent invests. Formally,

1. **Sorting** Let $f(0) = 0$ and $0 < f_1 < 1$.[12]

I also assume that investment is risky. While the return on one-sided investment increases in the players' type, when at least one of the players is a negative type, one-sided investment still leads to a loss for the investing party relative to the outside option.

2. **Risky Investment** Let $g(0,0) = 0$ with $g_1, g_2 > 0$ and $g(\theta_i, \theta_{-i}) < 0$ if $\min\{\theta_i, \theta_{-i}\} < 0$.

The above assumptions imply that negative (opportunistic) types always prefer to stay out, and positive (reciprocal) types always prefer to invest if their opponent is a positive type and stay out if their opponent is a negative type. In the specification above, the distinction between reciprocal and opportunistic types matches the preference between mutual investment versus the outside option. The analysis below generalizes to the case in which a player - independent of type - also receives a direct benefit $b > 0$, when her opponent

[12] For negative types, it is sufficient to assume that $f(\theta_i) > \theta_i$.

invests.[13] Specifically, when $b + \theta_{\min} > 0$, all types prefer mutual investment relative to the outside option. Yet all other aspects of the game remain the same. In this specification, if all types were commonly known to be negative, the game would correspond to a Prisoner's Dilemma (PD) with a dominant strategy of not investing. In contrast, if all types were commonly known to be positive, the game would correspond to a pure Coordination Game (CG) with a dominant strategy of investing. The problem of social investment then arises because a player faces uncertainty about her opponent's type, hence, about the optimal strategy to employ.

When interpreting the game below, I will sometimes implicitly employ the assumption that $b = 0$ (dating game), and sometimes that $b > -\theta_{\min}$ (trade partnership). In the former case, only positive types benefit from mutual investment relative to the outside option. In the latter case, all types prefer mutual investment to the outside options.

3.2 Main Example - At the Bar

Consider a simple specification. 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:

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

,where $\gamma \to 1$, and, for simplicity, I restrict the domain of θ_i to $[1, -1]$. 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 $c > 0$. 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 make a move (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 values the match positively and rejects if she values negatively. 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, or embarrassment, or simply the cost associated with investing in a futile move. Although this setup describes a sequential-move game, since in the second stage there is always a dominant strategy, it is equivalent to the simultaneous game presented above. A similar interpretation can be given when a asking for or giving a job offer.

[13] That is, when her opponent invests, player i's payoff is either $\theta_i + b$ if she herself invests, or $f(\theta_i) + b$ if she herself does not invest.

♠ **Trust in Trade.** For this application, suppose that $b > -\theta_{\min}$. Partners need to invest in relationship-specific assets to maximize benefits from trade, Williamson (1979). While each party benefits from mutual investment, the return on one-sided investment depends on the type of one's partner. If one's partner is opportunistic (negative type who values mutual investment relatively less), 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. In contrast, if one's partner is reciprocal (positive type), 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, $f(\theta_i) + b > 0$, parties cannot reduce the risk of social investment by pre-pay communication.

♣ **Dissent** A member of an organization either disagrees (positive type) or agrees (negative type) with a norm or practice. She can decide to *voice* her concern and deviate from the norm (In) or *stay silent* and act loyal (Out). A person who agrees with the norm never wants to deviate. Dissent by someone who disagrees with the norm in front of a member who also disagrees with the norm leads to a benefit (a coalition of dissenters or a sense of liberation) for both. Dissent in front of someone who agrees with the norm leads to a loss for the dissenter relative to staying silent. The 'speaker' might be punished or ostracized leading to a loss of c. If $f(\theta_i)$ is positive for the negative type, the loyalist may even receive a reward for reporting the dissenter.[14]

3.3 Projection Equilibrium

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 players come to underestimate the probability that their opponents have matching preferences. Below, the operator E^{ρ} refers to the expected inference of a ρ biased player given the true distribution of type-dependent actions generated by the equilibrium behavior σ^{ρ}.

Proposition 3 *For any $\rho > 0$, there is a unique $\rho - IPE$. This is given by symmetric cutoffs,*

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

Furthermore,

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

II. $E_{\sigma^{\rho}}[E^{\rho}[\theta_{-i} \mid \theta_i > 0]] < 0$ *and is decreasing in c if $\rho > \widehat{\rho}$ where $\widehat{\rho} < 1$.*

[14] Recall that for $\theta_i < 0$ it is sufficient to assume that $f(\theta_i) > \theta_i$.

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

Let me first describe the intuition for the result on actions. Consider the bar example. By projecting information, an interested Judith exaggerates the extent to which her opponent Paul knows that she is interested. As a consequence, Judith underestimates the perceived risk that Paul faces when contemplating a move, and exaggerates the probability that Paul will enter if he is interested. Since Judith still does not know Paul's type, and because a reciprocated entry is almost as good as mutual 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 move. Since the game is perfectly symmetric, the same argument holds for Paul. In the limit, as projection becomes full, each player stays out for sure, even if interested, but both expect the other player to move for sure if interested.

Consider, now, the implications to inference, that is the way players update their beliefs about their opponent. Such inference determines trust and, thus, 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 cutoff than the cutoff he actually uses on average. Hence, when seeing Paul enter, Judith too often thinks that Paul entered only because he knew Judith would not reject him. When seeing Paul stay out, Judith is too convinced that Paul is not interested. Both of these effects on *conditional inferences* are increasing in the degree of information projection ρ. The distortion in the conditional belief following an entry by the opponent is increasing in c and the distortion in the conditional belief following out by the opponent is decreasing in c.

II. Misattribution of Loss The second inference result claims that positive types underestimate their opponents *on average* as well and the extent of this underestimation is increasing in c if players are sufficiently biased. By under-estimating how much risk her opponent faces, a positive type exaggerates the extent to which her opponent should invest if he is positive and not invest if he 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.

To see the comparative static with respect to c, consider first the fully biased case. Here, as long as $c > 0$, no player enters, yet upon observing their opponent stay out they each become more pessimistic and the degree of such excessive pessimism is greater the greater is the wedge between the perceived probability with which a positive type should enter and the actual probability. This wedge here increases in c because it leaves the former unaffected but decreases the latter. A continuity argument shows that the same holds if the bias is not full but sufficiently large hence allowing for the fact that there is a true positive probability of entry in the game.

The above points to a 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, if the bias is sufficiently high, an increase in the payoff risk associated with investing with a negative type always leads to a decrease in the expected ex-post beliefs about the interest, trustworthiness, or matching objectives of others.

III. False Antagonism Finally, in the same way that 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 his 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. Through interacting with others, each player will exaggerate the probability that others have the opposite preference compared from hers.

3.4 Dynamics

The probability that investment happens is decreasing in the loss c for any $\rho \in [0, 1)$. 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. Such a change in c over time could correspond to (i) a change in how formal the setting is where players interact, e.g., a wrong move is very costly in a formal environment and less costly in an informal one; (ii) a decrease in loss due to stronger legal institutions and greater enforceability contracts (which can guard against opportunism and substitute for trust); and (iii) a weakening of the disciplinary (societal or organizational) sanctions against dissent.

Let the dynamic interaction be characterized by a strictly decreasing sequence of losses $\underline{c} = \{c_t\}_{t=1}^{T}$ such that $c_1 < 1$, and T being finite. As will be clear, the fact that the sequence is decreasing is without loss of generality. For simplicity, I focus on myopic interaction, in which, in each period, people care only about the payoff of that period. Nevertheless, at each t each player fully recalls the history of his or her past interactions and this is common knowledge.

In this context, the psychologically natural assumption is that a player 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 had 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, in fact, 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* $\Pr^0(M \mid \underline{c}) = 1 - c_T$*.*

In the Bayesian case, matching is constrained efficient: as the payoff risk associated with social investment becomes zero, all positive matches are formed. Furthermore, matching here is history-independent. In contrast, the next proposition shows that given *any* positive degree of information projection, the reverse can hold. Even as c vanishes, no matches are formed. Furthermore, an extreme form of false uniqueness follows: even if all types are positive, they all 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 (False Uniqueness) *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 skeptical about the type of their opponents. The logic again follows from the differential attribution of non-entry to self and to others.

