2014

Projection Equilibrium: Definition and Applications to Social Investment and Persuasion (old version)

Kristof Madarasz, *London School of Economics and Political Science*

Projection Equilibrium: Definition and Applications to Social Investment and Persuasion

Kristóf Madarász (LSE)

First Circulated Version: May 2013. This Version: January 2015 (minor revision 10/15)[1]

[1]Would like to thank audiences at Arizona, Bonn, Columbia, Harvard, Princeton, Stockholm, UCL, UCLA, UC San Diego, Yale, Wash U, LSE, Berlin Behavioral Seminar 2011, CEU 2013, European Behavioral Economics Meeting Berlin 2013, ESSET Gerzensee 2013, Stanford Institute for Theoretical Economics 2015, as well as, Pedro Bordalo, Peter Bossaerts, Colin Camerer, Jeff Ely, Ignacio Esponda, Erik Eyster, Marina Halac, Philippe Jehiel, Xavier Gabaix, Navin Kartik, George Loewenstein, Marek Pycia, Drazen Prelec, Matthew Rabin, Luis Rayo, Adam Szeidl, Balazs Szentes, Andrei Shleifer, Tomasz Strzalecki, Jörgen Weibull for comments. All errors are mine. Contact: k.p.madarasz@lse.ac.uk, or London School of Economics, Houghton Street, London, WC2A 2AE UK.

Abstract

People exaggerate the extent to which their information is shared with others. I incorporate such information projections into the solution of Bayesian games, whereby people wrongly think that if they can condition their strategy on an event, others can as well. In the context of investments into a social asset, people misattribute the lack of trust by others due to differences in information to others having antagonistic preferences. Even if everyone prefers mutual investment, no one invests, but each comes to believe that none else values investment. In the context of communication, the model predicts credulity: persuasion by an advisor with a known incentive to exaggerate the truth, nevertheless, induces uniformly inflated expected posteriors. Credulity results when receivers have sufficient, but limited, financial education and the conflict is limited. An increase in the former, as well as, a decrease in the latter, can systematically lower receiver welfare. I extend the model to incorporate ignorance projection and apply it to common-value trade. The model predicts that sellers engage in too much truth-telling and buyers underappreciate selection, and provides a better fit of the data than BNE or cursed equilibrium.

Perspective Taking, Norms, Distrust, Financial Advice, Credulity, Trade, Under-Bluffing

1 Introduction

People fail to appreciate the full extent to which others act using different information than they do. While the typical assumption in economics is that people perceive informational differences in an unbiased manner, on average, the evidence shows that the typical person misperceives these differences, in that, she too often acts as if others had access to the information she did. Such informational projections – empathy gaps in informational perspective taking – may shape strategic interactions in key economic settings analyzed via Bayesian games. The goal of this paper is then to incorporate such informational projections into the solution of Bayesian games.

Direct evidence for such information projection comes from a variety of domains. It dates back to the work of Jean Piaget, e.g., Piaget and Inhelder (1948). His work initiated a literature pointing to egocentric biases in people's 'theory of mind', that is, their insufficient tendency to attribute different *beliefs* to others than what they themselves hold. In a now classic study, Wimmer and Perner (1983) demonstrate that young children too often act as if lesser-informed others shared their superior information. Birch and Bloom (2007) showed that the same kind of mistake is present among Yale undergraduates in slightly more complex tasks.[1] Such robust phenomena, as the *curse of knowledge*, (Camerer et al., 1989; Newton, 1990; and Krueger et al., 2005)*;* the *hindsight bias (*Fischhoff, 1975), the *outcome bias* (Baron and Hershey, 1988), and the *illusion of transparency*, or the *spotlight effect,* (Gilovich et al., 1998, 2000) are all consistent with the idea that people act as if they exaggerated the probability with which others knew their private information.[2] Madarász (2012) offers a more extensive review of the evidence and introduces a notion of information projection into monotone inference problems. In a strategic context, Samuelson and Bazerman (1985) study behavior in common-value bilateral trade and find that both sellers and buyers act as if they ignored the informational asymmetry that existed between them.

A key issue when introducing such biased beliefs into strategic settings is that here higher-order perceptions, that is, each person's theory of others theory of her, may matter. Accounting for such considerations, I offer a parsimonious, but fully specified and portable model allowing one to study its strategic implications both empirically and theoretically. To illustrate these, I consider applications to problems of social investment, persuasion, and common-value trade.

Model Section 2 presents the model of information projection equilibrium. I consider a Bayesian games with partitional information where people receive different information about the state. A person who projects information misperceives her opponent's strategy set and has

[1] These studies are often referred to '*false belief tasks',* see also Baron Cohen et al. (1985) linking it to autism. The phenomenon is also discussed in the context of *mentalization*, with the idea that teaching perspective-taking is the fundamental common feature among the many versions of adult psychoteraphy, see, e.g., Allen, Fonagy, Baterman (2008).

[2] For an overview see, Epley et al. (2004).

an exaggerated belief that if she can condition her behavior on the knowledge of an event, so can her opponent. The extent of this false belief is characterized by the parameter $\rho \in [0, 1)$.

As mentioned, incorporating information projection into Bayesian games, one needs to note that when players form such biased forecasts of the behavior of others, one needs to specify not only how each player thinks about the information of her opponent, but also how she thinks the opponent would behave based on that information, which in turn, depends on a player's view of her opponent's view of herself. To model projection in a parsimonious manner, I distinguish between the real and the projected versions of a player. The real version of a player conditions his strategy on his true information. The projected (super) version of a player, who is real only in his opponent's imagination, instead conditions his strategy on his and this opponent's joint information. While in reality only real versions exist, if a player projects to degree ρ, she believes that her opponent is such a projected super version as opposed to the real version with probability ρ.

Two properties characterize the model. First, in equilibrium, projection is all encompassing: a player assigns probability ρ to her opponent being the projected version who knows everything she does, including the fact that she is her real version playing her real strategy. If the true game is poker where, in reality, each player only knows the value of his or her own card, a biased Judith believes that with probability ρ Paul is super and knows the value of her card and, when he does so, he also knows the true fact that she does not know the value of his card. Projected Paul is then believed to best-respond to Judith's real strategy given such information. Second, in equilibrium, a player's belief about how her opponent might behave is consistent with how this opponent actually behaves. Each player assigns probability $1-\rho$ to her opponent being the real version, and thus, to her opponent's true real strategy. Despite players being biased, nothing happens in equilibrium that would be inconsistent with what players think might happen.

These two properties imply that the predicted behavior is consistent with an interpretation whereby people act as if they *partially* anticipated the biases of others. Due to consistency, each player acts as if she anticipated that her opponent was biased and projected onto her. Due to all-encompassing projection, exactly proportional to the extent that she herself projects, a player underestimates the true extent of this. All such differences in higher-order perceptions are solely a function of the degree of projection, allowing the model to provide a tight characterization. After presenting the model, I establish existence and present some basic properties.

Social Investment In Section 3, I apply the model to the problem of social investment. Partnerships in trade, friendships, or the formation of political associations, require people to pool resources and invest into a jointly owned social assets. Such investments are typically risky because people face uncertainty regarding others' preferences. Investing with someone who has matching goals and would reciprocate investment (a positive type) is a source of gain. Investing with someone who is opportunistic and would not never recipricate investment is a

source of loss. A key determinant in this setting is trust: the belief that one's partner is the former as opposed to the latter type.

To illustrate, consider a dating example. Two people are sitting at a bar. Each is privately informed about his or her own value of a match. Each can independently decide to make a move (invest) or not make a move (not invest). If neither makes a move, they get their outside options. If both invest, a match is formed. If only one of them makes a move, the other accepts if interested, and rejects otherwise. In case of a rejection, the proposer incurs a loss such as pain incurred when being rejected. Investment is risky because neither player knows whether the other is interested or not. A similar problem arises when a buyer and a seller need to decide whether or not to invest into a relationship-specific asset, not knowing whether the other party is reciprocal or opportunistic, or when a person needs to decide whether to voice his dissent or stay silent in front of a co-worker, not knowing whether his peer also disagrees with the prevailing norm (status-quo), or instead is instead loyal and want to punish or block deviations.

By projecting information, a person fails to appreciate the extent to which another person faces the same kind of uncertainty about her preferences as she does about his preferences. In the dating context, an interested Judith too often thinks that Paul should know that she is interested. As a result, she comes to exaggerate how often Paul should make a move if he is interested. Since Judith still does not know Paul's preferences, she now finds it relatively more important to protect herself from the loss in case Paul were to reject her. At the same time, an interested Paul, in a symmetric situation, reasons similarly. Therefore, neither invests, but both conclude that the other is not interested. Even if they have the same preferences and behave identically, they will *always* conclude that the other party has antagonistic preferences.

In particular, two mistakes always arises in equilibrium. First, conditional on *any* outcome in the game, a positive type underestimates her opponent's preference for mutual investment. Second, players develop false antagonism on average: a positive type on average concludes that her opponent is negative, and a negative type concludes on average that her opponent is positive. On average, a person who opposes a norm will come to conclude that her neighbor supports the norm, and a person who is loyal to the norm will come to be too suspicious that her neighbor is against the norm. Social investments under projection leads to a ex ante predictable false negative correlation between the direction of own's preferences and the preferences of others.

As a corollary, in a setting with repeated encounters, I specify environments in which even when continued interaction leads to fully efficient matching under Bayesian assumptions, it leads to no investment, given any positive degree of repeated information projection. Even if all players are positive and value mutual investment, none invests, but all come to believe that everyone else is opportunistic almost surely. Such false uniqueness is the consequence of the above differential attribution of identical behavior to oneself and to others. I discuss comparative static implications in the context of various applications and relate the predictions

to evidence described in psychology under the rubric of 'pluralistic ignorance', e.g., Prentice (2007).

I conclude by generalizing the setting and show that underestimation and false antagonism remain true both if initial investments are substitutes or complements relative to reciprocated investments. Hence, to the extent that player's valuation of a social asset increases in her perception of how much her opponent values mutual investment, the model predicts too little trust and a general undervaluation of social assets.

Persuasion In Section 4, I apply the model to a simple problem of strategic communication. Bayesian communication under unbiased expectations, has two general properties: by providing information it improves the expected welfare of receivers, and it is purely informative in that a receiver is never fooled, on average. In an environment with a commonly known conflict of interest and distribution of the payoff relevant state, the model, nevertheless, implies a systematic violation of both of these properties.

A sophisticated advisor sends a cheap talk message to a biased receiver about a statement being true or false. A doctor advises a patient, a broker and investor about the suitability of a drug or a financial product for the receiver. The advisor's preferences are misaligned towards claiming that the statement is true. Receivers have private information about the cost at which they can verify the advisor's recommendation. Differences in such costs might reflect private information about one's financial expertise or cost of accessing additional sources to evaluate the sender's advice. If the receiver decides to verify, he learns whether the sender lied or not, and the sender suffers a loss (of business) in case she did.

A biased receiver who projects information exaggerates the probability with which the sender knows the receiver's actual cost of verification. In turn, in equilibrium the receiver exaggerates the extent to which the sender's incentive to lie is tailored to his privately known cost as opposed to the publicly known distribution thereof. In equilibrium this will imply, that the lower is the receiver's cost of verification, the more confident he will be that a positive recommendation by the sender is truthful.

While in equilibrium receiver types with the highest financial expertise always check and learn the truth, types with sufficient but not the highest expertise will always be overconfident and overinvest in the asset. Types with little or no expertise on the matter will be (weakly) in disbelief and may underinvest in the asset. I show, however, that given any degree of projection, as long as the conflict between the parties is sufficiently high, or the asset is sufficiently complex to evaluate, the model, nevertheless, predicts *uniform credulity*: all receiver types are always too optimistic when hearing the advisor's positive recommendation. The act of persuasion itself predictably inflates ex ante expected posteriors of all receiver types and leads to overinvestment.

Understanding the mechanism through which such credulous investments arise is potentially key for evaluating such commonly advocated policies as capping the conflict between a financial advisor and an investor or improving the financial education of investors. In fact, the

model predicts that it is the very presence of limited conflict and sufficient financial education which drives financial advice towards credulity. Key comparative static predictions establish sufficient conditions under which decreasing verification costs (improving financial education), or decreasing the conflict, will strictly lower receiver welfare. I conclude by partially endogenizing the conflict and the complexity of the asset by invoking the seller of the asset. I show that any partial cap on the conflict may have very limited effectiveness: given any degree of projection, the receiver may have a strictly positive willingness to pay for advice ex ante which reduces his welfare. In fact, the seller-optimal way to induce uniform credulity will minimize the conflict and ensure that the asset is not too difficult to evaluate.

Projection equilibrium Section 5 combines information projection – the underappreciation of the positive side of the information gap – with a notion of *ignorance projection* – the underappreciation of the negative side of the information gap whereby a person acts as if she also exaggerated the probability with which if she does not know an event her opponent does not know it either. Here, a player projects both what she knows and what she does not know, exaggerating the probability that her opponent conditions his strategy on exactly the same set of events as she does. I derive implications of the combined model of projection equilibrium for the classic problem of trade with common values, as in Akerlof (1970). Consistent with the evidence – e.g., Samuelson and Bazerman (1985) and Holt and Sherman (1994) – when a privately informed seller has all the bargaining power, the model 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, the model 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 with the evidence and show that a very small degree of projection robustly provides a better fit with the data than BNE or cursed equilibrium.

2 Model

This section develops the model. 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. Player i's information about ω is given by a standard information partition $P_i : \Omega \to 2^{\Omega}$; her finite action set is A_i; and her bounded payoff is $u_i(a, \omega) : A \times \Omega \to \mathbb{R}$, where the action profile $a \in A = \times_i A_i$. In short, the game is summarized by the tuple $\Gamma = \{\Omega, \pi, P_i, A_i, u_i\}$.[3]

To introduce information projection, I express the joint information of the two players i and j. Specifically, consider the following correspondence:

[3] The setup immediately extends to the case in which the set of available actions depends on the state, that is, $A_i(\omega) \neq A_i(\omega')$ for some $P_i(\omega) \neq P_i(\omega')$.

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

This correspondence is also partitional and describes the coarsest common refinement of the two players' partitions – that is, the information distributed between these two players. Note that this partition is the unique one to capture the players' joint information.[4] If an event, $E \subseteq \Omega$, is known at a state ω by either of the players, this event is also known at that state under P^+. Conversely, any event that is known in a given state under P^+ is known given the pooled information of the two players. This joint information, thus, corresponds to the natural object to capture the idea of information projection; it will imply that a person who projects her private information has an exaggerated belief that whenever she can condition her strategy on an event, so can her opponent.