Note that using Corollary 3, 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 beliefs is also confirming in the sense that players' beliefs about their opponent are never explicitly contradicted. I now describe some implications of the above results and link them to findings and mechanisms described in the psychological 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.[15]

3.5 ♣ Organizations: Dissent

Let me first illustrate some consequences of the above results for social or organizational dissent. Although the interactions above describe bilateral situations, they can be applied

[15] Note that the logic in these co-ordination games differs from the logic of herding on other people's action in sequential social learning in many ways. In herding models there are no direct strategic interactions, rather people infer their own preferences from the actions of others. Here people know their own preferences, but behave as others do because they falsely believe that others have the *opposite* preferences as they do.

Bénabou (2013) offers a mechanism where people's desire to not process publicly available bad news is enhanced by others unwillingness to process bad news in an environment with team production. In contrast to that mechanism people here become too pessimistic about their prospects rather than too optimistic.

to such bilateral interactions taking place pair-wise between all members of a community.

Norm Falsification If there is a potential loss associated with expressing dissent with an existing norm relative to staying silent, people who oppose the norm will come to exaggerate the public support for that norm. In effect, they wrongly interpret the silence of others as their loyalty. Even if no one supports the norm, no one deviates, and everyone eventually loses almost all hope that anyone would deviate.

The results may also matter for understanding disciplinary organizations that sanction dissent. Proposition 3 implies that an organization can maintain loyalty by initially having high sanctions: the greater is the sanction - due to institutionalized terror, humiliation, risk of dismissal - the more members will come to believe that there is genuine support for the practice against which dissent is costly. Such high sanctions, c, can then be replaced with lower sanctions and *further* increase positive types' pessimism about others' willingness to dissent.

Shy Revolutions Corollary 3 allows for a comparative static with respect to the sequence $\{c\}$. A sufficiently large drop in c will lead to an unexpected - relative to positive types' expectations - increase in the fraction of people who dissent (In). To see this, consider the case in which c drops to zero from the one round to the next. First, positive types become too skeptical about others opposing the norm, q^t can be arbitrarily close to 0. At the same time, because the loss associated with investing with the wrong type becomes zero, all positive types express a preference against the norm. Generalizing this argument, since positive types underestimate their opponents following non-entry, in a dynamic context, they underestimate the likelihood of entry that will follow given a *sufficiently large* drop in c, from period t to $t+1$. The positive players' prediction at the 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 information - due to static information projection. Second, they underestimate the likelihood of entry because of 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.

Evidence for Pluralistic Ignorance The predictions of the model are consistent with the 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 might help explain the strategic forces that might lead to this effect.[16]

In an illustrative study, Miller and McFarland (1987) asked students to evaluate their

[16] The description of this phenomenon dates back at least to Hans Christian Andersen's famous tale of "The Emperor's New Clothes."

understanding of a very difficult text. In one condition (the unconstrained condition), students were given the opportunity to stand up in front of their group-mates and ask for clarification from the experimenter in a nearby office.[17] In the other condition (the constrained condition), there was no such option. Given the fact that the task was very difficult, bordering on incomprehensible, most types would have benefited from clarification. The results showed that while no one left the room in the unconstrained condition, students' evaluation of their own understanding of their relative performance was significantly lower in the unconstrained than in the constrained condition. Furthermore, subjects rated themselves lower than they rated the average group member, in the unconstrained condition, but not in the constrained condition.

Related evidence on the support for norms comes inter alia from Prentice and Miller (1993). They 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. Students overestimated how comfortable others were and also how universal such a support was. Similar effects were found in the context of racial segregation with white males exaggerating how much other white males supported racial segregation, O'Gorman (1979). Kuran (1995) argues that pluralistic ignorance plays a role in social change and revolutions often come as a surprise. For example, a year after the collapse of the Berlin Wall, former citizens of the German Democratic Republic were surveyed about whether they had expected that such a change was possible, and, despite the benefit of hindsight, 76% of respondents indicated that they were totally surprised of the possibility of such a change.[18],[19]

3.6 ♠ Trade: Legacy of Mistrust

The next Corollary has implications for how the legacy of weak institutions might matter for genuine trust. To see this, consider again repeated myopic encounters with two periods as before. Compare two cases. Suppose while the payoff loss associated with investing with the wrong type is the same in the second period for both cases, in the first period it is higher in case (i) than in case (ii). It follows that there will be overall less entry in case (i) than in case (ii).

Corollary 4 *Consider two decreasing sequences* $\underline{c}$ *and* $\underline{c}'$ *s.t.* $c_1 > c_1'$ *and* $c_2 = c_2'$*. If* $\rho > 0$*,*

[17] It was confirmed that asking for clarification in the first condition was embarrassing in case others did understand the text.

[18] See Kuran (1991). Elster (2007) also argues that Tocqueville (1856) considered the logic of pluralistic ignorance in his account of the French revolution.

[19] The result on silent revolutions differs markedly from a multiple-equilibrium logic of change. There people understand the possibility of moving from one regime to another which is thus a matter of coordination only. In the game studied here, each positive player instead becomes convinced that staying in, a bad situation from her perspective, is what others prefer, hence loses hope.

the overall probability of enter is smaller (sometimes strictly) under $\underline{c}$ than under $\underline{c}'$.

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. 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 latter community will be less likely to invest than members of the former community.

3.7 $\diamondsuit$ Mistaken Segregation

Finally, one can perform a comparative static with respect to the uncertainty that players face about each other's types. Let α be the probability with which a person can identify ex-ante whether her opponent has a positive or a negative valuation. Suppose that the realization of this event is private and independent from type and across players. If this probability is 1, there is no informational asymmetry and projection has no bite. In contrast, false antagonism perceived by positive types is maximal when this probability is 0, and decreases as this probability increases.

Corollary 5 *For any $\rho > 0$, $E_{\sigma^\rho}[E^\rho[\theta_{-i} \mid \theta_i > 0, \alpha]]$ is increasing in α.*

To illustrate the consequences, consider two groups A and B, and suppose that different members of the same group can read each other's type ex-ante with a higher probability than they can read the type of a member of another group. 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. The extent of segregation decreases when initial payoff risk associated with matching with an antagonistic type is decreased. In the presence of payoff risks and lack of sufficient information, however, the extent of mistaken segregation potentially increases as people's opportunity to interact with members of the other group increases.

Shelton and Richeson (2005) provide suggestive evidence in the context of interracial friendship formation. They show that both White and Black students at Princeton and at U. Mass desired having more interracial 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 for prejudiced Whites, who did not want interracial friendships.

3.8 Investment Games

Let's return to the more general case presented in the Setup. A formal distinction between complement and substitute investments is needed here. I call investments *complements* if the game exhibits increasing differences: given positive types, the difference between the return on investing with an investing opponent, $\theta_i - f(\theta_i)$, as opposed to a non-investing opponent, $g(\theta_i, \theta_{-i}) - 0$, is increasing in type. In contrast, investments are *substitutes* if this differential 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 $\theta > 0$. Investments are substitutes if $\theta_i - f(\theta_i) - g(\theta_i, \theta_{-i})$ is decreasing in θ_i for all $\theta > 0$.*

In the main example above, the parameter $\gamma \in (0, 1)$ governed the substitutability of investments. In the case in which $\gamma \to 1$, investments were almost perfect substitutes. A one-sided initial investment implemented almost the same payoffs as mutual investments, conditional on both players being positive 'reciprocal' types. For all $\gamma > 0.5$, investments remain substitutes - $\theta_i(1 - 2\gamma)$ is decreasing in θ_i – and the statements of Proposition 3 continue to apply. There is a unique equilibrium with a cut-off increasing in ρ.

Going to the Bar Consider the bar example, but assume that $\gamma < \frac{1}{2}$. Investments are now complements. As $\gamma \to 0$, only mutual investment leads to positive gains relative to the outside option. A reciprocated investment implements the same payoff as the outside option for the investing party. Returning to the bar example, this could relate to a situation where each player needs to decide whether to go to a bar or not. Only if *both* go, do they have the opportunity to spend the night together. If they are both interested, there is still no shame cost following one-sided investment, but by the time the non–investing party learns about her partner's interest and shows up, the night is already almost gone, leaving them with a payoff of $\gamma\theta_i \to 0$. If the opponent is non-interested, the investing party still incurs a shame cost of c.

The proposition below implies that the implications of information projection on actions depends on whether investments are substitutes or complements. All equilibria are again in cut-off strategies. When investments are substitutes, as in 'at the bar' example, this mistakes leads to too little entry relative in the unique symmetric BNE. When investments are complements, as in the 'going to the bar' example, this mistake may lead to over-entry. Here, there are potentially multiple symmetric equilibria, and while the lowest one decreases in ρ, higher ones might as well increase in it. Note that4, since each player exaggerates the probability with which her opponent should act in accordance with her type, the qualitative inference results remain the same under both complement and substitute investments. In the proposition below, I call a game a linear if $g_2 = g_{11} = f_{11} = 0$ if $\theta_i > 0$.

Proposition 4 *For any $\rho > 0$, all equilibria are given by cut-off strategies:*

(i) If investments are substitutes, there is a unique symmetric equilibrium and it is increasing in ρ.

(ii) If investments are complements, and the game is linear, then there are at most two symmetric equilibria, and the lowest equilibrium cut-off is 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 equilibria, $E^{\rho}[\theta_{-i} \mid a_{-i}, \theta_i > 0] \leq E^{0}[\theta_{-i} \mid a_{-i}, \theta_i > 0]$ for all $a_{-i} \in A_{-i}$.

The logic 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, then the best response functions are increasing in the opponent's cutoff, and the perceived return on entry is now potentially exaggerated, because a positive type exaggerates the probability that her opponent enters. As a result, information projection may lead to over-entry relative to the Bayesian case in games with complement investments.[20]If investments are substitutes, then the perceived return on entry is underestimated for the same reason as before, and information projection leads to under-entry in games in the unique symmetric equilibrium.

False Antagonism (iii) All projection equilibria will exhibit *false antagonism*. 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 of her opponent staying out and over-infers from entry.

Undervaluation of Social Assets (iv) Note that all positive types come to underestimate the type of their opponent both if the opponent enters and if the opponent stays out. Thus, even if a match is formed, the parties are too pessimistic about how much their opponent values the investment going forward. This can diminish the extent of costly positive investment into the match, which pays off only if one's partner also values the match at a sufficiently high level. As a consequence, people will underestimate the return on investing in such social assets.

False Consensus on Actions Finally, note that while the proposition predicts false antagonism when *inferring* preferences, this is not inconsistent with a false consensus effect on actions in this setting with independent types.[21] In fact, any positive type who enters exaggerates the probability that her opponent will enter, relative to the truth. Any negative

[20] The higher equilibrium, if exists, will increase in ρ in a linear game.

[21] Despite the term'false consensus effect', e.g., Ross, Greene and House (1976), the realization of one's signal is often the best predictor of the signal realization of others for there to be a positive correlation between one's action and the predicted action of the opponent. Hence the label 'false' might often not be warranted. See, e.g., Dawes and Mulford (1996).

type who does not enter will exaggerate the probability that her opponent will not enter, relative to the truth. Of course, positive types who do not enter do the reverse. Crucially, as Proposition 4 shows, when embedded in strategic interaction, such exaggeration happens in conjunction with developing false uniqueness, which can remain self-confirming through repeated interactions, Corollary 3.

4 Persuasion

In this section, I study a simple sender-receiver problem with costly state verification. A financial adviser makes a recommendation to an investor on whether to buy or sell an asset. I show that in a setting with commonly known conflict of interests between the parties, information projection predicts credulity: receivers believe good news too much, and sophisticated senders take advantage of such credulity, which leads 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.

Comparative static predictions show how a small change in the market size can push communication from inducing correct average beliefs and investment to inducing inflated average beliefs and, on average, too much investment into an uncertain asset. While in the Bayesian case, communication is always weakly beneficial, under information projection, persuasion can uniformly lower the welfare of receivers.

The problem of credulity in financial advice, and in persuasion more generally, is a widely recognized issue in economics. In the context of the recent financial crisis, many has called attention to the inflated credit ratings, e.g., The United States Senate Permanent Subcommittee on Investigations (2011). While an incentive to inflate recommendations exists, a puzzle remains of why such inflated credit ratings, which are often strategically distorted, are nevertheless believed by investors. For example, Malmendier and Shanthikumar (2007, 2014) provide evidence that small investors take positive recommendations too literally and fail to sufficiently discount the extent to which these are strategically inflated. Understanding the strategic mechanism that leads to such credulity and its implications is then key.[22]

As mentioned, information projection in a common value environment can lead to an exaggeration of such correlations if valuations are positively related conditional on the state. It will lead to the opposite when valuations are negatively related conditional on the state.

[22] Exogenously invoked naive types who always take recommendations at face value have been considered, see e.g., Kartik, Ottaviani and Squintani (2007).

4.1 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 the prior on this state to be symmetric.[23]

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 $[0, \infty)$, cdf $F(c)$. Heterogeneity in costs may reflect differences in 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 to 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 the receiver decides to check the recommendation and finds out that 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.

4.2 Bayesian Case

Consider, first, the BNE. This equilibrium has a simple structure: 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.

Proposition 5 *Let $\rho = 0$. 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, $E[y_c^*] = \frac{1}{2}$.*

[23] No qualitative result depends on this symmetry assumption.

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 being transmitted. Note that, communication, on average, is neutral: each receiver's type equilibrium belief follows a martingale. This means that the ex-ante expected ex-post investment (belief) is the same as the true 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.

4.3 Biased Persuasion

Consider an unbiased sender ($\rho_S = 0$) and a biased receiver ($\rho_R = \rho$). Persuasion is no longer neutral. Rather, information projection leads to two kinds of mistakes: *credulity* by some types and *disbelief* by some other types. The former means that persuasion itself predictably increases average beliefs. The latter means that persuasion itself predictably decreases average beliefs. Under credulity, belief updating forms a sub-martingale process, and expected ex-post investment is higher than the ex-ante expected investment, i.e., $E[y_c^*] > \frac{1}{2}$. Here, the receiver takes a positive recommendation too much at face value. 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., $E[y_c^*] < \frac{1}{2}$. Here, the receiver is too skeptical when hearing a positive recommendation. The next proposition describes the unique $\rho-$IPE of the game.

Proposition 6 *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 is credulous and overinvests;

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

(iv) if $c > c_3^\rho$, the receiver never checks and is (weakly) in disbelief;

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

In an IPE, each receiver type exaggerates the extent to which the sender knows her type, the realization of c, and, thus, the probability with which the sender tailors the message to how easy or difficult it is for the receiver to check its validity. The imaginary projected sender who knows the receiver's type lies less often to a receiver with higher financial expertise than to a receiver with a lower financial expertise. While, in reality, the sender lies to each type with the exact same probability, different types will, thus, have different beliefs about this probability. 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. As a result, receivers with low financial expertise, $c > c_2^\rho$, will exaggerate the probability that a high recommendation is a lie, while receivers with medium financial expertise, $c \in [c_1^\rho, c_2^\rho]$, will exaggerate the probability that a high recommendation is not a lie - they are credulous. Receivers with high financial expertise would also be credulous, except that they always check and, thus, figure out the truth.

The logic of the equilibrium divides the receiver's types into three regions. If p^{ρ} is the equilibrium probability with which the sender lies, then type c_1^{ρ} is the type who is indifferent between checking and not checking when he perceives the probability of lie to be $(1-\rho)p^{\rho}$. Similarly, type c_2^{ρ} is the type who is indifferent between checking and not checking when he perceives the probability of a lie to be $(1-\rho)p^{\rho}+\rho$. Types below c_1^{ρ} then always check and types above c_3^{ρ} never do. In the Bayesian case $c_1^0 = c_3$, however, for any $\rho > 0$, this no longer holds, and types between these types, mix between checking and not checking.