To incorporate such information projection into games in a parsimonious manner, I distinguish between two versions of each player i. The real *regular* version of i conditions her strategy on her true information. Formally, she chooses a strategy from the set

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

The fictional *projected* (super) version of player i – who is real only in the imagination of player j – conditions her strategy on the joint information of i and j. 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. Fictional, projected versions enter only into people's beliefs about each other.[5]

Information projection by a player j corresponds to a mistaken belief that, with probability $\rho \in [0, 1)$, player i is a projected, as opposed to a regular, version. For ease of notation below, I first assume the degree of projection to be common across players, but then immediately extend the definition to heterogeneous projection.

Notation Below, the operator $\circ$ denotes the mixture of two probability-weighted lotteries. The operator BR denotes the standard best-response operator; its subscript always refers to the set of strategies over which the there indexed player maximizes her expected utility; its

[4] Note that the unique knowledge operator $K : 2^\Omega \to 2^\Omega$ associated with partition $P_i(\omega)$, is $K_i(E) = \{\omega \mid P(\omega) \subseteq E\}$ describing the set of events know at a given state ω for each state. The knowledge operator corresponding to the joint information P^+ is also uniquely defined by $K^+(E) = \{\omega \mid \cap_i P_i(\omega) \subseteq E\}$.

[5] Note that, since all payoff relevant facts are encoded in ω, to the extent that player i has information about her own taste, or the taste of her opponent, she projects her private *information* about preferences as well. For example, in a common-value environment with positively affiliated valuations, this may imply exaggerating the correlation between one's own conditional valuation and the conditional valuation of one's opponent. See Section 5 for further discussion.

argument refers to this player's belief about her opponent's strategy to which she wishes to best respond.

2.1 Definition

Below, σ^ρ describes the predicted strategy profile of the players – the strategy profile of the real, regular versions. Since, in reality, people can condition their true strategies only on the information that they truly have, this strategy profile is an element of the true strategy space. It is supported by a profile σ^+ describing the conjectured behavior of the projected versions.

Definition 1 *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.

$$\sigma_i^\rho \in BR_{S_i}\left\{(1-\rho)\sigma_{-i}^\rho \circ \rho\sigma_{-i}^+\right\} \tag{2}$$

2.

$$\sigma_{-i}^+ \in BR_{S_{-i}^+}\{\sigma_i^\rho\} \tag{3}$$

If $\rho = 0$, each player has correct forecasts about the behavior of the opponent, and the predictions of IPE collapse to that of the BNE for Γ. If $\rho > 0$, each real player mistakenly assigns probability $(1-\rho)$ to the actual strategy of her opponent and probability ρ to the strategy of her projected opponent. Here, each player has potentially mistaken forecasts and assigns probability ρ to her opponent's playing a strategy which is also conditioned on her own private information and is a best response to her true strategy. I now describe the two defining features of the model.

All-encompassing Projection First, projection is all-encompassing: the real player i believes that her projected opponent knows that she is regular for sure. This is reflected in Eq. (3). In other words, a biased player believes that her projected opponent knows everything she knows. If the true game is poker, in which each player truly sees only his/her own card, by projecting information, Judith believes that with probability ρ Paul knows both the value of her card and the fact that she does not know the value of his card. Since a player always knows what she herself knows, this feature implies, consistent with the spirit, that projection is not based on an arbitrary distinction between the content of a person's private information and her information about what she knows, but applies to both of these equally.

Consistency Second, each player's expectation about her opponent's play is consistent with how her opponent actually plays. This is reflected in Eq.(2). Each regular player assigns probability $1-\rho$ to her opponent's behaving in the way that this opponent always behaves. Thus, in equilibrium, nothing happens that explicitly contradicts a player's theory of how her opponent may behave. This remains true even if players observe joint payoffs. The deviation

from BNE is simply that, despite potential evidence to the contrary, based on her egocentric projection, Judith nevertheless expects something to happen that may never happen or may happen with a different probability than expected.

Heterogeneity The definition extends immediately to differentially biased players. Heterogenous projection is described by a vector ρ with a potentially different ρ_i replacing ρ in Eq. (2) for each i. If $\rho_i = 0$, then player i is unbiased. Given the consistency property, an unbiased player is fully sophisticated and has correct forecasts about her opponent's strategy given her information.

Partial Anticipation The above two properties directly imply that the predicted behavior is consistent with an interpretation whereby players partially anticipate the biases of others. Each player anticipates that her opponent projects onto her, but exactly proportional to that extent that she herself projects, she underestimates the extent of this. Given all-encompassing projection, Judith believes that, with probability ρ, Paul has correct beliefs about her strategy. Given consistency, Judith believes that, with probability $1-\rho$, Paul wrongly believes that she knows the value of his card with probability ρ. In sum, Judith expects on average Paul to believe that Judith knows the value of his card with probability $\rho - \rho^2$. Instead, Paul believes that Judith knows the value of his card with probability ρ.[6]

Under heterogeneous projection, Judith – player j – acts as if she expected Paul – player i – to assign probability $(1-\rho_j)\rho_i$ to Judith being the projected super version on average. Her underestimation is, thus, proportional to the degree of her own mistake ρ_j. If $\rho_j = 0$, Judith is unbiased, which then implies that she fully anticipates Paul's misperception.

Evidence In a companion paper, Danz, Madarász, and Wang (2014) directly test the above tight partial anticipation aspect of the model of projection. They find not only that people project onto others, consistent with earlier evidence, but also that people partially anticipate the projection by others onto them. By considering heterogenous projection, their design allows one to measure both jointly and separately the extent to which people project and the extent to which people underappreciate the projection of others. Consistent with the logic of the model, they find that the degree to which subjects project and the degree to which subjects

[6] In a similar fashion to ones described above, one can construct the real players' iterative higher-order perceptions about, say, player j being a super version. Here, the same pattern will hold. Define the higher-order iterative perceptions of real players i and j about player j's being her projected super version inductively, as follows: the first-order belief is the probability that player i assigns to j being super; the second-order belief is the probability that real player j assigns to the *expected* probability that player i assigns to player j being super, etc. The k^{th} element of this sequence is always given by $\sum_{s=1}^{k}(-1)^{s+1}\rho^s$. For example, when $k=3$, the real Paul believes that in expectations Judith expects Paul to expect Judith to be super with probability $\rho - \rho^2 + \rho^3$. He, thus underestimates Judith's underestimation. In this sequence, (i) the sub-sequence of odd elements (containing those elements where k is an odd number) is decreasing in k, (ii) the sub-sequence of even elements is instead increasing in k. In words, real Judith's assessment is increasing and real Paul's is decreasing in k. Furthermore, (iii) each odd element is larger than the subsequent even element, and (iv) both the even and the odd subsequences converge to $\rho/(1+\rho)$. Hence, for all higher orders of 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.

underestimate the projection of others are remarkably similar. [7]

2.2 Discussion

Let me turn to some of the basic properties of the model. The first claim establishes existence.

Proposition 1 *For any Γ and ρ, a $\rho - IPE$ exists.*

The next corollary points out that the model delivers differential predictions only to the extent that players are differentially informed.

Corollary 1 *If $P_i(\omega) = P_j(\omega)$ for all ω, the set of $\rho-IPE$ for Γ is independent of ρ.*

While in games with symmetric information projection has no bite, it affect predictions already in games with one-sided private information, ones where, say, P_j strictly refines P_i. Furthermore, already there, given all-encompassing projection, the degree to which each player projects matters. This is true because even under heterogenous projection, the lesser-informed player's degree of projection ρ_i governs her anticipation of the projection of her opponent. The next claim shows that any BNE of Γ which is an ex post equilibrium – an equilibrium where no player has an individual incentive to deviate even after observing the state – is also an information projection equilibrium. In contrast to a $\rho-$IPE, an ex post equilibrium often does not exist, but when it does, projecting information does not affect such predictions.

Proposition 2 *If a BNE is an ex post equilibrium, then it is also a $\rho-IPE$ for all ρ.*

As an example, bidding one's own valuation in a standard second-price auction is an ex-post equilibrium, hence, such a prediction is projection proof. A converse of the above claim is not true. Even if all BNE of a game are ex-post equilibria, IPE can extend the set of predictions and lead, for example, to illusory coordination.[8]

Related Literature This paper relates to other approaches that study players who exhibit an explicitly wrong theory of others' behavior. In particular, Jehiel (2005) and Jehiel and

[7] Technically, Danz et al. (2014) can only estimate the model of projection equilibrium as introduced in Section 5. As will be described, the structure of higher-order perceptions, and hence the qualitative nature of partial anticipation is unchanged. In their setting, the homogenous joint and the heterogeneous role-specific estimates all amount to an estimated $\rho \in [0.27, 0.29]$. Furthermore, the exploratory power of the homogenous specification and the one allowing for role-specific projections are also very close.

[8] To illustrate, consider the following game with a symmetric prior and the state being the column player's private information:

ω_1	R	L		ω_2	R	L
T	$1,1$	$0,0$	,	T	$1,1$	$0,0$
B	$-3,3$	$-3,3$		B	$0,0$	$2,2$

The unique BNE is $\{T, R(\omega_1), R(\omega_2)\}$. In contrast, if ρ is sufficiently high, there is a $\rho-$IPE where the row player plays T and the column player plays R in state ω_1 and L in state ω_2.

Koessler (2008) study a general framework of analogy-based expectations equilibrium. Eyster and Rabin (2005) study the notion of cursedness. The identifying assumption in all of these approaches is that a player has coarse but correct expectations about the strategy of her opponent, on average Instead, the key assumption here is that a player has wrong expectations about her opponent's strategy, on average. The model, thus, systematically violates the identifying assumption of these approaches. Since information projection applies through the misperception of the opponent's strategy set, it also clearly differs from level-k approaches in Bayesian games – e.g., Crawford and Irriberi (2008) – because these maintain the assumption that people have a correct understanding of informational differences.

Note that the logic of information projection also differs markedly from the logic of cursedness. A cursed player underappreciates the extent to which her opponent conditions his choice on his own private information. Specifically, a cursed player has correct expectations about the distribution of her opponent's type, conditional on her own type, but expects that with some probability χ her opponent plays the same strategy independent of his realized 'type'. Instead, information projection points to an exaggeration of the extent to which a player thinks that her opponent conditions his choices also on her information. Furthermore, while she forms wrong beliefs about the information of others, she, nevertheless, forms fully coherent beliefs about the strategy of her opponent, given her structured misperception. In Section 5, when incorporating ignorance projection into the model as well, I return to the link between ignorance projection and cursedness.

Finally, the above model describing the psychological wedge in players beliefs of perceived informational differences versus the true informational differences is consistent with multiple different ways that such a wedge in interim beliefs may be generated. For example, $\rho - IPE$ could be described as a heterogenous prior BNE where initially each player assigned zero probability to herself becoming a projected (super) version, but assigned probability ρ to her opponent becoming a projected (super) version. The model then specifies a clear structural restriction in the way such priors might be heterogenous as a function of the true data generating process.[9]

3 Social investment

I now apply the model to the problem of social investment. Efficient outcomes between trading partners, Williamson (1979), friendships, the formation of a political or social associations, typically require partners to pool resources and invest into a joint social asset. The return on one's investment into such an asset depends on how much one's partner also values investing into

[9] An earlier extended version of this paper, Madarasz (2014), also formulates a model of private information projection equilibrium describing a more naive approach, perturbing only first-order beliefs relative to an underlying BNE, hence, describing a more radical departure from the standard model. Here, people fully fail to anticipate the projection of others and form potentially fully misspecified forecasts of the behavior of others. It also considers applications of this alternative approach to zero-sum games and auctions.

this asset. Investing with someone who has matching goals and would reciprocate investment is a source of gain. Investing with someone who is opportunistic or does not value the asset, hence would not never himself invest, is a source of loss. Investment is therefore risky whenever people face uncertainty regarding the goals and preferences of others. A key component of such interactions is then trust: the belief that people form about the goals and preferences of others.

For example, in the context of trade, when contracts are incomplete, trust is a key determinant of efficient exchange. As Arrow (1972) argued, "virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence." Trust between people is also key for production in large organizations, e.g., La Porta et al. (1997), Algan and Cahuc (2010). Similarly, many social and political outcomes – such as change from a norm or status quo, or intergroup conflict – are dependent on people's perception of others preferences - whether they would support change or whether they prefer integration over segregation.

3.1 Setup

Consider a social investment problem. Each player i has a type θ_i corresponding to her own valuation of mutual investment. After each player observes her valuation privately, players decide independently whether to invest (enter) or not (stay out). If both invest, each realizes her valuation. If both stay out, each gets an outside option normalized to zero. The rest of the game is described as follows:

$$\begin{array}{cccc} & \text{In} & & \text{Out} \\ \text{In} & \theta_1 \ , \ \theta_2 & & g(\theta_1, \theta_2) \ , \ f(\theta_2) \\ \text{Out} & f(\theta_1) \ , \ g(\theta_2, \theta_1) & & 0 \ , \ 0, \end{array} \tag{4}$$

where each θ_i is i.i.d. with a uniform density on $[\theta_{\min}, \theta_{\max}]$. I impose no restrictions on these values, except that $\theta_{\min} < 0 < \theta_{\max}$.

The key distinction below will be between positive and negative types. I make two substantive assumptions. First, a positive type is reciprocal and prefers mutual investment to her opponent investing alone. A negative type is opportunistic and has the reverse preference; conditional on her opponent investing, she prefers not to invest. Specifically:

1. Sorting Let $f(0) = 0$, and $f_1 < 1$.[10]

Second, I assume that investment is risky: if at least one of the players is a negative type (opportunistic), one-sided investment leads to a loss for the investing player relative to the outside option.

[10] If $\theta_i < 0$, $f(\theta_i) > \theta_i$ is sufficient.

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

Finally, I also impose monotonicity assumptions on g, in particular, $g(0, \theta_{-i}) = 0$ and $g_1 \geq 0$ if $\theta_i > 0$, and $g_2 \geq 0$. Two remarks are in order.

▼In the above normal form, each positive type prefers mutual investment to the outside option, and each negative type has the reverse preference. This can be relaxed. Specifically, suppose that each player – independent of her own type and action – receives a benefit b whenever her opponent invests.[11] For example, if $b > |\theta_{\min}|$, mutual investment now Pareto dominates the outside option for *any* type profile. Here, if players were always negative types, the game would be a Prisoner's Dilemma with a dominant strategy outcome of mutual Out and a social optimum of mutual In. If players were always positive types, the dominant strategy outcome would be mutual In. The analysis below holds for any $b \geq 0$.

▼ In the above normal form, a player's payoff may depend on her opponent's type. Specifications of this can be reinterpreted by considering a sequential game in which all payoffs depend only a player's own type and the action profile. Specifically, assume that players first play the above game, leading to the same payoffs in all cases, except when only one of the players, say i, invests. Now $-i$ can decide whether or not to reciprocate investment, $-i$'s payoff still being $f(\theta_{-i})$. If a positive $-i$ always reciprocates, and a negative one never does, the assumptions on $g(\theta_i, \theta_{-i})$ can be satisfied by virtue of player $-i$'s second action. The specification employed in the main example below allows for this interpretation, hence, I invoke it.