True Leakage Note that the results on expected beliefs and average distorted investments do not depend on leakage of a receiver's type per se. Instead, it is the consequence of the exaggeration of such leakage. If there is some true commonly known probability α with which the sender can privately figure out the receiver's level of expertise, then all Bayesian properties continue to hold. The structure of the equilibrium is the same as in Proposition 6, but no type is credulous and no type is in disbelief. Instead persuasion is again *neutral,* on average. In contrast, projection implies the exaggeration of α to $\alpha+(1-\alpha)\rho$. The statements in Proposition 6 continue to hold for any $\alpha > 0$ and $\rho > 0$.

4.4 Credulity

Let's turn to the comparative statics. To describe these, consider first the impact of an unobservable increase in the bias. Part (v) of Proposition 6 claims that this decreases both the set of types who always check, i.e., c_1^{ρ} decreases, and the set of types who never check, i.e., c_3^{ρ} increases. To see the logic, suppose that in contrast after an increase in the degree of projection, c_3^{ρ} decreased. This must mean that the probability of the sender lying must have decreased since type c_3^{ρ} always thinks that the projected sender always lies to her. This then implies that c_1^{ρ} must also decrease. While this would mean that the total incentive for the real sender to lie is increased, implying that c_3^{ρ} must have decreased, which leads to a contradiction.

The consequence is that if the bias is sufficiently high, all types become at least weakly credulous, and a positive measure of types becomes strictly credulous. In sum, everyone (at least weakly) over-invests and persuasion increases aggregate beliefs; also lowering the welfare of all receiver types. I refer to this case as *uniform credulity.* The next proposition shows that for any given degree of projection such uniform credulity holds, provided, the conflict of interest, B, is sufficiently high. Furthermore, the larger is the degree of projection, the lower is the minimally necessary B inducing it.

Proposition 7 *In the unique $\rho - IPE$,*

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 only 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 (6) 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, maintaining the assumption that F has full support, brings about uniform credulity. I now turn to these predictions in more detail.

Consider, first, a change in the conflict of interests. In the Bayesian case, a change in B leads to decreased information transmission, p^0 increases, but has 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.

Consider now a change in complexity or financial education, F. In the Bayesian case, the cost of processing information does not affect the expected level of confidence. 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 mostly concentrated on 0, then everyone learns the truth. As the product becomes more complex, in that it is more costly to check the sender's recommendation, the probability of overly optimistic average beliefs increases. If F is sufficiently high, however, such that for virtually no types would it be rationalizable to check, then IPE again induces mostly correct average beliefs. Here, the sender always lies and almost all receivers expect this. Hence, a downward shift in F can *lower* investors welfare. Greater financial education can increase credulity. It is the presence of a mass of types with some, but not full financial education that drives the market for financial advice towards exuberance in this model.

4.5 Endogenous Conflict and Belief Bubbles

In the analysis above, the conflict of interest B was exogenous. Let me now endogenize this benefit by invoking the owner of the asset. The owner of the asset pays the sender (rater) or continues business with a financial advisor conditional on positive recommendations. This is fully transparent. In other words, the seller pays B to the sender (rater) for sending a high message, and this is common knowledge. What is the optimal commonly known transfer, bribe, that the seller who understands the behavior of the sender and the receiver in equilibrium would want to offer?

Suppose that the seller's profit is simply given by the aggregate demand for the asset minus the bonus it pays:

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

where γ is the size of the market or a measure of the profit/markup on each unit of demand. Recall that $E_{c,\theta}[y^*]$ simply reflects the expected aggregate demand for the asset in the population, which is nothing else here, but the ex-ante expected posterior of investors.

Consider the Bayesian case. Here, by virtue of the martingale property of Bayesian communication, the seller's ex-ante 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 implies that the optimal bribe B is zero.

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

Consider the biased case. The proposition below shows that if the markup is sufficiently high, then the seller always wants to offer a positive bribe 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 8 *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 $E_{c,\theta}[y^*] = 0.5$.*

A key feature of the optimum here is that if γ is sufficiently high, there is always 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 allowing for greater margins -or more complex - such as in the case of a financial innovation.

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 the asset. Similarly, after a decrease in the profitability parameter below

$\underline{\gamma}(\rho)$ the seller withdraws payment from the rater. Optimistic beliefs burst, and investors' average expectations become realistic again. 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.

Welfare: **Caps versus Disclosure** The analysis above has implications for 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. The solution in such models is mandatory disclosure: highlight the conflict and thereby eliminate naïveté. In contrast, here, the conflict is always common knowledge, but credulity persists despite this fact. Disclosure of the conflict is ineffective. Instead, capping the bribe B, or increasing S, 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; 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 a low level of financial sophistication.

5 Projection Equilibrium

The model above focused on information projection only. A logical counterpart of information projection is *ignorance projection:* the wrong belief that if one cannot condition her strategy on an event, one's opponent cannot either. Direct evidence on such ignorance projection is much more sparse than that for information projection. Thus, it might be a weaker force. Nevertheless, the model introduced here 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 projection equilibrium.[24]

If player j projects both her information and ignorance onto player i, she exaggerates the probability that her opponent can condition his strategy on the *same* events she can. Formally, consider the strategy set

[24] It is possible to incorporate information and ignorance projection simultaneously but separately. Here ignorance projection alone would correspond to a false belief that one's opponent conditions his strategy only on information that is truly commonly known by the players, i.e., the finest common coarsening of the two information partitions. I leave such an extension to future work.

$$S_i^j = \{\sigma_i(\omega) \mid \sigma_i(\omega) \in \Delta A_i \text{ measurable w.r. to } P_j(\omega)\}, \tag{9}$$

denoting the strategies that the projected player i could use if she had exactly the same information about ω as j. One can then state the definition of a projection equilibrium in perfect analogy with that of information projection equilibrium.

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

1.

$$\sigma_i^\rho \in BR_{S_i}\{(1-\rho)\sigma_{-i}^\rho \circ \rho\sigma_{-i}^i\} \tag{10}$$

2.

$$\sigma_{-i}^i \in BR_{S_{-i}^i}\{\sigma_i^\rho\} \tag{11}$$

Note that the definition again satisfies the *all-encompassing* and limited *consistency* properties, as before. The former is true because the projected version of a player again thinks that her opponent is regular for sure. If the true game is poker, a biased Judith thinks that, with probability ρ, Paul knows *her* cards only, and also knows that those are the only cards Judith knows. In other words, Judith thinks that, with probability ρ, Paul knows *exactly* what she knows. The existence of projection equilibrium follows the same logic as that of IPE. Equivalent versions of Proposition 2 and Corollary 1 continue to hold. Similarly, the structure of higher-order perceptions is the same as before. Paul thinks that Judith is like him w.p. ρ (first-order); Judith underestimates this and thinks that, on average, Paul wrongly thinks that Judith is like him w.p. $\rho - \rho^2$ (second-order); Paul exaggerates this and thinks that, on average, Judith thinks that Paul thinks that Judith is like him w.p. $\rho - \rho^2 + \rho^3$ (third-order); and so on. Higher-order perceptions then again all converge to $\rho/(1+\rho)$. Finally, the definition immediately extends to heterogeneous projection in the same way as before.

5.1 Common-Value Bargaining

Let me illustrate the predictions of projection equilibrium by applying it to bilateral trade with common values, Akerlof (1970). The seller has an item of quality $q \in \mathbb{R}^+$ and values it at q. The buyer's valuation is $w(q) > q$, so it is common knowledge that there are benefits from trade. Quality is distributed according to a continuous density π, and its realization is observed only by the seller.

An extensive experimental literature starting with Samuelson and Bazerman (1985) documents systematic deviations from BNE in this setting, and there is a theoretical literature providing different accounts of the buyers' anomalous behavior in this setting, e.g., Eyster and Rabin (2005), Esponda (2008). The rest of this Section shows that when the buyer has the bargaining power, projection equilibrium predicts the classic winner's curse. When the seller has the bargaining power, it predicts non-altruistic truth-telling. While in the latter setting, the predictions of cursed equilibrium coincide with those of BNE, I compare the model's predictions with CE in the former setting.