3.2 Main Example

The following simple specification helps highlight the main results and intuitions. Suppose that

$$\begin{array}{cccccc} \theta_i, \theta_{-i} \geq 0 & \text{In} & \text{Out} & \text{else} & \text{In} & \text{Out} \\ \text{In} & \theta_1, \theta_2 & \gamma\theta_1, \gamma\theta_2 & \text{In} & \theta_1, \theta_2 & -c,\ f(\theta_2) \\ \text{Out} & \gamma\theta_1, \gamma\theta_2 & 0,\ 0 & \text{Out} & f(\theta_1), -c & 0,\ 0 \end{array} \tag{5}$$

where $\gamma \to 1$, $c > 0$.[12] The following examples describe some applications.

$\diamondsuit$ **At the Bar**. ($b = 0$). Judith and Paul are sitting at a bar. Player i values a match with the other at θ_i. Each player decides to make a move (in) or not (out). If both make a move, a match is formed. If both stay out, each gets the outside option. If only one player, say Judith, makes a move, Paul, can accept or reject it. If Paul is interested, a positive type, he accepts, and a match is formed, leading to a payoff of $\gamma\theta_i$. If Paul is not interested, a negative type, he rejects and receives $f(\theta_{-i})$. Judith, however, now incurs a cost of c associated with the shame or embarrassment of being rejected, or simply with the cost of investing in a futile move.

[11] Formally, if $-i$ chooses In, i's payoff is $\theta_i + b$ if i also chooses In, and $f(\theta_i) + b$ if i chooses Out.
[12] The fact that $\gamma < 1$ ensures that $f' < 1$, since if $\theta_i > 0$, $f(\theta_i) = \gamma\theta_i$.

♠ **Trust in Trade.** ($b+\theta_{\min} > 0$) Trading partners need to invest in a relationship-specific asset to maximize benefits from trade (Williamson 1979). While each would benefit from mutual investment, an opportunistic party benefits more if only the other party invests. This leads to the classic *hold-up problem*: if one party believes that her partner is opportunistic (negative) and would not reciprocate her initial investment, she has no incentive to invest either, forgoing benefits from trade. In contrast, if the parties are reciprocal (positive), one-sided investment is always reciprocated. Here, c can be interpreted as the loss from being help up. This loss may then increase in the extent to which ex ante promises are not enforceable ex post.[13] A similar situation may arise in negotiations, for example when a lender and a borrower needs to negotiate over sovereign debt, and the lender can decide whether or not to restructure the debt and the borrower can decide whether or not to implement economic reforms.

♣ **Costly Dissent** A member of an organization either disagrees with (a positive type), or agrees with (a negative type) an existing norm or business practice. When two members meet, each can voice dissent or deviate from the norm (choose in), or stay silent and act loyal (choose out). If a member agrees with the norm, he acts loyal. If he disagrees, he gains from explicit dissent if he expresses it in front of someone who also disagrees with the norm. They might form a coalition or merely experience a sense of liberation. When dissenting in front of a loyalist, however, the dissenter experiences a loss of c. The loyalist might ostracize or report him, causing the dissenter to be punished, fired, or persecuted.

3.3 Equilibrium

The next proposition characterizes the predictions. Below, $E^{\rho}_{\sigma^{\rho}}$ refers to the mean of a ρ-biased player i's posterior belief in equilibrium σ^{ρ} conditional on i's type and a realized action profile . The operator $E^{0}_{\sigma^{\rho}}$ refers to the true ex ante expected mean operator, given the true distribution of behavior generated by σ^{ρ} and the prior. Finally, E_0 refers to the prior mean.

Proposition 3 *For any ρ, there is a unique $\rho - IPE$. Player i enters iff $\theta_i > \theta_i^{*,\rho}$ where*

$$\theta_i^{*,\rho} = \sqrt{\frac{nc}{1-\rho}}.$$

where $n = |\theta_{\min}|$. Furthermore,

I. for any $\rho > 0$, and $\theta_i > 0$, $E^{\rho}_{\sigma^{\rho}}[\theta_{-i} \mid \theta_i, a] < E^{0}_{\sigma^{\rho}}[\theta_{-i} \mid \theta_i, a]$, given any $a \in A$,[14]

II. for any $\rho > 0$, $E^{0}_{\sigma^{\rho}}[E^{\rho}_{\sigma^{\rho}}[\theta_{-i} \mid \theta_i, a]] < E_0[\theta_{-i} \mid \theta_i]$ if $\theta_i > 0$,

[13] Since, here, $f(\theta_i) + b > 0$ for all θ_i, before they invest, each type has an incentive to convince her opponent that she is reciprocal because she benefits from her opponent investing. Hence, pre-play communication does not reduce the uncertainty.

[14] In the contingency where Judith chooses in and Paul chooses out, if Judith also observes her own payoff her underestimation is only weakly true. In all other cases, even when observing her own payoffs ex post, her underestimation is strict for any $c, \rho > 0$.

III. for any $\rho > 0$, $E^0_{\sigma^\rho}[E^\rho_{\sigma^\rho}[\theta_{-i} \mid \theta_i, a]] \geq E_0[\theta_{-i} \mid \theta_i]$ *if* $\theta_i < 0$.

By projecting information, each player underestimates the uncertainty her opponent faces about her preferences. This then implies the following results on actions and inference about others.

Under-Entry Equilibrium is given by cutoff strategies. To describe the prediction on actions, consider the bar example. By projecting information, an interested Judith exaggerates the extent to which Paul knows that she is interested. Since the projected Paul always enters if interested, she exaggerates the probability with which Paul will make a move. Since Judith still does not know whether Paul is interested, and because her payoff from waiting and reciprocating Paul's move if he invests is almost as high as from making a move initially, conditional on Paul being positive and reciprocating it, it now becomes relatively more important for her to stay out and reduce the risk of being shamed in case Paul were to reject her. By symmetry, the same holds for Paul. Hence, an increase in projection decreases each reciprocal player's willingness to invest, thus, the overall chance of an investment. As projection increases, no type ever invests; but each positive type increasingly expects the other party to invest if interested.

Underestimation If Judith is a positive type (trustworthy trading partner), she will always be too pessimistic about Paul's valuation of mutual investment relative to the truth given any $\rho > 0$. When seeing Paul make a move, Judith too often thinks that Paul invested because he knew that she would at least reciprocate such an investment; hence she underestimates the extent to which a move is good news about Paul's interest. When seeing Paul stay out, she is too convinced that Paul is not interested, as opposed to fearing being held up. Since a positive party believes that her opponent uses a lower average cutoff than he actually does, she underestimates him in all contingencies that arise in equilibrium.

False Antagonism The model predicts not only distortions in conditional inference, but also in average (ex ante expected sense) inference as well. Specifically, a positive type underestimates her opponent type not only conditionally but also on average. Since the decision to invest is positively correlated with Paul's interest, Judith comes to underestimate Paul's interest on average. In contrast, a negative Judith exaggerates the probability that Paul will adjust his action to her preferences, and stay out even if he is positive. Thus, she over-infers from Paul's entering, and under-infers from Paul staying out and overestimates Paul's interest, on average. In the unique $\rho-$IPE, beliefs no longer follow a martingale. Instead, information projection introduces a false *negative* correlation between one's own type and the perceived type of her opponent: on average, a player *always* come to conclude that others are less similar to her than she originally thought they were.

3.4 Dynamics

The willingness to invest is always decreasing in c, that is, the cost of being held up. It is, then, natural to consider settings in which the opportunity repeats itself. Consider a dynamic repetition of the exact same game over time t, except assume a changing value of c. Specifically, consider a strictly decreasing sequence $\underline{c} = \{c_t\}_{t=1}^{T}$.[15] For simplicity, I focus on myopic repetition: in each period t, players care only about the payoff of that period, but are able to recall the history of past interactions. In this context, the natural psychological assumption is that players project to some extent at the beginning of each new encounter independent of the history. That is, at the beginning of each t, each player believes in some probability $\rho > 0$ that her information privately leaks at that period t.[16]

Suppose that in each round, players play according to the unique $\rho-$IPE in that round. Let then $\Pr^{\rho}(M \mid \underline{c})$ be a measure of efficiency describing the true ex ante probability that, conditional on both players being positive, at least one party invests by the end of the sequence $\underline{c}$. Note that whenever at last party invests, the relevant uncertainty is resolved, and from thereon each type has a dominant strategy. Finally, let $q^{\rho}_{\underline{c}}$ be the true ex-ante expected posterior probability that a positive player assigns to her opponent being positive by the end of sequence $\underline{c}$, and let q_0 be the corresponding prior belief assigned to this.

Corollary 2 *Suppose $\rho = 0$. For any $\underline{c}$, $\Pr^{0}(M \mid \underline{c},) = \max\{1 - c_T n/(\theta^2_{\max}), 0\}$ and $q^{0}_{\underline{c}} = q_0$.*

In the Bayesian case, matching is efficient: as the loss from being hold up goes to zero, all positive types always match. Furthermore, as c_T vanishes, players always correctly learn whether their opponents have similar or opposite attitudes. Here, matching is also history-independent and depends only on the last element of any sequence $\underline{c}$.

The next corollary shows that given *any* positive degree of projection, the reverse can hold. Here, as c_T vanishes, no matches are formed, but with an ex ante probability of one, each positive type comes to wrongly conclude that her opponent is almost surely a negative type.

Corollary 3 (False Uniqueness) *For any $\rho, \tau > 0$, there exists a strictly decreasing $\underline{c}$ with $c_T = \tau$ such that $\Pr^{\rho}(M \mid \underline{c}) = 0$ and $q^{\rho}_{\underline{c}} \leq \tau$. If this is true for $\underline{c}$, it is also true for any $\underline{c}'$ such that $c'_t \geq c_t$ for all t.*

The above result specifies environments where even if c_t vanishes, no type ever enters, but each positive type concludes that her opponent must be a negative. Key to this corollary is that information projection leads to the differential attribution of identical behavior to self and others. While each positive type attributes her own lack of entry to her fear of her opponent

[15] Such a decrease could correspond to: (i) a wrong move being less costly in an informal than in a formal environment; (ii) an improvement in the enforceability of ex ante promises; and (iii) weakening disciplinary actions following reported dissent

[16] The results hold a fortiori if $\underline{c}$ is not strictly decreasing.

being opportunistic, she attributes the identical behavior of her opponent to the opponent actual being opportunistic.

The statement of the corollary focuses on the beliefs and actions of positive types. At the same time, in the above limit, negative types maintain correct views about their opponents. This is true because the projected version of a negative type's opponent now behaves the same way as her real opponent, hence, a negative type correctly makes no inferences from her opponent's behavior. Finally note that, here, beliefs of all types are fully self-confirming.

3.5 ♣ Discussion

Let me turn to a discussion of the above results in the context of some applications. Although the interactions above describe bilateral situations, they can be equally applied to such bilateral interactions taking place pair-wise between all members of a community.

■ **Norm Falsification** The prevalence of a norm, or the persistence of the status quo, is based not only on how many people actually support it, but often also, on how many people believe others support it. In the context of dissent, the above results imply that those who oppose the norm, misattribute the silence of others to their genuine loyalty. When speech is free, $c = 0$, people learn the truth about the attitudes of others; when such 'speech' is not free, Proposition 3 predicts a *systematic* wedge between the privately held support for a norm and the perceived public support for this norm: those who privately oppose the norm, will predictably come to exaggerate the public support for the norm. Under the conditions of Corollary 3, even if the vast majority opposes the norm – such as homophobia, affirmative action, political correctness, Stalin's leadership, or a corrupt business practice – the majority, nevertheless, comes to believe that the majority supports the norm.[17]

■ **Disciplinary Organizations** The predictions may matter for understanding organizations that sanction dissent. Such organizations will not only maintain loyalty, but create in its members a false sense of loyalty of other members. The proof of Corollary 3 implies that such sanctions c can be *gradually* removed over time without risking dissent. For example, the leadership of the organization can, now, save on the cost of running a disciplinary organization, which might increase in c, because the deterrent due to the size of c can gradually be replaced by the increasing pessimism of those who would like to change the status quo. Self-censorship may then outlive effective censorship due to the misattribution of the silence of others to their loyalty, and will lead to apathy as opposed to change.[18]

[17] The logic differs markedly from that of herding in sequential social learning. First, the setups differ. In sequential social learning, there is no direct strategic problem since players' payoffs do not depend on their opponents' actions (or types), which are used only to make inferences. Second, the identifying assumption of rational social learning is that, *on averagem,* players develop correct beliefs about others by virtue of the martingale property of Bayesian beliefs. Third, in a herd, rational or irrational, people come to believe that they act on the *same* expected preferences, as opposed to the *opposite* preferences, as others do.

[18] In his classic work, Hirschman (1970) discusses the critical role of voice versus loyalty for orga-

■ **Shy Revolutions** The above results also allow for a non-monotone comparative static with respect to the sequence $\underline{c}$. First, consider a sequence that satisfies Corollary 3. By the end of round T, all positive types are almost surely convinced that everyone else is negative. Suppose, now, that in round $T+1$, there is a drop and $c_{T+1} = 0$. Now, all efficient matches are formed. Such unexpected mass investment or dissent against the leadership comes as a great surprise potentially to almost all.

To see this more generally, consider what happens as c_t drops over time. There are two effects. First, due to the mechanical projection from one round to the next, positive types exaggerate the probability of entry by others. Since they are too surprised by how little entry there is relative to their expectations, they become too pessimistic. Second, due to the dynamic consequence of projection, positive types come to underestimate the fraction of others who are also positive types. As a consequence, given a small drop from c_t to c_{t+1}, the negative wedge between the private support for the norm and the private perception of the public support for it is still reinforced. If the drop from c_t to c_{t+1} is sufficiently large, however, the second effect dominates. Now all positive types are too surprised by seeing the large fraction of people who invest (dissent) which then reduces, or even completely eliminates, this wedge.

■ **Distrust** In the context of trade, the results imply a psychological hold-up problem. Exactly when trust is key, $c > 0$, trustworthy trading partners, failing to see enough the investment by others, will come to believe that their trading partners are likely to be opportunistic. Even if almost all people in a community are reciprocal, each such person may predictably come to conclude that all others are selfish or there for short-term gains. When such beliefs are passed on from one generation to the next, a gradual improvement of institutions, a decrease in c to some level, may have little or no effect on forming efficient partnerships due to mistaken distrust. A more dramatic improvement, leading to this level of c more quickly, may, nevertheless, lead to much higher investment.

■ **Mistaken Segregation** False antagonism above is the result of the joint presence of asymmetric information and limited perspective taking due to projection. This allows for a further comparative static. Let there be two groups, A and B. Suppose that the distribution of types is independent of group membership. Suppose also that in-group members can read each other's attitudes (preferences) ex ante but not the attitudes of out-group members, perhaps due to cultural differences. Proposition 3 then implies that each person will come to conclude that members of his or her own group are more likely to have matching objectives, to want to be friends when one wants to be friends, or to have similar attitudes towards the norm, than members of the other group. Creating an opportunity, such as interaction in an informal setting, where a wrong move is less costly may then reduce such views about the antagonistic preferences of others.

---

nizational change. Here, the above logic then implies a form of organizational apathy when voice is costly.

■ **Evidence** The above predictions are broadly consistent with a series of empirical observations discussed in the psychological literature under the rubric of pluralistic ignorance, introduced by Allport (1924) and summarized by Prentice (2007) as "the phenomenon that occurs when people erroneously infer that they feel differently from their peers, even though they are behaving similarly."[19] In an illustrative lab study, Miller and Mcfarland (1987) present students with a comprehension task of a difficult text. In the unconstrained treatment, students, seated in small groups, had an option to publicly leave the room, and seek clarification from an aid outside. In the constrained treatment, no such option was present. Although none actually left the room in the unconstrained treatment, there, students rated their own relative ability significantly lower than in the constrained treatment. Such an effect is consistent with the interpretation that students attributed their own lack of asking questions to the fear of embarrassment, but that of others to their superior comprehension.

In the context of norms, Prentice and Miller (1993) document a systematic wedge between the private support for a norm and the perception of the public support for it, and show that Princeton undergraduates rated the average comfort of others, including that of their friends, with the amount of alcohol consumption on campus, as significantly higher than their own and, hence, than that of reality. O'Gorman (1975) documents a similar wedge in the context of white males preference towards forced racial segregation, based a survey conducted in 1968.[20] In the context of social and political change, Kuran (1995) and Elster (2007) argue that such a wedge may have contributed to underestimating the demand for such charge prior to major large-scale social upheavals.[21] In the context of friendship formation, Shelton and Richeson (2005) show that students at Princeton and U. Mass desired having more interracial friendships (the same was not true for same-race friendships) but, consistent with the above logic, attributed the lack of their own initiative to the fear of rejection and the lack of initiative by members of the other racial group to their lack of interest.

3.6 Investment Games

Let me return to the more general case introduced in the beginning of this section. To characterize the implications of the model, a distinction between complement and substitute initial investments (entry) is needed. Initial investments are *substitutes* if, conditional on both play-

[19] Broadly consistent anecdotal evidence comes from H.C. Andersen's work on the Emperor's New Cloth.

[20] Based on a national survey form 1968 the ratio of the perceived fraction of whites who preferred segregation, as estimated by whites, versus the true fraction was close to 3. For example, 40 % of those in the West North Central of the US claimed that most whites favor segregation, although only 16 % of the same respondents were segregationists. Matza (1964) argues about the presence of a similar wedge amongst gang members towards gang violence.

[21] Examples listed include the unexpected popular support for the Solidarity Movement in Poland in the elections of 1981; the revolutions of Eastern Europe in 1989; for example, a year after the Fall of the Berlin Wall, over 70% of those surveyed said they were totally surprised by such a change; the Iranian Revolution of 1979, or the French Revolution.

ers being positive, the return on initial entry is higher if the opponent stays out than if the opponent enters initially. Investments are *complements* if the reverse is true.

Definition 2 *Investments are substitutes if* $\theta_i - f(\theta_i) < g(\theta_i, \theta_{-i})$ *whenever* $\min\{\theta_i, \theta_{-i}\} > 0$. *Investments are complements if* $\theta_i - f(\theta_i) > g(\theta_i, \theta_{-i})$ *whenever* $\min\{\theta_i, \theta_{-i}\} > 0$.

In the main example, $\gamma \in (0, 1)$ governs the degree of substitutability. If $\gamma \to 1$, investments are (almost) perfect substitutes. Conditional on Judith and Paul being interested in each other there is no real payoff differences in the scenario where they make a move simultaneously versus where one player only reciprocates the move of the other. In fact, for all $\gamma > 0.5$, initial investments remain substitutes, and *all* qualitative statements of Proposition 3 continue to hold. If $\gamma < 0.5$, investments are complements.

To illustrate the complements case with $\gamma = 0$, consider a different bar example. Suppose Judith and Paul each needs to decide whether or not to go to the bar. Only if both go do they have the opportunity to enjoy each other's company. If only one goes, the other learns about this. If this other is not interested all is the same as before. If they are both interested, there is still no shame cost if only one shows up initially, but, by the time the initially non–entering party learns about this and rushes to the bar, the night is gone, leaving them with essentially their outside option for the night.

In most important economic settings, initial investments are likely to be substitutes since the return on investing with a trustworthy opponent relative to receiving the outside option is likely to be higher than the return on being first to invest relative to reciprocating one's opponent's investment. For a more complete analysis, however, below, I analyze both cases.

Proposition 4 *For any* $\rho > 0$, *equilibria are given by cutoff strategies.*

1. *If investments are substitutes, there is a unique symmetric equilibrium, and this is increasing in* ρ.

2. *If investments are complements and* $g_2 = 0$, *all equilibria are symmetric, and the lowest one is decreasing in* ρ.

3. *In all equilibria,* $E^{\rho}_{\sigma^\rho}[\theta_{-i} \mid \theta_i, a] < E^{0}_{\sigma^\rho}[\theta_{-i} \mid \theta_i, a]$ *given any* $a \in A$ *and* $\theta_i > 0$.

4. *In all equilibria,* $E^{0}_{\sigma^\rho}[E^{\rho}_{\sigma^\rho}[\theta_{-i} \mid \theta_i]] < E_0[\theta_{-i} \mid \theta_i]$ *given any* $\theta_i > 0$ *and* $E_0[\theta_{-i} \mid \theta_i] \leq E^{0}_{\sigma^\rho}[E^{\rho}_{\sigma^\rho}[\theta_{-i} \mid \theta_i]]$ *given any* $\theta_i < 0$.

If investments are substitutes, a player's willingness to enter decreases in the likelihood that her opponent enters. Since, by projecting information a positive type exaggerates the probability with which her opponent enters if he is also positive, her own willingness to enter is decreased. This leads to under-entry in the unique symmetric equilibrium. If investments

are complements, a player's willingness to enter increases in the likelihood that her opponent enters. The perceived return on entry is now potentially exaggerated since a biased positive type exaggerates the probability with which her opponent enters. In this case, however, multiple symmetric equilibria may exist. The lowest equilibrium, the one with the highest probability of entry, now decreases in ρ leading to over-entry relative to the lowest BNE. The second-lowest equilibrium, if exists, however, increases in ρ leading to under-entry relative to the second-lowest BNE.

Importantly, the dynamic inference properties described in Proposition 3 continue to hold in *all* equilibria, both if initial investments are substitutes and if they are complements. Any equilibrium exhibits underestimation by positive types in any contingency, and false antagonism by all types, on average.[22] A positive type exaggerates the probability with which her positive opponent enters, and her assessments are too negative both conditionally and on average. A negative type underestimates the probability with which her opponent enters, and her assessments are too positive, on average.

Undervaluation of Social Assets The fact that in all equilibria a reciprocal type comes to underestimate her opponent's valuation in all contingencies implies a general failure of trust: those who would invest into the social asset become too skeptical about how much others would themselves want to invest into it. Even if a match is formed, each player will underestimate how much the other actually values the relationship. Such beliefs are often critical for a person's willingness to protect the relationship or further invest in it. To the extent that one's valuation of the joint asset increases in one's belief of how strong is one's partner's preference for mutual investment, projection leads not only to false antagonism, but to the undervaluation of social assets.[23]

4 Persuasion

I now turn to second application of study the implications of information projection to strategic communication. Persuasion is central to many settings. While an incentive to distort advice may often exist, a puzzle remains as to why people may fail to discount strategically distorted recommendations sufficiently. For example, Malmendier and Shanthikumar (2007, 2014) provide evidence that small investors take positively strategically distorted positive recommendations too much at face value and make biased investments. Della Vigna and Kaplan (2007) offer evidence in the context of politics and show that access to Fox news affects voting

[22] Again in the contingency where only one player enters underestimation maybe weak.

[23] Note that while the model predicts false antagonism in *inference,* this logically goes hand-in-hand with a '*false consensus'* effect in the *prediction* of action: a non-investing negative type exaggerates the probability with which her opponent will not enter either, and an entering positive type exaggerates the probability with which her opponent will enter as well. A large the current model may help to impose some discipline on such empirical investigations and offer structured and testable predictions.

decisions.[24] Following the recession of 2008, many have argued that such beliefs may have contributed to the financial crisis.

The identifying property of Bayesian communication is that receivers are never fooled on *average*. When the martingale property of correct Bayesian beliefs holds in a BNE, communication is informative but never by itself shifts the ex ante expected posteriors. In this section, I consider a classic sender-receiver game without commitment. A sophisticated sender provides advice to a receiver (investor) whether a proposition is true or false. The receiver can verify the sender's recommendation at a privately known cost c. Although the sender has a commonly known incentive to claim that the proposition is true, the model nevertheless predicts uniform credulity. If the conflict is sufficiently large, or it is sufficiently difficult to verify the sender's message, persuasion always leads to uniformly exaggerated ex ante expected posteriors.

Understanding the mechanism leading through which persuasion may inflate expectations and lead to credulity is potentially key. For example, in the UK, regulation since 2013 aims at capping the direct commission that financial advisors may receive from the producers of the asset which can be interpreted as a cap on the conflict between the sender and the receiver.[25] Similarly, many have argued about the role that financial education can provide in eliminating biased investor decision. Key comparative static predictions of the model imply that in fact it is the presence of only limited conflict and sufficient financial education of receivers which drives receivers to make biased and credulous decisions. Increasing financial education, or lowering the sender's incentive to lie, will often *fuel* such credulous beliefs and lower receiver welfare. By endogenizing the conflict and the complexity of the asset, when invoking its producer, I also describe why imposing *any* partial cap on the conflict may only have limited effectiveness and still allow for receivers to want to pay for advice which then only reduces their welfare relative to the case of no advice.

4.1 Setup

Timing. The sender is privately learns whether a proposition is true, $\{\theta = 1\}$, or false, $\{\theta = 0\}$. She then provides advice via cheap talk. A financial expert recommends to a mutual-fund manager whether to buy or sell a security; a lobbyist recommends to a politician whether or not to support a given policy; and a doctor recommends to a patient whether or not this patient should take a certain drug. Only the sender knows the truth value of the proposition.

Upon receiving advice, the receiver can verify the sender's statement at some cost c *privately* known to the receiver. If she verifies the message, she learns the value of θ. If she does not, she knows only the recommendation. Finally, the receiver takes an action y. For simplicity, I

[24] Exogenously invoked naive receiver types who always take recommendations at face value or ignore conflicts of interest have been considered by e.g., Kartik, Ottaviani, and Squintani (2007). Instead, here, credulity arises endogenously in equilibrium as a joint function of the conflict between the sender and the receiver and the extent to which the receiver is financially educated.

[25] See, for example, https://www.handbook.fca.org.uk/instrument/2011/2011_54.pdf .

assume the prior on θ to be symmetric.[26]

Verification. The receiver's cost of verification c is drawn according to a cdf $F(c)$ with a strictly positive density over $[0, \infty)$. Its realization is the receiver's private information. Differences in costs may reflect private information about the receiver's financial expertise, or his cost of accessing sources that allow him to assess the truthfulness of the sender's advice. A higher distribution of costs, a first-order stochastic increase in F (an increase in F, henceforth), corresponds to a lower distribution of receiver expertise or lower financial education. Equivalently, it can be interpreted as greater complexity of the asset to be evaluated.

Investment. Following the sender's recommendation and the receiver's verification decision, the receiver takes an action $y \in [0, 1]$ to maximize her expected utility. This action could correspond to the fraction of investment made in a given portfolio, the amount of resources invested in promoting or blocking a policy. To keep the analysis fully transparent, I assume that the optimal action simply equals the receiver's posterior confidence – that is his posterior that the proposition is true $\{\theta = 1\}$. This is captured by the standard assumption that the receiver's payoff is determined by the loss function:

$$u_r(y, \theta) = -(y - \theta)^2. \tag{6}$$

Conflict of Interest. The conflict between the sender and the receiver is such that the sender gets a bonus $B > 0$ – potentially from the seller of the asset to be introduced later – anytime she issues a positive recommendation – independent of the true state θ. At the same time, if the receiver decides to check the sender's recommendation and finds out that the sender has lied to him, the sender, the doctor or the lobbyist incurs a cost such as a loss of business, leading to a loss of $S > 0$. Without loss of generality, I normalize $S = 1$. Hence, B is always interpreted in proportional terms relative to S. Furthermore, to make the analysis non-trivial, I also assume that $B < 1$.[27] All of the above is common knowledge.[28]

Welfare When discussing receiver welfare (welfare, henceforth), I take the standard ex ante expected perspective. The receiver's welfare is given by the expected loss minus the potential verification cost incurred taking expectations, given the *true* distribution of actions in equilibrium.

4.2 Bayesian Case

Consider the unbiased case. In the unique equilibrium, the sender tells the truth if $\theta = 1$ and lies with probability p^0, if $\theta = 0$. The receiver checks a positive recommendation iff her cost

[26] All qualitative results hold for any strictly positive non-symmetric prior.

[27] If $B > 1$, the sender has a dominant strategy to issue a positive recommendation in each state.

[28] By assuming that the sender only receives a noisy binary signal about the truth-value of the proposition, one can ensure that absent verification the receiver never fully learns whether the sender lied or not even ex post while preserving the qualitative results presented in this Section.

is below a certain threshold c^0 and never checks a negative recommendation, but believes it fully. Below, $E_\theta[y_c^{*,0}]$ denotes the true ex ante expected equilibrium confidence (investment) of receiver type c. I denote the prior confidence by $\overline{y}$.

Proposition 5 *Let $\rho = 0$. In the unique equilibrium, the receiver checks iff $c \leq c^0(F,B)$, and the sender lies with probability $p^0(F,B) > 0$. An increase in F or B increases $c^0(F,B)$ and $p^0(F,B)$. Communication is neutral, $E_\theta[y_c^{*,0}] = \overline{y}$ for all c.*

Neutrality. In equilibrium, each type either checks a positive recommendation or discounts it proportional to the true lying probability. To maintain balanced incentives, a greater conflict, or lower receiver expertise (more complexity) induces more lying and more checking. A key feature of the BNE is that persuasion is neutral: the ex ante expected confidence of each type is the same as his prior. This is a direct and general consequence of the martingale property of Bayesian equilibrium beliefs. Although advice is valuable to the receiver, Bayesian communication is purely informative and never shifts *average* posterior beliefs.