Buyer Offer Suppose that the buyer makes a take-it-or-leave-it offer (TIOLI) that the seller can accept or reject. To derive $\rho-$PE, consider the buyer's perceived expected utility when making an offer of p_b. This is given by

$$(1-\rho)\Pr(p_b \geq q)(E_\pi[w(q) \mid q \leq p_b] - p_b) \text{ if } p_b < E_\pi[q],$$

$$(1-\rho)\Pr(p_b \geq q)(E_\pi[w(q) \mid q \leq p_b] - p_b) + \rho(E_\pi[w(q)]) - p_b) \text{ if } p_b \geq E_\pi[q].$$

where $E_\pi[q]$ is the ex-ante expected quality of the object. This is true because both the informed seller and the fictional projected seller - who is equally uninformed as the buyer - have a dominant strategy. The the latter accepts p_b if it is greater than $E_\pi[q]$.

Seller Offer For simplicity, I only derive the predictions of a full projection equilibrium where $\rho \to 1$. Here, the seller always (bids arbitrarily close to) $w(q)$. Generically, such a pure-strategy fully-revealing equilibrium will not be a BNE since a seller with a lower quality who recognizes that the buyer is uninformed would want to deviate and bid higher. Thus, information projection predicts 'excessive' truth-telling here; the fact that the seller holds the buyer to her reservation value makes it evident, however, that the source of this is not a form of altruism.

5.1.1 Multiplicative Lemons Problem

Holt and Sherman (1994) consider a problem where $w(q) = mq$ and π is uniform on $[q_0, q_0 + r]$. It is easy to show that projection equilibrium is discontinuous in ρ: if $\rho < \rho^*$, it is the same as the BNE, and if $\rho > \rho^*$, $p_b^\rho = E_\pi[q]$. Hence, relative to the Bayesian prediction, it is sufficient to represent it by ρ^*.

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

	$[r]$	$[q_0]$	$[m]$	$b(\chi=0)$	$b(\chi=1)$	$b(\rho>\rho^*)$	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).

In all conditions examined by ER (2005), as long as $\rho > 0.07$, 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 very low. In the winner's curse condition - where people overbid relative to the Bayesian prediction - the relevant cut-off is $\rho^* = 2\%$. 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' bids concentrate on the point predictions of projection equilibrium provides further support. The intuition for why such a small degree of ignorance projection is sufficient to induce the buyer to bid the seller's unconditional expected valuation results from the familiar logic that a buyer's utility gain from trade with selection is much smaller than the buyer's utility gain in the absence of selection. Hence, if the buyer mistakenly puts positive weight on the seller being symmetrically uninformed, then acting optimally given this contingency is much more beneficial than under the scenario where the seller's acceptance is subject to selection.

A variant of the above specification is where $q_0 = 0$ and $r = 1$. This game is studied experimentally by Ball, Bazerman, and Carroll (1991), who allow for multiple rounds of learning. Here projection equilibrium predicts that if $\rho > \rho^* = 0.12$, then $b^*(\rho) = \frac{1}{2}$ and 0 otherwise. The Table below summarizes the results:

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

Cursedness versus Ignorance Projection The logic of cursedness and projection differs not only in the point predictions – and that the former smoothly spans an interval, while the latter is concentrated on two points only – but also in the equilibrium beliefs about the perceived probability of acceptance. A fully cursed buyer - the prediction that, in this parametric class, is closest to the data - believes that her offer should be accepted independent of q. In contrast, a buyer who projects her ignorance, *understands selection* and assigns a potentially very high probability to its presence. Note that while cursed equilibrium would predict a positive bid even if $m < 1$, projection always predicts a bid of 0 in this case.

5.1.2 Additive Lemons Problem

Samuelson and Bazerman (1985) consider a specification where $w(q) = q + 30$ and π is the uniform on $[0, 100]$. They report data in a more aggregate form. The unique prediction of the model here is given by

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

Samuelson and Bazerman find that offers bunch at 50 with 70% of bidders bidding in the $[50, 80]$ interval and a positive fraction bidding above 60. The sellers follow their dominant strategy to accept prices above their reservation value. Note that under correct Bayesian expectations bidding above 60 leads to a negative earning and hence it is a dominated strategy. In contrast, bidding below 80 can be rationalized by some degree of projection combined with social preferences, as these all lead to positive perceived expected earnings in a projection equilibrium.[25]

Samuelson and Bazerman also study the bargaining protocol where the seller makes a single TIOLI. They show two facts. First, roughly half of the sellers bid a value that is equal to the conditional reservation value of the object to the buyer – that is $w(q)$. Thus, they reveal the quality of the object, but leave no surplus for the buyer. Second, the overwhelming majority of the remaining bids are in the interval $[q, w(q)]$. Third, buyers accept such bids close to uniformly such that, given the buyers' behavior, sellers *underbid*.

All of the above facts are consistent with projection equilibrium. For a sufficiently high ρ, there exists a *fully* separating pure-strategy projection equilibrium where the seller's bid $p_s^\rho(q) = q + 30$ for all q and buyers accept for sure. No such pure equilibrium exists if $\rho = 0$. A sufficient condition for full revelation is $\rho \geq 10/13$. Here, $(1 - \rho)130 \geq 30$, and then by monotonicity, it follows that no seller type has the incentive to deviate from $p_s^\rho(q) = q + 30$. Hence, given the buyers' true behavior, sellers leave no rent, but bid lower than the profit-maximizing bid. In sum, the model simultaneously predicts over-bidding by buyers who are then subject to the winner's curse and under-bidding and truth-telling by sellers in this common-value bargaining setting.

Valuations or Information The above evidence is also consistent with the idea that projection is based on information. If sellers projected their *valuations* instead of their information about the state, i.e., q, then they should bid significantly below the buyers' valuation $w(q)$ simply by respecting perceived dominance.

[25] Fudenberg and Peysakhovich (2013) provide careful evidence for additive lemons problem with buyers interacting with computerized sellers. Their focus is on learning, but one can relate the predictions of this model to their Experiment 1. In the case where $w(q) = q + 3$, the prediction of PE is $b(\rho) = 3$ if $\rho \leq 1/16$ and $b(\rho) = 5$ otherwise. The average empirical bid in the baseline case was 5.148. In the case where $w(q) = q + 6$, the predictions of PE is $b(\rho) = \frac{6-16\rho}{1-\rho}$ if $\rho \leq 1/11$ and $b(\rho) = 5$ if $\rho \geq 1/11$. The average empirical bid in the baseline case here was 6.02.

5.2 Multi-Player Extension

So far, I, considered only two-player games. The model can readily be extended to N players. The key aspect of this extension is that, now, each player i has a collection of fictional opponents. Each such fictional opponent j shares player i's information, or in the case of projection equilibrium, player i's ignorance as well. Since the information of player k 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. The fictional version of j is now specific to a given player i. Let then $S^i_{-i} = \times\prod_{j\neq i} S^i_j$ be the strategy set of the fictional opponents who are real in player i's imagination. I denote the generic element of this set by σ^i_{-i}.

Projection is again *all-encompassing*. In equilibrium, player i thinks that the projected versions of her opponents have the exact same beliefs about all other players as she does and that they also know player i's actual state-contingent strategy. Player i's expectations are again consistent - in a limited way - in that she assigns probability $1-\rho$ to her opponents all playing the strategy that they truly play. This then corresponds to the following formal statement.

Definition 5 *Consider a game Γ. A strategy profile $\sigma^\rho \in S = \times\prod_{i=1}^N S_i$ is a projection equilibrium if for all i there exists $\sigma^i_{-i} \in S^i_{-i}$ such that*

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

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

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

The above defines the multi-player extension of projection equilibrium. The extension of information projection equilibrium is analogous and differs only in that each S^i_j is replaced by $S^{i+j}_j = \{\sigma_i(\omega) \mid \sigma_i(\omega) \in \Delta A_i \text{ measurable w.r. to } P_j \cap P_i\}$.

6 Conclusion

In this paper, I introduced informational projections into a class of Bayesian games. I developed a model containing only information projection and one combining the presence of both information and ignorance projection simultaneously. Beyond the problems considered in this paper, the model is likely to deliver a number of novel predictions in many other settings, e.g., bargaining with two-sided incomplete information. In a companion paper, Madarász (2014b), I extend the model to a dynamic bargaining setting. Dynamic extensions may also help understand anomalies in the evidence on social learning.