4.3 Persuasion under Projection

Consider, now, a biased receiver ($\rho_R = \rho$) and an unbiased, thus sophisticated, sender ($\rho_S = 0$). Persuasion is no longer neutral. Instead, it leads to two kinds of mistakes: *credulity*, whereby a positive recommendation is taken too much at face value by the receiver, and *disbelief*, whereby a positive recommendation is interpreted with too much skepticism by the receiver. In the former case, persuasion successfully inflates confidence, on average – belief updating forms a sub-martingale. In the latter case, persuasion effectively decreases confidence, on average – belief updating forms a super-martingale.

Proposition 6 *For any $\rho > 0$, equilibrium is unique. There exist $0 < c_1^\rho < c_2^\rho \leq c_3^\rho$, such that*
(i) for $c < c_1^\rho$, persuasion is still neutral and $E_\theta[y_c^{*,\rho}] = \overline{y}$;
(ii) for $c \in [c_1^\rho, c_2^\rho)$, credulity holds and $E_\theta[y_c^{*,\rho}] > \overline{y}$;
(iii) for $c \in (c_2^\rho, c_3^\rho)$, disbelief holds and $E_\theta[y_c^{*,\rho}] < \overline{y}$; and
(iii) for $c \geq c_3^\rho$, (weak) disbelief holds and $E_\theta[y_c^{*,\rho}] \leq \overline{y}$.

To provide intuition, note that by projecting information, each receiver type exaggerates the extent to which the sender tailors the truthfulness of her strategy to his privately known type, as opposed to the commonly known distribution thereof. The receiver too often thinks that the advisor sees through him and knows how costly it would be for him to verify the recommendation. The real sender, however, knows only the distribution of the verification costs. Hence, the receiver exaggerates the extent to which the sender's message is conditioned on his privately known type, which in equilibrium implies that the biased receiver's perception of the probability with which the sender lies to him is decreasing in his cost c: the easier

it actually is for the receiver to verify the message, the more he thinks the sender must be truthful.[29]

In equilibrium, the real sender lies with probability p^ρ. The projected sender is perceived to lie according to a continuous function $p^+(c)$ strictly increasing only on a positive interval, $[c_1^\rho, c_3^\rho]$. Hence, there always exists a type $c_2^\rho \in [c_1^\rho, c_3^\rho]$ for whom the projected version's lying frequency matches the real sender's. This type develops correct beliefs on average despite projection. At the same time, all types below c_2^ρ believe a positive recommendation too much, and all types above are too skeptical. In turn, the verification strategy is such that types below c_1^ρ always check and believe that the projected sender never lies to them; types in $[c_1^\rho, c_3^\rho]$ check probabilistically, consistent with their increasing belief about the extent to which the projected sender lies to higher types; types above c_3^ρ never check, consistent with their belief that the projected sender always lies to them.

In sum, receivers with the highest expertise (lowest costs) always check and make correct decisions. Persuasion predictably boosts the average confidence of middle cost types, those with some but not full expertise who overinvest on average. Finally, the opposite holds for types with little or no expertise, persuasion decreases their confidence, on average.

4.4 Uniform Credulity

A key feature of the equilibrium predictions is that while credulity is always present, disbelief is only a limited phenomenon. Below, I refer to the case in which all receiver types are at least weakly credulous and a strictly positive measure is strictly credulous as *uniform credulity*. Here, persuasion unambiguously increases average receiver confidence, causing all types to (at least weakly) overinvest. The next result claims that given any degree of the bias, such uniform credulity always holds provided that the conflict is sufficiently high or verification is sufficiently complex.

Proposition 7 *For any $\rho > 0$,*

1. *If $B \geq \overline{B}(\rho, F)$, uniform credulity holds. Furthermore, $\overline{B}(\rho, F)$ is decreasing in ρ with $\lim_{\rho \to 1} \overline{B}(\rho, F) = 0$.*

2. *There exists $\overline{F}(\rho, B)$ such that for any $F \geq \overline{F}(\rho, B)$, uniform credulity holds. Furthermore, if an $\overline{F}(\rho, B)$ satisfies this for ρ, it also does for any $\rho' > \rho$.*[30]

[29] **True Leakage** The results on belief distortions do not depend on the perception of 'leakage' per se, but on *exaggeration* of such a perception due to projection. To see this, suppose that there was a true commonly known probability $\alpha \in [0, 1]$ with which the sender privately learned the receiver's type before making a recommendation. Although the equilibrium, here, would have a similar structure to that in Proposition 6, persuasion would, nevertheless, still be neutral for all types.

[30] Since first-order stochastic dominance is a partial order, of course, multiple such $\overline{F}(\rho, B)$ exist.

To provide intuition, note that as the conflict increases, the sender's incentive to lie increases. To maintain balanced incentives, this leads to more types checking and less information is transmitted via advice. Disbelief, however, is limited by the extent to which there is information transmission. If the conflict is sufficiently large, there is never enough incentive to provide balanced incentives to counterbalance the sender's incentive to lie, and disbelief no longer applies. At the same time, due to projection, all receiver types for whom it is rationalizable to check underestimate their incentives to check. Hence, all types in $[c_1^\rho, c_3^\rho]$ strictly overinvest, on average. Since an increase in F also makes it harder to provide incentives for the sender, the logic, here is analogous.

4.5 Welfare

I now turn to welfare. In the Bayesian case, a decrease in conflict or an increase in financial education (a decrease in complexity) increases the informativeness of advice and receiver welfare. In contrast, the presence of projection can reverse these comparative static results. The next result establishes sufficient conditions for this to be the case.

Proposition 8 *If $\rho = 0$, welfare is decreasing in B or F. For any $\rho > 0$,*

1. *If $B \geq \overline{B}(\rho, F)$, then an increase in B strictly increases receiver welfare.*
2. *If $F \geq \overline{F}(\rho, B)$ and $B < 0.5$, then an increase in F which does not change $F(\frac{1-\rho}{(2-\rho)^2})$ strictly increases receiver welfare.*

For any $\rho \geq 0$, welfare is maximal if the conflict is zero or verification is always free. In the unbiased case, welfare is monotone decreasing in these key economic variables. This no longer holds under any positive degree of projection. In fact, here, it is the combination of limited conflict and sufficient, but not full, financial expertise by receivers which creates the most scope for overly optimistic investments. Specifically, credulous types check too little in equilibrium because (a) they underestimate the value of checking, and (b) they make biased investment decisions in the absence of checking. A change that decreases their perception of the sender's incentive to lie relative to the truth, leads to bolder investment choices, and represents a negative force for welfare. Hence, a lower conflict and greater financial education can each decrease welfare. Specifically,

Comparative Static with B. A decrease in the conflict in the Bayesian case raises welfare because (i) it increases information transmission, and (ii) it induces less checking, which saves on the verification costs incurred. Given projection, less checking (iii) also leads to more-biased investments. Whenever uniform credulity holds, all receiver types (at least weakly) underestimate the sender's true lying probability and check too little relative to the value of checking. Now, a decrease in B does not affect the amount of information transmitted but

induces less frequent verification. Less checking then increases over-investment and reduces welfare.

Even if uniform credulity does not hold, the welfare of *some* types will always increase in the conflict as long as $\rho > 0$. Since types right above c_1^ρ are credulous and make the most distinctly overoptimistic investment choices, they enjoy a discontinuously lower expected utility than types just below c_1^ρ, who always check. Since an increase in B increases c_1^ρ, it improves the welfare of these originally most credulous types discretely. At the same time, it does not change the welfare of types below the original c_1^ρ. The overall welfare effect here, however, depends on further assumptions.

Comparative static with F In the Bayesian case, greater financial education (lower complexity) increases receiver welfare because (i) it mechanically decreases verification costs, and (ii) and makes it easier to provide truth-telling incentives which increases information transmission. Given projection, there is again a third effect: (iii) lower verification costs create more scope for credulity since, all else equal, the receiver's confidence when hearing a positive recommendation is decreasing in c due to projection. If uniform credulity holds, an increase in F lowers the perceived but not the real amount of information transmitted, tampering credulous expectations. If the conflict is not too great, then there is always sufficiently little checking such that the benefit of more checking following an increase in F by credulous types is always higher than the loss due to higher costs of checking. Holding constant the set of types who always check under uniform credulity, greater financial education now must reduce welfare. The same may hold even if uniform credulity is not satisfied.

4.6 Endogenous Conflict and Complexity

An implication of Proposition 8 is that under *any* positive degree of projection advice might reduce receiver welfare relative to the case of no advice: the welfare from the receiver simply acting on his prior about θ. Furthermore, the receiver might even be willing to pay for such advice ex ante. This all holds in a setting where the conflict and the distribution of θ is common knowledge. As mentioned, a complete ban on the conflict will certainly eliminate all distortions. At the same time, given projection any effectively only partial cap on the conflict may only have limited effectiveness. To demonstrate the above points, I now briefly endogenize the conflict and the complexity by invoking the seller of the asset.

So far, the conflict and complexity have been exogenous. Suppose, now, that, ex ante, before the resolution of any uncertainty, the seller – the manufacturer of the drug or the asset – pledges to pay the sender B whenever the sender makes a positive recommendation to the investor. As before, the chosen value of B is common knowledge. Suppose that the seller's expected profit is simply the ex ante expected investment – the aggregate demand for the asset – times a markup γ minus the transfer to the sender:

$$R(\rho, B) - B = \gamma E_{c,\theta}[y_c^{*,\rho}] - B. \tag{7}$$

What is the optimal B that an unbiased, hence sophisticated seller would want to offer to the unbiased sender?

In the Bayesian case with an unbiased receiver, the seller-optimal conflict is always zero. Since persuasion is neutral providing a bonus is a pure cost for the seller. Instead, given projection, an increase in B can increase aggregate beliefs. By limiting information transmission, the seller can now induce credulity and boost demand. Let $B^*(\rho, F)$ denote the seller-optimal bonus.

Corollary 4 *If $\rho = 0$, then $B^*(0, F) = 0$. If $\rho > 0$, then $B^*(\rho, F) > 0$ whenever $\gamma > \overline{\gamma}(\rho, F)$.*

Finally, suppose the seller can also affect the complexity of the asset or how much to educate receivers. As an extreme assumption, consider a seller can pick any B and any F satisfying full support. This is, of course, unrealistic, since typically the ability to manipulate F is limited. While a full analysis on the seller-optimal joint design of F and B is beyond the scope of the current analysis, let me conclude with a partial one. Suppose the seller wanted to implement uniform credulity. Here, the seller can maximize profit by minimizing the size of the bonus B and making sure that most of the mass of F is concentrated at a sufficiently high, but not too high, level of financial education.

Corollary 5 *Let $\rho > 0$. The seller-optimal way of implementing uniform credulity (i) minimizes $B > 0$ and (ii) concentrates F on $c_1^\rho = (1-\rho)/(2-\rho)^2$. Here, receiver welfare is decreasing in ρ and is always lower than without advice.*

As long as uniform credulity holds, the cutoffs c_1^ρ and c_3^ρ are independent of F and B. The confidence in a positive recommendation, conditional on not checking, is decreasing in c. Verification frequency increases, thus, overinvestment decreases in the conflict. Hence, the seller-optimal design of uniform credulity minimizes $B > 0$ and concentrates F on c_1^ρ – lowest type who does not perceive to have a strict incentive to check – but makes sure that there is a positive chance that the receiver might not find it rationzaliable to check. By making sure, that most receivers face sufficiently low costs of verification, but not too low so that they still do not strictly prefer to check, the seller now takes full advantage of the uniform credulity at the lowest possible cost.[31]

The table below illustrates the seller's and the receiver's expected payoffs in three cases: (i) without advice, (ii) in the seller-optimal design in the unbiased case, and (iii) in the limit of the above constrained seller-optimal setting for $\rho > 0$.

[31] More precisely, consider an F which is essentially zero for $c < c_1^\rho$, jumps to $1 - \varepsilon$ at c_1^ρ, for some $\varepsilon > 0$, and reaches 1 only for some $c > c_{\max}$.

	Receiver's welfare	Seller's expected profit
no advice	$-\frac{1}{4}$	$\frac{\gamma}{2}$
advice, $\rho = 0$	0	$\frac{\gamma}{2}$
advice, $\rho > 0$	$-\frac{1-\rho+0.5\rho^2}{(2-\rho)^2}$	$\frac{\gamma}{1+(1-\rho)}$

In the unbiased case, the seller cannot raise expected confidence, but information is useful for receivers on average. In the biased case, the receiver's welfare is affected discontinuously: for any $\rho > 0$, it is now lower with advice than without advice. Finally, the seller's expected profit increases and the receiver's welfare decreases in ρ.[32]

Paying for welfare reducing advise. Consider now the ex ante interpretation of the model whereby the receiver wrongly believes, before observing c, that his realized cost will privately leak to the sender with probability ρ. For any $\rho > 0$, such a receiver is now willing to pay a positive amount for advice which then predictably reduces his welfare.

Maximal Conflict Returning to the assumption that $B < 1$, note that when $B > 1$, it becomes common knowledge that the sender has a dominant strategy to issue a positive recommendation in all states. This ensures that the receiver is never exploited. In short, in this admittedly stylized setting, a partial cap on the conflict between the sender and the receiver may still allow for exploitative advice, a sufficiently large conflict never does.[33]

5 Projection Equilibrium

Above, I focused on projecting information only. Direct evidence provides strong support for this phenomenon. A logical counterpart of information projection is *ignorance projection:* the mistaken belief that if one cannot condition her strategy on an event, her opponent can not do so either. Taken together, they imply full projection, that is, that a person too often thinks that others can conditional their choices on the knowledge of exactly the same events as she can. Existing evidence directly supporting such distinct ignorance projection is sparse, which may suggest that in many domains it may be a weaker force. The technology introduced in Section 2, however, immediately allows one to incorporate the *joint* presence of information and ignorance projection. I now turn to the resulting solution of projection equilibrium.

If player j projects both her information and her ignorance onto player i, she exaggerates

[32] **Commitment** This settings considers strategic advice without commitment. In contrast, a recent literature considers the benefits for the seller to commit ex ante to a disclosure rule about θ, e.g., Rayo and Segal (2010). Note, however, that, here the seller prefers no such commitment. It is exactly the *lack* of such commitment which allows the seller to take advantage of the receiver's endogenously arising credulity.

[33] **Commitment** Finally, above I considered strategic communication without commitment. A recent literature considers the benefits of committing to an ex ante disclosure rule about θ by the seller, e.g., Rayo and Segal (2010). Note, however, that, here the seller prefers no such commitment. It is exactly the *lack* of such commitment which allows the seller to take advantage of the receiver's endogenously arising credulity.

the probability with which player i can condition his strategy on the same set of events as she can. Formally, the projected version of player i – who is real in the imagination of player j – now chooses a strategy from the set:

$$S_i^j = \{\sigma_i(\omega) \mid \sigma_i(\omega) \in \Delta A_i \text{ measurable w.r. to } P_j(\omega)\}. \tag{8}$$

In each state, this fictional projected version of i knows the events player j knows and only those events. The first part is due to information projection, the second is due to ignorance projection. I now state the definition of a projection equilibrium in a manner analogous to that of information projection equilibrium.

Definition 3 *A strategy profile $\sigma^\rho \in S_i \times S_j$ is a ρ-projection equilibrium of Γ if there exists $\sigma^\pm = \{\sigma_i^j, \sigma_j^i\} \in \{S_i^j \times S_j^i\}$ such that for all i,*

1.

$$\sigma_i^\rho \in BR_{S_i}\{(1-\rho)\sigma_{-i}^\rho \circ \rho\sigma_{-i}^i\} \tag{9}$$

2. and

$$\sigma_{-i}^i \in BR_{S_{-i}^i}\{\sigma_i^\rho\}. \tag{10}$$

The definition satisfies the same two properties as before: (i) projection is all-encompassing and (ii) it satisfies the limited consistency property. The projected version of Paul, as imagined by Judith, still knows that Judith is regular for sure as before. If the true game is poker, a biased Judith now wrongly thinks that with probability ρ Paul knows the value of *her* card and also that she does not know the value of his card. The difference is that Judith now also act as if the projected Paul did not know that value of his own realized card either, but, instead faced the same uncertainty about this as she did. Instead, in equilibrium Judith thinks that the projected Paul best–responds to Judith's strategy knowing the same set of cards as Judith does. Finally, Judith again expects Paul to play his actual strategy with probability $1-\rho$.

Note that equivalent versions of Propositions 1 & 2 and Corollary 1 continue to hold. Similarly, the structure of higher-order perceptions is the same as before; people again display partial sophistication about the biases of others.[34] The definition also immediately extends to heterogeneous projection in the same way as before.

Nested Model In many settings information projection is likely to occur even in the absence of ignorance projection. Importantly, though, one can nest the models and incorporate information projection and combined projection simultaneously but separately. Specifically, suppose each player i assigned probability ρ^κ to her opponent conditioning her strategy on S^+ and probability ρ^υ to her opponent conditioning her strategy on S_i^j and believing that j is

[34] As before, Paul thinks that Judith is projected with probability ρ; Judith underestimates this and thinks that, on average, Paul thinks that Judith is the projected version with probability $\rho - \rho^2$ etc.

regular with probability $1-\rho^{\kappa}-\rho^{\upsilon}>0$. Furthermore, maintain the assumption that projection is all-encompassing: the real player i believes that each projected version of j knows that i is real for sure. If $\rho^{\upsilon}=0$, the joint model collapses to that of information projection equilibrium. If $\rho^{\kappa}=0$, the joint model collapses to that of projection equilibrium.

5.1 Trade

As the final application, I apply projection equilibrium to the classic problem of common-value trade, the kind introduced by Akerlof (1970). The informed party, the seller, values the object of quality q at q. The uninformed party, the buyer, values it at $w(q)$. If $w(q)>q$, it is common knowledge that there are benefits from trade. Quality is drawn from a density π, and its realization is observed privately by only the seller.

This fundamental setting has not only found many applications, but has been studied experimentally starting with Samuelson and Bazerman (1985) and a literature following it. The remainder of this Section derives the model's predictions to this problem and relates it to the evidence. I also compare the empirical fit of the model with that of BNE and cursed equilibrium often motived by and applied to such key problems with adverse selection.

5.1.1 Additive Lemons Problem

Samuelson and Bazerman (1985, S&B henceforth) study two protocols. In the seller-offer game, the privately informed party, the seller, has the bargaining power and names the price $p_s^{\rho}(q)$ which the uninformed party can accept or reject. In the buyer-offer game, the uninformed party, the buyer, names a p_b, which the informed can accept or reject.[35] I follow their setting, where π is uniform on some $[a,b]$ with mean $\overline{q}$, and $w(q)=q+x$ with $x>0$.

Seller-Offer. A key feature of the BNE of this game is that different qualities can never be sold at difference prices for sure (or with the same probability). If they were, the seller would have a strict incentive to bluff, that is, to always name the higher of the two prices. Under projection equilibrium, however, a biased seller exaggerates the probability with which the buyer could detect such a bluff. This increases the scope for truth-telling. The following proposition describes the way this can hold.

Proposition 9 *For any $\rho\geq 0$, there exists a ρ projection equilibrium where $p_s^{\rho}(q)=q+x$, and the buyer accepts any price below $\overline{p}=\min\{\frac{x}{1-\rho},\overline{q}+x\}$ for sure and any higher price p with probability $e^{-(p-\overline{p})/x}$.*

In equilibrium, the seller's price fully reveals the quality. Yet, such no-bluffing is not altruistic; the seller always leaves the buyer with no rent, holding the buyer to his conditional (to

[35] Since in this sequential-move game, not all offers may be on the equilibrium path, I assume that the standard restriction of perfectness holds off the equilibrium path.

the buyer initially unknown) reservation value. Finally, all goods below some medium quality one are exchanged for sure, and higher quality items are bought with decreasing probability.

The above truthful equilibrium is supported by the strategies of the fictional projected versions. The projected buyer – who knows q – accepts a price p iff $p \leq q + x$. The projected seller – who does not know q, but knows that the buyer does not know q either – bids $\overline{q} + x$. The former implies that the real seller has a lesser incentive to bluff. The latter puts a bound on the highest price that can be accepted for sure. The bound on $\overline{p}$ is then determined by whether the IC constraint due to the deviation of the real seller, or that of the projected seller binds. The result relies on the joint presence of information and ignorance projection. This is true because by projecting her ignorance, the buyer also believes that despite knowing that the buyer does not know q, nevertheless, the projected seller does not base deviations on the realization of q.

Two qualitative properties characterize the predictions. First, the seller engages in non-altruistic truth-telling. Second, the seller under-bids relative to buyers' actual acceptance behavior. Both of these match the evidence closely. In particular, S&B study the case where $a = 0, b = 100$, and $x = 30$. They find that the the most common bidding strategy is $p_s(q) = q + 30$. [36] They also find that the privately informed sellers significantly underbid relative to what their payoff maximizing strategy would be, given the buyers' *actual* acceptance behavior. In particular, the acceptance probability in the data is fairly flat for any price below 80, but declines more sharply after that.[37]

Finally, if $\rho = 0$, the seller's maximal revenue is attained in the equilibrium in which the seller sells at a single price of $p = 60$, (Samuelson 1984). It is easy to show that if ρ is sufficiently high, the above equilibrium generates higher revenue than this Bayesian optimal one. Here, the fact that projection makes it easier for the seller to tell the truth in a (perceived) incentive-compatible way, then raises his expected revenue and also efficiency

Buyer-Offer. In the buyer-offer game, a ρ-biased buyer, acts as if she believed that with probability ρ the seller also does not know the realization of q. Her perceived expected utility from bidding p_b is, then

$$\begin{aligned} &(1-\rho)\Pr(q < p_b)[E_\pi[w(q) \mid q < p_b] - p_b] \text{ if } p_b \leq \overline{q} \\ &\rho[E_\pi[w(q)] - p_b] + (1-\rho)\Pr(q < p_b)[E_\pi[w(q) \mid q < p_b] - p_b] \text{ if } p_b > \overline{q}, \end{aligned} \tag{11}$$

since both the seller has a dominant strategy: the real seller accepts p_b iff $p_b \geq q$; the projected seller accepts p_b iff $p_b \geq \overline{q}$. In the above specification of S&B, this implies the following claim.

[36] The overwhelming majority of the other bids satisfied $p_s(q) \in [q, q + 30]$.

[37] In particular, the acceptance probability in the data is fairly flat for any price below 80 but declines more sharply after that. Calibrating the above result, for any $\rho > \frac{5}{8}$, $\overline{p} = 80$ consistent with this.

Claim 1 *In the unique ρ projection equilibrium, the buyer's bid is given by*

$$p_b^\rho = 30 \text{ if } \rho \leq 1/16 \text{ and } p_b^\rho = 50 \text{ if } \rho > 1/16$$

The model's prediction is discontinuous in ρ.[38] If ρ is small, the prediction is identical to that of BNE. If ρ is larger than 6.2 %, the buyer bids the seller's unconditional value, that is the mean quality, $\overline{q}$. Given that there is negative selection, the buyer significantly overbids and achieves a much smaller expected payoff than in the unbiased case, she is subject to the classic 'winner's curse'.

In line with the model's unique prediction, S&B find that the most common bid is 50. Furthermore, in their data, less than 17% of bids are in $[30, 40]$; and most other bids are in the $[50, 80]$. A non-trivial fraction of bids are above 60. Since under correct expectations, bidding above 60 leads to strictly negative expected earnings for the buyer this cannot be rationalized by the presence of social preferences. In contrast, bidding below 80 still leads to positive perceived earnings under projection for any $\rho > 1/16$.

In the buyer-offer game, cursed equilibrium,$CE(\chi)$, also predicts plausible deviations from the BNE. The predictions of cursed equilibrium span the interval $[30, 40]$ as a function of the degree of cursedness, with 40 being the fully cursed prediction. Projection equilibrium, thus, matches the data in both treatments better than CE and BNE for any $\rho > \frac{1}{16}$.

5.1.2 Multiplicative Lemons Problem

Holt and Sherman (1994) test a multiplicative specification of this problem in which $w(q) = 1.5q$ and π is uniform on $[q_0, q_0 + r]$. They focus only on the buyer-offer game. Table 1 characterizes the predictions of projection equilibrium in the three conditions studied experimentally and calibrated by Eyster and Rabin (2005). Below, the average empirical bid is $\overline{b}$; the unbiased BNE predictions corresponds to $b(\chi = 0)$, the fully cursed prediction to $b(\chi = 1)$, with CE(χ) spanning the interval between these two. The unique ρ-projection equilibrium has the same threshold structure as before. If $\rho \leq \rho^*$, it is identical to the unbiased BNE. If $\rho > \rho^*$, the buyer bids $b(\rho > \rho^*)$ independent of ρ.

[38] Given a fixed ρ^{υ}, the predictions are unaltered by considering any feasible $\rho^{\kappa} \geq 0$.

	$[r]$	$[q_0]$	$[m]$	$b(\chi=0)$	$b(\chi=1)$	$b(\rho>\rho^*)$	ρ^*	$\overline{b}$
No Curse	2	1	1.5	2	2	2	0	2
Winner's Curse	4.5	1.5	1.5	3	3.5	3.75	0.02	3.78
Loser's Curse	0.5	0.5	1.5	1	0.81	0.75	0.07	0.74

Table 1: Holt and Sherman (1994), Eyster and Rabin (2005).

As Table 1 shows, in *all* treatments predictions of projection equilibrium match the data almost perfectly. In the winner's curse condition, where the buyer's optimal bid is lower than $\overline{q}$ this is true for any $\rho > 2\%$. In the loser's curse condition, where $\overline{q}$ is lower than the optimal bid, this is true for any $\rho > 7\%$. In the no curse condition, $\overline{q}$ is the optimal bid and the predictions of all these models are the same. Hence, for all $\rho > 7\%$, the predictions of projection equilibrium robustly match the data.[39] The reason that such a small degree of projection leads to such substantial deviation is that the gain from trade conditional on selection is much smaller than the gain from trade without selection.[40]

Cursedness versus Ignorance Projection In the buyer-offer game, both cursedness and ignorance projection imply plausible deviations from the predictions of BNE. Their logic and the predictions differ substantially. A cursed buyer has correct expectations about the seller's information, but wrongly thinks that with probability χ the seller plays the same strategy irrespective of q. A buyer who projects her ignorance acts as if she wrongly believes that with probability ρ the seller only knew π but not the realization of q, but then has fully coherent beliefs about how such a projected seller would behave. This implies that a fully cursed buyer, the prediction in this class closest to the empirical behavior, acts as if he believed that his offer should be accepted independent of q. The predictions of projection equilibrium almost perfect matching the data, instead allow the buyer to believe in a very high positive correlation between the seller's acceptance decision and the value of q. It also implies, as ER (2005) note, that $CE(\chi)$ predicts a strictly positive bid even if $m < 1$, that is, even if the buyer always values the object strictly less than the seller does. Projection equilibrium, here predicts a bid

[39] Ball, Bazerman, and Carroll (1991) study a close variant of the above specification and also allow for multiple rounds of learning. Here, the relevant threshold is $\rho > \rho^* = 0.12$

r	q_0	m	$b(\chi=0)$	$b(\chi=1)$	$b(\rho>\rho^*)$	$\overline{b}$
1	0	1.5	0	0.375	0.5	0.55

[40] As mentioned, in a different context, Danz et al. (2014) estimate a $\rho = 0.28$.

of 0.[41]

Projecting Valuation Finally, the above data is inconsistent with the idea that players mistakenly thinks that others have the same valuations as opposed to the same information as they do. Note that informed sellers bid the buyers' higher conditional valuations, and uninformed buyers bid the sellers' lower unconditional valuations. They act as if they exploited binding individual rationality constraints ignoring informational differences.

5.2 Multi-Player Extension

Lastly, consider the extension of the model to N players. Below, I formally define the extension for projection equilibrium. The extension for information projection equilibrium is perfectly analogous. Key to this extension is that, now. each player i has a collection of projected opponents. In addition, since the information of players i and j differ, the projected version of player k, as imagined by player i, differs from the projected version of k, as imagined by player j. The former conditions his strategy on player i's information. The latter conditions his strategy on player j's information.

As introduced in Section 5, let S_k^j be the strategy set of the projected version of player k real in player j's imagination. I denote its generic element of this set by σ_k^j. Let $S_{-j}^j = \prod_{k \neq j} \times S_k^j$ be the strategy set of the $N-1$ fictional opponents who are real in player j's imagination. I denote the generic element of this product set by σ_{-j}^j. Lastly, denote by $\sigma_{-(j \cup i)}^j$ the restriction of the strategy profile σ_{-j}^j containing all of its elements except for σ_i^j.

In the definition below, projection is again all-encompassing: each projected opponent of j knows that j is real. Furthermore, each player believes that with probability ρ all her opponents are projected versions and with probability $1-\rho$ they are all regular versions. Finally, each projected opponent of j has the same beliefs about the distribution of the versions of the rest of the players as j does. If the true game is poker, a biased Judith thinks that Paul knows the value of her card with probability ρ, as before. Judith, now, also thinks that such a projected Paul also believes that with probability $1-\rho$ Sam knows the value of Sam's card only and with probability ρ Sam knows the value of Judith's card only. Finally, the same consistency property holds as before.