7 Appendix

Proof of Claims 1&2. The derivation of the private$\rho - IPE$ follows directly from the BNE of the game. Let $p = \pi(\omega_w)$

$p< \frac{1}{2}$	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	$\frac{1-\omega_w}{2-\omega_w}$
	$\frac{1}{2}a \circ \frac{1}{2}b$	$\frac{1}{2}a \circ \frac{1}{2}b$			$\frac{1}{1+\omega_w}a\circ\frac{\omega_w}{1+\omega_w}b$	$\frac{1}{1+\omega_w}a\circ\frac{\omega_w}{1+\omega_w}b$	

The defender's winning probability in case of no private information is $\frac{1-p\omega_w}{2-p\omega_w}$.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, $\frac{1-\omega_w p}{2-\omega_w} > \frac{1-\omega_w}{2-\omega_w}, \frac{1-\omega_w p}{2}$. .

Proof of Claim 3. 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 $\sigma^+_{-i}(\theta_i, \theta_{-i})$. If the fictional version of $-i$ has $\theta_{-i} < \theta_i$, then he will bid higher than before, as long as the lowest value of the support of $\sigma^0_i(\theta_i)$ is lower than θ_{-i}. Suppose his bid is unchanged otherwise. If $\theta_{-i} \geq \theta_i$, then he will now bid lower, but bid the highest value of the support of $\sigma^0_i(\theta_i)$. Consider now player $i's$ best response:

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

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 even if player i bidder bids above the support of $\sigma^0(\theta_i)$, for any θ_i. At the same time, bidding lower than under σ^0, the probability of winning is lower than in the case where $\rho = 0$. Hence bidding below the support of $\sigma^0_i(\theta_i)$ cannot be a private information projection play.[26] 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. This incentive can only be reinforced by higher informed types. The discontinuity for any $\rho > 0$, follows from the fact that the interval over which θ_i was supposed to mix given σ^0 has positive measure. .

Proof of Proposition 1. The existence of IPE (PE) follows from Kakutani's theorem. Consider IPE. The mapping from a real to a perceived strategy profile $\sigma^\rho \rightarrow (\sigma^\rho, \sigma^+)$ 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

[26] This follows from the fact that the relevant derivative did not decrease at any point.

(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 that is BNE of Γ there there exists $\sigma^+ = \sigma$ which satisfies the definition of a IPE since $BR_{S_i^+}\{\sigma^\rho_{-i}\} = BR_{S_i}\{\sigma^\rho_{-i}\}$ for any ρ. By the same token, the reverse is true because if a player is indifferent between two strategies in equilibrium, then there exists a BNE where this player mixes between them. The logic immediately extends to projection equilibrium .

Proof of Proposition 2. Suppose that σ^0 is a BNE and it is also a strict ex-post equilibrium in Γ. It must then be true that for any $\sigma_i^+ \in BR_{S_i^+}(\sigma^0_{-i})$, $\sigma_i^+ = \sigma_i^0$ hence $\sigma^\rho = \sigma^0$ is a $\rho - IPE$ for any ρ .

Proof of Proposition 3. It follows from the proof below that equilibrium is always in cut-off strategies. If θ_{-i} is the cutoff of player $-i$, then player i is indifferent at θ_i satisfying

$$\rho(\theta_i - \gamma\theta_i - c) + (1-\rho)((1-\theta_{-i})(\theta_i - \gamma\theta_i) + \theta_{-i}(\gamma\theta_i) - c) = 0$$

Taking $\gamma \to 1$, one obtains that for each i,

$$\theta_i = \frac{c}{\theta_{-i}(1-\rho)}$$

Solving these equations, it follows that $\theta_i^{*,\rho} = \min\{\sqrt{c/(1-\rho)}, 1\}$ for each i and the solution is thus unique.[27] The result on conditional inferences holds since $\theta_{-i}^{*,\rho} > \theta_{-i}^+ = 0$ for all $c, \rho > 0$ and decreasing in ρ.

For the result on average inference, let $\Pr^\rho(in)$ be the perceived probability by player i with which a positive type expects her opponent to enter given the prior on types. From the martingale property of beliefs it follows that

$$\overset{\rho}{\Pr}(in)E_1^\rho[\theta_{-i} \mid \theta_i > 0, a_{-i} = in] + (1 - \overset{\rho}{\Pr}(in))E_1^\rho[\theta_{-i} \mid \theta_i > 0, a_{-i} = out] = E_0[\theta_{-i} \mid \theta_i]$$

Since $\Pr^\rho(in)$ is always greater than the true $\Pr(in)$, as long as $c > 0$, the extent of the underestimation is given by

$$[\Pr(in) - \overset{\rho}{\Pr}(in)][E_1^\rho[\theta_{-i} \mid \theta_i > 0, a_{-i} = in]] - E_1^\rho[\theta_{-i} \mid \theta_i > 0, a_{-i} = out]] < 0$$

To show the comparative static with respect to c we need to consider the expected inference of a positive type. For simplicity, let's denote by $b = \theta_{-i}^* = \theta_i^*$ the cut-off of a regular player used in equilibrium. It follows from the above that this cut-off is increasing

[27] In what follows, for convenience, I omit denoting the upper-bound o1on $\theta_i^{*,\rho}$, but assume that it binds.

in c. Consider now the expected posterior of a positive type who enters. This is given by

$$E[\theta_{-i} \mid \theta_i > 0, a_i = in] = \frac{1-b}{2}\left(\frac{(1-\rho)(1-b)(1+b)/2 + \rho/2}{(1-\rho)(1-b)+\rho}\right) + \frac{b}{2}\left(\frac{b}{2}\right) + \frac{1}{2}\left(-\frac{1}{2}\right)$$

since if her opponent does not enter, such a player has correct posteriors after observing her payoff and thus learning whether the opponent is positive or negative. Consider now the expected posterior of a positive type who does not enters. This is given by

$$E[\theta_{-i} \mid \theta_i > 0, a_i = out] = \frac{1-b}{2}\left(\frac{(1-\rho)(1-b)(1+b)/2 + \rho/2}{(1-\rho)(1-b)+\rho}\right) + \frac{1+b}{2}\left(\frac{(1-\rho)(b+1)(b-1)/2 - \rho/2}{(1-\rho)(b+1)+\rho}\right)$$

The expected posterior conditional on player i being positive is then $(1-b)E[\theta_{-i} \mid \theta_i > 0, a_i = in] + bE[\theta_{-i} \mid \theta_i > 0, a_i = out] =$

$$E[E^{\rho}[\theta_{-i} \mid \theta_i > 0] = \overbrace{-\frac{1}{4}\frac{b\rho}{1-b^2(1-\rho)^2}}^{I}\overbrace{\left(b^2\rho + (b-b^2-b^3)(1-\rho)+1\right)}^{II}$$

To show that this smooth function decreases in b if ρ is sufficiently large, note that if $\rho \to 1$, $E[E^{\rho}[\theta_{-i} \mid \theta_i > 0] = -\frac{1}{4}b\left(b^2+1\right)$. Here of course $b = 1$ must hold for all $c > 0$, hence $E[E^{\rho}[\theta_{-i} \mid \theta_i > 0]$ can also be expressed as $-\frac{1}{4}\frac{2\rho}{2-\rho}$. Consider now the impact of a change in b when ρ is bounded away from 1. The sign of the derivative of the first term equals

$$-\frac{\rho}{\left(b^2\rho^2 - 2b^2\rho + b^2 - 1\right)^2}\left(b^2(1-\rho)^2+1\right) < 0$$

and hence it is always negative. The second term, however, increases in b if ρ is sufficiently large, such that $2b\rho + (1-2b-3b^2)(1-\rho) > 0$. Hence if ρ is sufficiently large, the comparative static prediction follows .