Definition 4 *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*

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

where each $\sigma_j^i \in \sigma_{-i}^i$, for any $j \neq i$, is such that

$$\sigma_j^i \in BR_{S_j^i}\{\sigma_i^\rho, (1-\rho)\sigma_{-(i \cup j)}^\rho \circ \rho\sigma_{-(i \cup j)}^i\}.$$

[41] For a related reason, CE is not defined for games where the action set depends on the state.

Note first that if $\rho = 0$, the above definition collapses to that of a BNE for Γ. The extension of information projection equilibrium to N players is analogous and differs in that each S_j^i above is replaced with $S_j^{i+j} = \{\sigma_i(\omega) \mid \sigma_i(\omega) \in \Delta A_i \text{ measurable w.r. to } P_j \cap P_i\}$.

6 Conclusion

A wealth of direct evidence shows that people engage in limited perspective-taking and project their information (and possibly also their ignorance) onto others. This paper incorporates this phenomenon into Bayesian games by proposing a fully specified model using a canonical and transparent framework. Accounting for this robust mistake may then shed novel light on a variety of economic problems and help establish the extent to which it matters empirically.

Projection in social perception will matter in contexts other than those described in this paper. It is likely to shape negotiations and bargaining behavior, the outcomes of contests, information aggregation in committees, or markets with private information. A context, where the wedge between true and perceived differences maybe particularly important is incentive design. In mechanism design, a key concern is the optimal provision of rents to privately informed agents. Projection will affect the demand for information rents and, thus, modify key incentive compatibility constraints. This will affect the shape of optimal mechanisms and the scope for truth-telling and efficiency as in Section 5.

In this vein Madarász (2014b) extends the model to sequential bargaining with observable moves. This paper shows that the presence of even minimal projection can significantly alter the seller-optimal way to sell an object, Myerson (1981). Dynamic haggling without commitment now dominates posted prices and the model predicts a complete reversal of the Coasian property of bargaining. As bargaining becomes smooth, the uninformed seller is able to extract all benefits from trade. I also show that the existing evidence rejects the Bayesian comparative static results but, instead, is consistent with the model. Dynamic extensions of the model to social learning or consumers' perception of the value of their privacy and firms' dynamic contracting responses to projection-based misperceptions maybe fruitful.

The portability of the model allows for an immediate evaluation of its empirical implications in a variety of domains. Danz, Madarász and Wang (2014) find strong support in the context of an agency problem. The current compares the precise predictions to data from common-value trade. Importantly, by providing a disciplined hypothesis for understanding potential inferential mistakes in social cognition, where such findings may sometimes lack a well-formulated ex ante hypothesis, the model may help provide a more unified explanation and careful empirical evaluation of these findings, and provide testable predictions as to when they might or might not occur.

7 Appendix

Proof of Proposition 1. The existence of IPE (PE) follows immediately from the existence of a BNE given Kakutani's theorem since the best-response correspondences are upper hemicontinous and convex .

Proof of Corollary 1. If $P_i = P_j$, then $P^+ = P_i = P_j$. Hence, for any $\sigma \in BNE(\Gamma)$, $\sigma^+ = \sigma$ satisfies the definition of a $\rho-$IPE for any ρ since $S_i^+ = S_i$. By the same token, any σ^ρ which is a $\rho - IPE$ of Γ must also be a BNE of Γ. The logic immediately extends to projection equilibrium .

Proof of Proposition 2. Suppose σ^0 is a BNE and is also an ex-post equilibrium. Then, for each i

$$u_i(\sigma_i^0(\omega), \sigma_{-i}^0(\omega), \omega) \geq u_i(\sigma_i'(\omega), \sigma_{-i}^0(\omega), \omega) \text{ for all } \sigma_i'(\omega) \in \Delta A_i \text{ in each } \omega \in \Omega.$$

Consider now $\sigma_i^+(\omega) = \sigma_i^0(\omega)$ for each ω. It follows that $\sigma_i^+ \in BR_{S_i^+}(\sigma_{-i}^0)$. Hence, given $\sigma^+ = \sigma^0$, it follows that σ^0 is a $\rho - IPE$ for any ρ. The logic immediately extends to projection equilibrium .

Proof of Proposition 3. The proof of Proposition 4 shows that equilibrium is in cut-off strategies. To simplify notation, let $x = \theta_{\max}$. If θ_{-i}^ρ is player $-i$'s cutoff, then player i is indifferent between In and Out at θ_i^ρ satisfying

$$\rho(x(\theta_i^\rho - \gamma\theta_i^\rho) - nc) + (1-\rho)((x - \theta_{-i}^\rho)(\theta_i^\rho - \gamma\theta_i^\rho) + \theta_{-i}^\rho(\gamma\theta_i^\rho) - nc) = 0 \tag{12}$$

Solving for θ_i^ρ, one obtains

$$\theta_i^\rho = \frac{cn}{x(1-\gamma) + \theta_{-i}^\rho(1-\rho)(2\gamma - 1)}.^{42} \tag{13}$$

Substituting in the symmetric equation for θ_{-i}^ρ, then taking $\gamma \to 1$, the unique interior solution is $\theta_i^\rho = \sqrt{nc/(1-\rho)}$.

1. If $\theta_i > 0$, then $\theta_{-i}^\rho > \theta_{-i}^+(\theta_i) = 0$ for all $c, \rho > 0$. This implies strict underestimation given any $a \in A$ since the perceived average cutoff used by $-i$ is strictly lower than one used by real $-i$. This still holds also if player i observes her own payoff, except when $(a_i, a_{-i}) = (in, out)$. Here, underestimation is weak since whenever i learns that $-i$ is positive, i knows that $-i$ must have been the real version, thus develops correct beliefs.

2. Let $\Pr(a_{-i} = in \mid \theta_i)^\rho$ be the perceived probability that type θ_i assigns to player $-i$'s entering in equilibrium – $\Pr(in)_{\theta_i}^\rho$ for short below. Similarly, $\Pr(in)_{\theta_i}$ will be the corresponding

[42] I adopt the convention that when no interior solution exists, then $\theta_i^\rho = \theta_{\max}$.

true probability. For each θ_i, the martingale property of beliefs holds with respect to this *perceived* probability in equilibrium. Hence, by the law of total probability,

$$E_0[\theta_{-i}|\ \theta_i] = \Pr(in)^{\rho}_{\theta_i} E^{\rho}_{\sigma^\rho}[\theta_{-i}|\ \theta_i, a_i, a_{-i}= in] + (1-\Pr(in)^{\rho}_{\theta_i}) E^{\rho}_{\sigma^\rho}[\theta_{-i}|\ \theta_i, a_i, a_{-i}= out]\,, \quad (14)$$

where a_i is the action taken by θ_i in equilibrium. Let

$$\Delta(\theta_i) \equiv E^{\rho}_{\sigma^\rho}[\theta_{-i}|\ \theta_i, a_i(\theta_i), a_{-i}= in]] - E^{\rho}_{\sigma^\rho}[\theta_{-i}|\ \theta_i, a_i(\theta_i), a_{-i}= out]$$

denote the difference between the conditional mean estimates of type θ_i when observing player $-i$ entering versus staying out. Note that $\Delta(\theta_i) > 0$ holds for all θ_i. Consider now average inference.

$$E_0[\theta_{-i}|\ \theta_i] - E^0_{\sigma^\rho}[E^{\rho}_{\sigma^\rho}[\theta_{-i}|\ \theta_i]] = \Delta(\theta_i)(\Pr(in) - \Pr(in \mid \theta_i)^{\rho}). \quad (15)$$

If $\rho = 0$, the RHS of Eq.(15) must be zero for any θ_i. Suppose $\rho > 0$. If $\theta_i > 0$, then $\theta^{\rho}_{-i} > \theta^{+}_{-i}(\theta_i)$, and $\Pr(in)^{\rho}_{\theta_i} > \Pr(in)_{\theta_i}$; hence, the RHS of Eq.(15) is strictly negative. If $\theta_i < 0$, then $\theta^{\rho}_{-i} \leq \theta^{+}_{-i}(\theta_i)$, and $\Pr(in)^{\rho}_{\theta_i} < \Pr(in)_{\theta_i}$; hence, the RHS of Eq.(15) is positive .

Proof of Corollary 2. Conditional on no entry by either players until round $t-1$, in round t, any player i's belief about $-i$'s type is given by a uniform density on $[\theta^0_{-i,t-1}, -n]$. It follows that the indifference cut-off of player i in round t, again *conditional* on no entry until then, using Proposition 3, must be $\theta^0_{i,t} = \sqrt{nc_t}$. Hence, $\Pr^0(M \mid \underline{c}) = \max\{1 - nc_T/x^2, 0\}$ since, given this inductive logic, by round T all types greater than $\sqrt{nc_T}$ must have entered .

Proof of Corollary 3. Let $q^{\rho}_{t,-i}$ be the probability that player i assigns to the event that $\theta_{-i} > 0$ conditional on no entry by either player until round t. Let $z^{\rho}_{t,-i}$ be the ex ante probability with which the real $-i$ enters in round t *conditional* on no entry by either player until round t. It follows from Proposition 3, that player i's cutoff in round t, conditional on entry till then, is

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

which is decreasing in $q^{\rho}_{t,-i}$ and increasing in c_t.

Suppose $c_t > (1-\rho)x(q^{\rho}_{t,-i})/(1-q^{\rho}_{t,-i})$ for all $t < M(\rho)$. Then $z^{\rho}_{t,-i} = 0$ for all $t < M(\rho)$. Furthermore, as long as $\rho > 0$,

$$q^{\rho}_{t,-i} = \frac{q^{\rho}_{t-1,-i}(1-\rho)}{(1-\rho) + (1-q^{\rho}_{t-1,-i})\rho} < q^{\rho}_{t-1,-i} \text{ for all } t < M(\rho) \text{ and } i \quad (16)$$

and the c_t sequence is strictly decreasing. Set $c_T = \tau > 0$. Since the belief sequence given by Eq.(16) converges to 0 as $M(\rho)$ goes to infinity for any $\rho > 0$, it follows, that, for any $\varepsilon > 0$,

there exists $m(\rho)$ such that $q_{m,-i} \leq \varepsilon$ if $m > m(\rho)$. It follows that there exists $m(\rho)$ such that

$$\theta^{\rho}_{i,m+1} = \frac{(1-q^{\rho}_{m,-i})c_T}{(1-\rho)\,(q^{\rho}_{m,-i} - z^{\rho}_{m,-i})} \geq \theta_{\max} \tag{17}$$

.

Proof of Proposition 4. To simplify notation, let $d = (\theta_{\max} - \theta_{\min})^{-1}$.

1. First, I show that equilibrium is in cutoff strategies. Note that the projected version of player $-i$ enters iff $\min(\theta_i, \theta_{-i}) > 0$. Let z_{-i} be the ex ante probability, given some strategy $\sigma_{-i} \in S_{-i}$ of real player $-i$, that real $-i$ enters. For any real type $\theta_i > 0$, the expected utility difference between entering versus staying out is then

$$\rho d(x(\theta_i - f(\theta_i)) + \textstyle\int_{\theta_{\min}}^{0} g(\theta_i, \theta_{-i}) d\theta_{-i}) + \\ (1-\rho)(z_{-i}(\theta_i - f(\theta_i)) + (1 - z_{-i}) E_{\theta_{-i}}[g(\theta_i, \theta_{-i}) \mid \sigma_{-i} = out]). \tag{18}$$

Differentiating this expression in Eq.(18) with respect to θ_i it follows that this difference is *strictly* increasing in θ_i for any given σ_{-i} since $f_1 < 1$ and $g_1 \geq 0$, for $\theta_i > 0$. Since indifference is attained when this expression equals zero, equilibrium must be in cutoff strategies.

2. Consider then the best-response functions. The function determining player i's cutoff is $\beta^{\rho}(\theta_{-i}) : [0, \theta_{\max}] \to [0, \theta_{\max}]$. It is defined on the positive domain and range only since negative types stay out in equilibrium. Note that $\beta^{\rho}(\theta_{-i})$ is continuous in $\theta_{-i} > 0$.The implicit function theorem then implies that the slope of $\beta^{\rho}(\theta_{-i})$, evaluated at some point $(\widehat{\theta}_i, \widehat{\theta}_{-i})$ on this curve, is

$$\overbrace{(1-\rho)(\widehat{\theta}_i - f(\widehat{\theta}_i) - g(\widehat{\theta}_i, \widehat{\theta}_{-i}) - d\textstyle\int_{\theta_{\min}}^{\widehat{\theta}_{-i}} g_2(\widehat{\theta}_i, \theta_{-i}) d\theta_{-i})}^{I} * \\ \overbrace{[\rho(x(1 - f'(\widehat{\theta}_i)) + d\textstyle\int_{\theta_{\min}}^{0} g_1(\widehat{\theta}_i, \theta_{-i}) d\theta_{-i}) + \\ (1 - \rho((\Pr(\theta_{-i} > \widehat{\theta}_{-i})(1 - f'(\widehat{\theta}_i)) + d\textstyle\int_{\theta_{\min}}^{\widehat{\theta}_{-i}} g_1(\widehat{\theta}_i, \theta_{-i}) d\theta_{-i})]^{-1}}^{II}. \tag{19}$$

Term II is strictly positive. Term I is strictly negative if investments are substitutes, and strictly positive if investments are complements and $g_2 = 0$.

3. By the intermediate value theorem, a symmetric equilibrium must exist since $\beta^{\rho}(\theta_{-i})$ is continuous, with $\beta^{\rho}(0) \leq \theta_{\max}$ and $\beta^{\rho}(\theta_{\max}) \leq \theta_{\max}$, and the players' best-response functions are mirror images on the 45-degree line. Consider substitute investments. Since $\beta^{\rho}(\theta_{-i})$ is strictly decreasing, there is a unique symmetric equilibrium. Consider complement investments. Here, all equilibria must be symmetric. This is true since, given that $\beta^{\rho}(\theta_{-i})$ is strictly increasing, $\widehat{\theta}_i = \beta^{\rho}(\widehat{\theta}_{-i}) > \beta^{\rho}(\widehat{\theta}_i) = \widehat{\theta}_{-i}$ cannot hold.

4. Consider the comparative static with respect to ρ. Consider cutoffs $(\theta^{\rho}_i, \theta^{\rho}_{-i})$ that

constitute a $\rho - IPE$ for a given ρ. Rewriting the equilibrium condition, using Eq.(18), one gets that

$$\overbrace{\rho[d\int_0^{\theta^\rho_{-i}} \theta_i^\rho - f(\theta_i^\rho) - g(\theta_i^\rho, \theta_{-i})d\theta_{-i}]}^{V} + \tag{20}$$

$$\Pr(\theta_{-i} > \theta^\rho_{-i})(\theta_i^\rho - f(\theta_i^\rho)) + d\int_{\theta_{\min}}^{\theta^\rho_{-i}} g(\theta_i^\rho, \theta_{-i})d\theta_{-i}] = 0, \tag{21}$$

First, note that the LHS is still increasing in θ_i^ρ. In addition, if initial investments are substitutes, Term V is negative. Holding $(\theta_i^\rho, \theta^\rho_{-i})$ fixed, the LHS of Eq.(20) is decreasing in ρ. Hence, for a fixed θ^ρ_{-i}, an increase in ρ must be compensated by an increase in θ_i^ρ. Hence an increase in ρ shifts the best-response function up.

If investments are complements, Term V is positive. Holding $(\theta_i^\rho, \theta^\rho_{-i})$ fixed, the LHS of Eq.(20) is now increasing in ρ. Furthermore, since $\theta^+_{-i}(\theta_i) = 0$ for $\theta_i > 0$, $\beta^\rho(0)$ is independent of ρ, and $\beta^\rho(0) > 0$, given investment risk. An increase in ρ, thus, shifts $\beta^\rho(\theta_{-i})$ downward for all $\theta_{-i} > 0$. Hence, the lowest equilibrium cutoff, the first intersection of $\beta^\rho(\theta_{-i})$ with the 45-degree line, is decreasing in ρ. The second intersection, if exists, is increasing in ρ since $\beta^\rho(\theta_{-i})$ is continuous in θ_{-i} and ρ.

Underestimation. Suppose that $\theta_i > 0$. Since $g(\theta_i, \theta_{-i}) < 0$ if $\min\{\theta_i, \theta_{-i}\} < 0$, and $g(0, \theta_{-i}) = 0$, it must be the case that $\theta^\rho_{-i} > \theta^+_{-i}(\theta_i) = 0$, hence, the statement follows from the proof of Proposition 3. **False antagonism**. Note that Eq. (15) still holds; hence, the statement follows from the proof of Proposition 3 .

Proof of Proposition 5. If $\rho = 0$, since the benefit of checking is strictly decreasing in c, the receiver adopts a cutoff checking strategy. The indifferent type is $c^0 = p^0/(1+p^0)^2$.[43] Let $c_{\max} = 1/4$. Since $B > 0$, $p^0(c^0)$ is uniquely determined by c^0 solving $c^0 = \min\{F^{-1}(B), c_{\max}\}$.

Proof of Proposition 6. Suppose that $\rho > 0$. Let $p^+(c)$ be the projected sender's lying probability if $\theta = 1$, given receiver type c. This $p^+(c)$ must increase in c without jumps. To see this, consider $c > c'$, but suppose that $p^+(c) < p^+(c')$. Since p^ρ does not depend on c, type c has a strictly lower incentive to check than type c'. Hence $p^+(c) = 1$, a contradiction. Furthermore, if a type c always checks, then $p^+(c) = 0$ and if a type c never checks, then $p^+(c) = 1$. Hence $p^+(c)$ must smoothly increase on some $[c_1^\rho, c_3^\rho]$ with $c_1^\rho < c_3^\rho$ and $p^+(c_1^\rho) = 0$ and $p^+(c_3^\rho) = 1$. Each $c \in [c_1^\rho, c_3^\rho]$ must play a mixed checking strategy to ensure that $p^+(c) \in (0,1)$ for $c \in (c_1^\rho, c_3^\rho)$.

[43] Since without checking $y = 1/(1+p^0)$, and the indifference condition is

$$c^0 = \frac{1}{1+p^0}(\frac{1}{1+p^0} - 1)^2 + \frac{p^0}{1+p^0}(\frac{1}{1+p^0})^2$$

.

There then exists $c_2^\rho \in (c_1^\rho, c_3^\rho]$ such that $p^+(c_2^\rho) = p^\rho$. Hence, if $c \in (c_1^\rho, c_2^\rho)$, $E_\theta[y^*] > \frac{1}{2}$ and if $c > c_2^\rho$, $E_\theta[y^*] \leq \frac{1}{2}$. .

Lemma 1 *The cutoff c_1^ρ is weakly decreasing and the cutoff c_3^ρ is weakly increasing in ρ.*

Proof of Lemma 1. Proceed by contradiction. Suppose that $\rho' > \rho$, but $c_3^{\rho'} < c_3^\rho$. Since $p^+(c_3^\rho) = p^+(c_3^{\rho'}) = 1$, $p^{\rho'} < p^\rho$ must then hold by monotonicity since c_3^ρ is increasing in p^ρ and in ρ separately. This also implies that $c_1^{\rho'} < c_1^\rho$ must also hold but then there is strictly less incentive to lie under ρ and $p^{\rho'} > p^\rho$, a contradiction. Hence, c_3^ρ is weakly increasing in ρ.

Suppose that $\rho' > \rho$, but $c_1^{\rho'} > c_1^\rho$. Since $p^+(c_1^\rho) = p^+(c_1^{\rho'}) = 0$ and $c_3^{\rho'} \geq c_3^\rho$ by the previous argument, it must be that $p^{\rho'} \leq p^\rho$, which then implies that $c_1^{\rho'} \leq c_1^\rho$, a contradiction. Hence, c_1^ρ is weakly decreasing in ρ .

Proof of Proposition 7. The sender's incentive condition, for any interior $p^\rho \in (0,1)$ is

$$B = F(c_1^\rho)/(1 - F(c_3^\rho) + F(c_1^\rho)). \tag{22}$$

An increase in B increases the LHS of Eq.(22). Holding ρ constant, c_3^ρ moves in the same direction as c_1^ρ in p^ρ. An increase in B must then increase c_3^ρ. Since $c_3^\rho \leq c_{\max}$, and $F(c_{\max}) < 1$, if B is sufficiently high, the equality can no longer hold; instead, $c_2^\rho = c_3^\rho$ binds and $p^\rho = 1$. This establishes $\overline{B}(\rho)$. Given Lemma 1, $\overline{B}(\rho)$ must decrease in ρ, because c_3^ρ increases in ρ.

To show the existence of a $\overline{F}(\rho)$, rewrite the sender's interior incentive condition as

$$B = F(c_3^\rho)B + F(c_1^\rho)(1 - B). \tag{23}$$

Consider an increase in F in the sense of fosd. Holding c_1^ρ and c_3^ρ constant, the RHS of Eq.(23) decreases. Hence, c_3^ρ must increase in F, but again $F(c_{\max}) < 1$. The second part of the result again follows from Lemma 1 .

Proof of Proposition 8. The $\rho = 0$ case is immediate. Suppose that $\rho > 0$. As long as uniform credulity holds, c_1^ρ and c_3^ρ do not depend on B or F. Consider now an increase in $B > \overline{B}(\rho, F)$. For each $c \in [c_1^\rho, c_3^\rho]$, the investment, conditional on a positive recommendation and not checking, is $y^{\rho,+}(c) = \frac{1}{1+\overline{p}^\rho(c)}$, where $\overline{p}^\rho(c)$ is given by the solution to $c = \frac{\overline{p}^\rho(c)}{(1+\overline{p}^\rho(c))^2}$, or, equivalently, by

$$\overline{p}^\rho(c) = -\frac{1}{2c}\left(4c - 2c + \sqrt{1-4c} - 1\right) \tag{24}$$

Hence, for any $c \in [c_1^\rho, c_3^\rho]$, the ex ante expected payoff, since, here, $p^\rho = 1$, is given by

$$\begin{aligned} E[u^\rho \mid c] &= B(-c) + (1-B)(-\frac{1}{2}(1 - y^{\rho,+}(c))^2 - \frac{1}{2}y^{\rho,+}(c)^2) \\ &= (2B-1)(c_{\max} - c) - c_{\max}, \end{aligned}$$

where the second equality follows when expressing $y^{\rho,+}(c)$ as function of c substituting in Eq. (24). It follows that the expected utility of a type $c \in [c_1^\rho, c_3^\rho]$ is decreasing in B. Finally, the payoff of a type $c < c_1^\rho$ is not changing in B and the same is true for $c > c_3^\rho$. Hence, receiver welfare is increasing in B.

Consider now an increase in F as long as $F > \overline{F}(\rho, B)$. Again the expected utility of types $c \notin [c_1^\rho, c_3^\rho]$ is unchanging. Consider now $c \in [c_1^\rho, c_3^\rho]$. If $B < 0.5$, $E[u^\rho \mid c]$ is strictly decreasing in c on $[c_1^\rho, c_3^\rho]$. Thus, an increase in F which leaves $F(c_1^\rho) = F((1-\rho)/(2-\rho)^2)$ unaffected, decreases receiver welfare .

Proof of Corollary 4. If $B \geq \overline{B}(\rho, F)$, then $R(\rho, \overline{B}(\rho, F)) > \overline{y}$. Hence, there exists $\overline{\gamma}(\rho, F)$ such that $\overline{\gamma}(\rho, F)[R(\rho, \overline{B}(\rho)) - \overline{y}] > \overline{B}(\rho, F)$.

Proof of Corollary 5. Consider uniform credulity. Here, for any F and B, $c_1^\rho = (1-\rho)/(2-\rho)^2$ must hold. Since $p^+(c)$ is increasing in c, $y^{\rho,+}(c)$ is maximal for this c_1^ρ. Furthermore, holding F fixed, $E[y_c^{*,\rho}]$ is maximal for $B = \overline{B}(\rho, F)$. Consider a sequence of F_n, such that each element satisfies the full support assumption, and for each element uniform credulity holds. Let F^* be the Dirac delta on c_1^ρ. It follows that there exists a sequence with $F_n \to_d F^*$ where $(1 - F_n(c_{\max}))\overline{B}(\rho, F_n) - F_n(c_1^\rho)(1 - \overline{B}(\rho, F_n)) > 0$ for all n and $\lim_{n\to\infty} \overline{B}(\rho, F_n) \to 0$ since $F_n(c_1^\rho)$ can be made arbitrarily small and $(1 - F_n(c_{\max}) > 0$ by assumption .

Proof of Proposition 9. Both the real and the projected buyer versions accept any price on the equilibrium path, they both reject any price greater than $b + x$. The projected seller (who is uninformed, but correctly believes that the buyer is uninformed) names a price of $\overline{q} + x$. Since the real buyer beliefs that the projected seller does not know q, no buyer version has an incentive to deviate.

Consider the real seller. If $q < \overline{p} - x$, deviating to any price below $\overline{p}$, leads to a perceived loss. This is true because $q + x \geq (1-\rho)\overline{p} + \rho q$ as long as $\overline{p} - q \leq x/(1-\rho)$. Deviating to some $p > \overline{p}$ generates an expected payoff of

$$(1-\rho)(pe^{-(p-\overline{p})/x} + q(1 - e^{-(p-\overline{p})/x})) + \rho q,$$

which is lower than $q + x$ because $pe^{-(p-\overline{p})/x} + qe^{-(p-\overline{p})/x} \leq q + x < q + x/(1-\rho)$. If $q > \overline{p} - x$, then naming a price of $p = q + x$ maximizes $pe^{-(p-\overline{p})/x} + q(1 - e^{-(p-\overline{p})/x})$.

Finally, the projected seller would never want to deviate to a price above $\overline{q} + x$ since $\overline{p} \leq \overline{q} + x$, or to a price below $\overline{p}$ since either $\overline{p}$ is accepted for sure or $x/(1-\rho) < \overline{q} + x$, but then, again, a price of $p = \overline{q} + x$ is optimal .

References

[1] Allport, Floyd (1924). *Social Psychology.* Boston: Houghton Mifflin.

[2] Allen, Jon, Peter Fonagy, and Anthony Baterman. (2008). *Mentalizing in Clinical Practice.* American Psychiatric Publishing, Washington DC.

[3] Akerlof, George. (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics,* 84: 488–500.

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

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

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

[11] Crawford, Vincent and Nagore Irriberi (2008). "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?" *Econometrica,* 75(6): 1721–1770.

[12] 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.

[13] 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.

[14] DellaVigna, Stefano and Ethan Kaplan. (2007) "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics*, 122: 1187-1234.

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

[16] Epley, Nicolas, Keysar Boaz, Leaf Van Boven, and Thomas Gilovich. (2004). "Perspective Taking as Egocentric Anchoring and Adjustment." *Journal of Personality and Social Psychology*, 87(3): 327-339

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

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

[19] 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.

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

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

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

[23] 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.

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

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

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

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

[28] Jehiel, Philippe and Frederick Koessler. (2008). "Revisiting Games of Incomplete Information with Analogy-Based Expectations." *Games and Economic Behavior*, 62: 533–557.

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

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

[31] Madarász, Kristóf. (2014a). "Projection Equilibrium: Definition and Applications to Social Investment and Persuasion." *Working Paper, LSE.*

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

[33] Malmendier, Ulrike, and Devin Shanthikumar. (2007). "Are Small Investors Naive about Incentives?" *Journal of Financial Economics*, 85(2): 457–89.

[34] Malmendier, Ulrike, and Devin Shanthikumar. (2014). "Do Security Analysts Speak in Two Tongues?" *Review of Financial Economics,* forthcoming.

[35] Mullainathan, Sendhil, Noeth Markus, and Antoinette Schoar. (2012). "The Market for Financial Advice: An Audit Study." *NBER Working Paper*

[36] Miller, Dale, and Cathy McFarland. (1987). "Pluralistic Ignorance: When Similarity is Interpreted as Dissimilarity." *Journal of Personality and Social Psychology*, 53(2): 298–305.

[37] Newton, Elizabeth. (1990). "Overconfidence in the communication of intent: Heard and unheard melodies." Unpublished doctoral dissertation, Stanford University.

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

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

[40] 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.

[41] Prentice, Deborah. (2007). "Pluralistic Ignorance." *In Encyclopedia of Social Psychology,* eds. Roy Baumeister and Kathleen Vohs, pp. 674-675, Sage Publications, Inc.

[42] Rayo, Luis and Ilya Segal (2010) "Optimal Disclsoure Policy." *Journal of Political Economy*, 118(5): 949-987.

[43] Samuleson, William F. (1984). "Bargaining under Asymmetric Information." *Econometrica*, 995-1006.

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

[45] Shelton, Nicole, and Jennifer Richeson. (2005). "Intergroup Contact and Pluralistic Ignorance." *Journal of Personality and Social Psychology*, 88(1): 91–107.

[46] Williamson, Oliver. (1979). "Transaction-cost Economics: the Governance of Contractual Relations." *Journal of Law and Economics*, 22(2): 233–261.

[47] Wimmer, Heinz and Joseph Perner. (1983). "Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception." *Cognition*, 13(1): 103–128.