Proof of Corollary 2. Suppose $\rho = 0$. Beliefs in round t, given no entry until the end of round $t-1$, are left-truncation of beliefs of period $t-1$, given the monotonicity of the equilibrium. This will correspond to a belief that the opponent's type is uniformly distributed on $[\theta_t, -1]$. Inductively, 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, i.e., $q_t = \frac{\theta^*_{t-1}}{1+\theta^*_{t-1}}$ given symmetry. Player i's indifference condition - determining the cut-off $\theta^*_{i,t}$ - in period t given θ^*_{t-1}, with $\theta^*_0 = 1$, is given by,

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

Solving these equation we get that

$$\frac{c_t}{\theta^*_{t-1}(1-\gamma)+2\gamma\theta^*_{-i,t}-\theta^*_{-i,t}} = \theta^*_{i,t}$$

Then substituting in the symmetric equation for $\theta^*_{-i,t}$, we get that

$$\frac{1}{4\gamma-2}\left(\theta^*_{t-1}(\gamma-1)+\sqrt{-4c+\theta^{*2}_{t-1}\gamma(\gamma-2)+8c\gamma+\theta^{*2}_{t-1}}\right)$$

taking $\gamma \to 1$, we get that

$$\theta^*_{i,t} = \sqrt{c_t}$$

Furthermore, it is easy to see that the efficiency probability by period t is given by $1-c_t$, since by then all types greater than c_t have entered. By virtue of the fact that c_t is decreasing it also follows that there is positive entry and hence an indifferent type in each round .

Proof of Corollary 3. Consider now $\rho > 0$. Let $q^\rho_{t,-i}$ be the probability player i assigns to player $-i$ having a positive type, conditional on no entry up to period t. Let $z^\rho_{-i,t}$ be the perceived probability that regular player $-i$ enters in period t conditional on no entry until the end of period $t-1$ and the sequence $\underline{c}$. The indifference condition in round t for i is given by

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

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

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

which is decreasing in $q^\rho_{t,-i}$ and in c_t. Note that underestimation implies that $q^\rho_t < \frac{\theta^*_{t-1}}{1+\theta^*_{t-1}}$ for all $t > 1$, since each player attaches some probability ρ to the event that her opponent was informed in round $t-1$ and would have entered with probability 1 if he was positive.

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 $N(\rho)$ hence $z^\rho_{-i,s} = 0$ for all $s < N(\rho)$. Furthermore, we have, suppressing the index of the player since both stays out and forms symmetric beliefs, that

$$q^\rho_{t+1} = \frac{q^\rho_t(1-\rho)}{q^\rho_t(1-\rho)+(1-q^\rho_t)} < q_t \text{ for all } t < N(\rho)-1$$

Set $c_T = \tau$. To show the proposition we only need to show that there exists some finite

$N(\rho)$ such that

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

solving the above for equality 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 $\tau, \rho > 0$. Since the above q_t sequence converges to 0 as $N(\rho)$ grows, for any $\varepsilon > 0$ there exists $q_{N(\rho)} \leq \varepsilon$. This proves the corollary .

Proof of Corollary 5. The result follows from the fact that while $\theta^*_{i,t=1} = \sqrt{c_1/(1-\rho)}$, in round 2, in the absence of entry in the first round, the entry cut-off is given by $\theta^*_{i,t=2} = \sqrt{c_2/(1-\rho)^2}$. If $c'_1 < c_2/(1-\rho)$ it follows that there is always more entry under $\underline{c}'$ than under $\underline{c}$. This is true because, here, there is no entry by any type in the second round under $\underline{c}'$, and under $\underline{c}$ all types that enter are strictly greater than $\sqrt{c'_1/(1-\rho)}$. In contrast, if $c'_1 \geq c_2/(1-\rho)$ the the set of types that enter are identical under the two scenarios.

To see that $\theta^*_{i,t=2} = \sqrt{c_2/(1-\rho)^2}$, consider the second period entry decision conditional on no entry in the first period. The indifference condition on θ_2, assuming an interior solution, is again given by

$$\rho(q^{\rho}_2\theta_2 - (1-q^{\rho}_2)c_2) + (1-\rho)(z^{\rho}_2\theta_2 + (q^{\rho}_2 - z^{\rho}_2)\gamma\theta_2 - (1-q^{\rho}_2)c_2) = \rho q^{\rho}_2\theta_2\gamma + (1-\rho)(z^{\rho}_2\gamma\theta_2)$$

where again given symmetry I suppressed the notation referring to player i or $-i$. Given the equilibrium cutoff θ_1, which is increasing in c_1, it follows that

$$q^{\rho}_2 = \frac{(1-\rho)(\theta_1+1)/2}{(1-\rho)(\theta_1+1)/2+\rho/2}\frac{\theta_1}{1+\theta_1}$$

and

$$z^{\rho}_2 = q^{\rho}_2\frac{\theta_1-\theta_2}{\theta_1}$$

Solving now for θ_2 and taking $\gamma \to 1$, we get that

$$\theta_2 = \frac{\theta_1}{(1-\rho)^2\,\theta_1}\sqrt{\frac{c_2\,(1-\rho)^2}{(\theta_1(1-\rho)+1)^2}}\,(\theta_1(1-\rho)+1)^2 = \frac{\sqrt{c_2}}{(1-\rho)}$$

.

Proof of Proposition 4. Note first that a super version of i always has a dominant strategy, she enters 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 regular player $-i$ enters given strategy σ^{ρ}_{-i} and the prior distribution on θ_{-i}. For any given type $\theta_i > 0$, the utility difference from entering versus staying out for type θ_i is

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

From $f' < 1$ and $g_1 > 0$, it follows that this difference is strictly increasing in θ_i for any given σ_{-i}. Hence the equilibrium must be in cut-off strategies corresponding to some $(\theta_i^{*,\rho}, \theta_{-i}^{*,\rho})$.

I. Consider the best-response correspondence $\beta(\theta_{-i})$ which determines the unique solution to $E^\rho[u_i(in) \mid \theta_i] - E^\rho[u_i(out) \mid \theta_i] = 0$ given a fixed θ_{-i} describing the cut-off of player $-i$. The solution is unique given the strict monotonicity of f and g in θ_i. Hence there is a best-response function. Furthermore, it is symmetric for the two players. Using the implicit function theorem, evaluated at some $(\widehat{\theta}_i, \widehat{\theta}_{-i})$ on the best-response curve, we obtain that along the best-response function

$$\frac{d\theta_i}{d\theta_{-i}} |_{(\widehat{\theta}_i, \widehat{\theta}_{-i})} =$$
$$\overbrace{(1-\rho)(\theta_i - f(\theta_i) - g(\theta_i, \theta_{-i}) - \int_{\theta_{\min}}^{\theta_{-i}} g_2(\theta_i, \theta_{-i}) d\theta_{-i})}^{I} *$$
$$\overbrace{\begin{array}{c}[\rho[(\Pr(\theta_{-i} \geq 0)(1 - f'(\theta_i)) + \int_{\theta_{\min}}^0 g_1(\theta_i, \theta_{-i}) d\theta_{-i}] + \\ (1-\rho)[\Pr(\theta_{-i} \geq \theta_{-i})(1 - f'(\theta_i)) + \int_{\theta_{\min}}^{\theta_{-i}} g_1(\theta_i, \theta_{-i}) d\theta_{-i}]]^{-1}\end{array}}^{II}$$

Note that term II is always positive. If investments are substitutes, term I is negative, hence $\beta(\theta_{-i})$ is downward sloping. If investments are complements, and $g_2 = 0$, term I is always positive, and hence $\beta(\theta_{-i})$ is upward sloping.

Given a linear investment problem, for all $\rho < 1$

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

where straightforward calculations show that $N \geq 0$. Since the sign of term III is always the same as the sign of term IV, both are positive if investments are complements, and negative if they are substitutes. As long as players enter with positive probability, this second-derivative is always positive in the relevant domain.

II. Consider substitute investments. A symmetric equilibrium must exists due to the fact that $\beta(\theta_{-i})$ and $\beta(\theta_i)$ are continuous and strictly downward-sloping with $\beta(0) \leq \theta_{\max}$ and $\beta(\theta_{\max}) \leq \theta_{\max}$, as long as there is entry, and are each-other's mirror images on the 45 degree line. Hence, given the intermediate value theorem, they must have an intersection with the 45 degree line. Furthermore, given that $\beta(\theta_{-i})$ is strictly decreasing, it follows that the symmetric equilibrium is unique.

III. Consider complement investments. Here all equilibria must be symmetric since $\beta(\theta_{-i})$ is continuous and strictly increasing, hence, it cannot be the case that for a given $(\theta_i^{*,\rho}, \theta_{-i}^{*,\rho})$, $\theta_i^{*,\rho} = \beta(\theta_{-i}^{*,\rho}) > \beta(\theta_i^{*,\rho}) = \theta_{-i}^{*,\rho}$. Furthermore, since $\beta(\theta_{\max}) \leq \theta_{\max}$, as long as there is entry, from the fact that $\beta(0) \leq \theta_{\max}$, it follows that $\beta(\theta_{-i})$ intersects the 45 degree line. Since $\beta(\theta_{-i})$ is increasing, if it is convex, it also follows that $\beta(\theta_{-i})$ can cross the 45 degree line at most two points in the relevant domain.

IV. Consider the comparative static with respect to ρ. We can re-write the equilibrium cut-off condition for $\theta_i^{*,\rho}$ as the one value of θ_i solving

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

IV.a. Consider substitute investments. Here Term V is negative. Hence, holding all else fixed, LHS is decreasing in ρ. Note that the LHS of the equation is also increasing in θ_i. This holds because $1-f' > 0$ for all $\theta_i > 0$ and $(1-\rho)\int_0^{\widehat{\theta}_{-i}} g_1(\theta_i, \theta_{-i})d\theta_{-i} + \int_{\theta_{\min}}^{0} g_1(\theta_i, \theta_{-i})d\theta_{-i} > 0$. Thus, holding $\theta_{-i}^{*,\rho}$ constant, the LHS is increasing in θ_i. An increase in ρ shifts the best-response function up and the symmetric equilibrium increases in ρ.

IV.b. Consider now complement investments. Here Term V is positive. Hence, holding all other terms fixed, the LHS is increasing in ρ. The LHS of the equation is again increasing in θ_i. This implies that the best-response function shifts down. At the same time, note that $\beta(0) > 0$ and is independent of ρ. Since $\beta(\theta_{-i})$ is an increasing function, given a linear game, it follows that $\beta(\theta_{\max})$ is decreasing in ρ. Hence if $\beta(\theta_{\max}) \leq \theta_{\max}$ for$\rho = 0$, the same holds for all $\rho > 0$. Thus, an equilibrium with entry exists for $\rho > 0$ if it existed for $\rho = 0$, and the lowest equilibrium cut-off is decreasing in ρ.

Underestimation. This claim follows directly from the fact that the cut-off of regular i is always higher than that of super i in equilibrium. Hence, all else fixed, a non-entering positive player $-i$'s posterior condition on entry of her opponent is lower than in the Bayesian case. Similarly, all else fixed, a non-entering positive player i's posterior conditional on non-entry by her opponent is also lower than in the Bayesian case.

False antagonism. Consider a positive type $\theta_i > 0$ who does not enter. This type always exaggerates the probability of entry by player $-i$ relative to the true probability with which player $-i$ stays out. 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, i.e., $E_\pi[E^\rho[(\theta_{-i} \mid \theta_i > 0)]] < E_\pi[\theta_{-i}]$ Consider a negative type $\theta_i < 0$, this type exaggerates the probability that her opponent stays out relative to the true probability. Hence she over-infers from entry and under-infers from out, and thus exaggerates the value of her opponent's type on average, i.e., $E_\pi[E^\rho[(\theta_{-i} \mid \theta_i > 0)]] > E_\pi[\theta_{-i}]$.

Proof of Proposition 5. 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 of Proposition 6. Note if $\rho > 0$, a pure-strategy equilibrium need no longer exist. To see this, let $p^+(c) : \mathbb{R}^+ \to [0, 1]$ - the probability which which the fictional informed sender lies given c. 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 $p^+(c)$ must strictly increase on some interval $[c_1^\rho, c_3^\rho]$ and be surjective. Hence for any p^ρ there exists $c \in [c_1^\rho, c_3^\rho]$ such that $c = p^\rho/(1 + p^\rho)^2$. Let this be c_2^ρ. Here, and only here, $p^+(c_2^\rho) = p^\rho$. 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, as long as the seller does not always lie, that the total mass of types who check over the types who do not check 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 must lead to an increase in both c_1^ρ and c_3^ρ .

Proof of Proposition 7. 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, $c > c_{\max} = 0.25$, 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 on ρ^* follow. Finally, since c_1^ρ and c_3^ρ are 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 of Proposition 8. 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 here 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] Andersen, H. C. (2001), *Fairy Tales from Andersen* Oxford University Press, Oxford.

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

[6] Ball, Sheryl, Max Bazerman, and John S. Carroll. (1991). "An Evaluation of Learning in the Bilateral Winner's Curse." *Organizational Behavior and Human Decision Processes*, 48:1-22.

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

[8] Bénabou, Roland. (2013). "Groupthink: Collective Delusions in Organizations and Markets." *Review of Economic Studies*, 80(2), 429-462.

[9] Biais, Bruno and Martin Weber. (2010). "Hindsight bias and Investment Performance." *Management Science,* 55: 1018-1029.

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

[11] 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.

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

[13] Danz, David, Kristóf Madarász, and Stephanie Wang. (2014). "The Biases of Others: Anticipating Informational Projection in an Agency Setting." mimeo LSE and U of Pittsburgh.

[14] 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.

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

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

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

[18] 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.

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

[20] Kartik, Navin, Marco Ottaviani, and Francesco Squintani. (2007). "Credulity, lies, and costly talk." *Journal of Economic Theory*, 134, 93–116.

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

[22] Kuran, Timut (1991) "Now Out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics*, 44(1), 7-48.

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

[24] Fudenberg, Drew and Alex Peysakhovich (2013). "Recency, Records and Recaps: Learning and Non-Equilibrium Behavior in a Simple Decision Problem." mimeo Harvard.

[25] 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*, 75(2): 332-46.

[26] Gilovich, Thomas, Victoria Medvec and Kenneth Savitsky. (2000). "The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own

actions and appearance." *Journal of Personality and Social Psychology*, 78, 211–222.

[27] 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.

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

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

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

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

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

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

[34] 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.

[35] Madarász, Kristóf. (2012). "Information Projection: Model and Applications." *Review of Economic Studies*, 79: 961–985.

[36] Madarász, Kristóf. (2014a). "Projection Equilibrium: Definition and Applications to Social Investment and Persuasion." *Working Paper, LSE.* http://works.bepress.com/kristof_madarasz/35

[37] Madarász, Kristóf. (2014b). "Bargaining under the Illusion of Transparency." *CEPR* Discussion Paper.

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

[39] Malmendier, Ulrike, and Devin Shanthikumar. (2014). "Do security analysts speak in two tongues?" *Review of Financial Economics,* forthcoming.

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

[41] Miller, Dale and Prentice, Deborah (1994). “Collective errors and errors about the collective.” *Personality and Social Psychology Bulletin*, 82, 213-225.

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

[43] Piaget, Jean and Bärbel Inhelder. (1948). *The Child’s Conception of Space.* Translated (1956). London: Routledge and Kegan Paul.

[44] 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.

[45] Prentice, Deborah. (2007). “Pluralistic ignorance.” *In Encyclopedia of Social Psychology.* Sage.

[46] Riley, John. (1989). “Expected Revenue from Open and Sealed Bid Auctions.” *Journal of Economic Perspectives*, 3(3): 41-50.

[47] 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.

[48] 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.

[49] Shelton, Nicole and Richeson, Jennifer. (2005). “Intergroup Contact and Pluralistic Ignorance.” *Journal of Personality and Social Psychology*, 88(1): 91-107.

[50] Williamson, Oliver. (1979). “Transaction-cost economics: the governance of contractual relations.” J*ournal of Law and Economics*, 22(2): 233-261.

[51] 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.