Limited Perspective Taking in Strategic Communication[1]

Kristóf Madarász[2]

LSE and CEPR

November 2023

[1] This paper subsumes Section 4 of Madarasz (2015) and the rest of that paper is subsumed by Madarasz, Danz, and Wang (2023). For this application, I would like to thank various seminar audiences and Joel Sobel for comments.

[2] Contact: k.p.madarasz@lse.ac.uk

Abstract

This paper incorporates limited informational perspective taking into a class of sender-receiver games. The sender wants to distort the truth, but the receiver can verify her message at some privately known cost. A projecting receiver exaggerates the chance that the sender knows his type and tailors the truthfulness of her message to his cost. Such biased second-order beliefs introduce a wedge between the actual and perceived informativeness of the sender's message and lead to credulity and disbelief in equilibrium. If the conflict is neither too large nor too small, and verification is neither too easy nor too difficult, persuasion is uniformly deceptive and receivers believe too much what senders want them to believe. Comparative statics are non-monotone and both a decrease in the conflict and an increase in verification costs systematically boost credulity and lower receiver welfare. When endogenizing the conflict by bringing in the seller of the asset, I show that under unbiased beliefs, the seller's profit is maximal with no conflict. Instead, under projection, a commonly-known conflict allows the seller to extract maximal profit with the receiver always being credulous on average, financial literacy backfiring, and advice becoming toxic.

Keywords: Perspective Taking, Communication, Second-order Beliefs, Deception, Financial Literacy, Consumer Protection.

1 Introduction

Strategic advice shapes outcomes in many economic and political domains. From financial investment and real estate purchases, to lobbying and medical choices, lesser-informed decision makers rely on recommendations of better-informed senders. Advice is typically disposable thus, even if there is a conflict between a sender and a receiver, under rational Bayesian inference, it should not systematically harm receivers. Nevertheless, it is often seen as an activity that is too persuasive in that it too often fools receivers into believing claims they should not.

For example, following the 2008 Great Recession and the subprime mortgage crisis, many argued that predatory lending practices and the sale of toxic mortgages were enabled by financial advice inducing overly optimistic consumer beliefs about the benefits of their investments. Such advice is then seen as systematically benefiting sellers of the asset, while hurting those buying them. While the incentive to distort advice in most contexts is well-understood, evidence suggests that receivers may, nevertheless, be persuaded too easily. In the context of financial advice, for example, Bergstresser et al. (2009) show that investors buy broker-recommended funds that deliver lower risk-adjusted returns than directly-sold funds even before subtracting the fee charged for advice. Similar credulity has been evoked in the context of fraudulent retail investment scandals, where those mis-selling funds, or running Ponzi schemes, do so successfully despite the relative low cost of uncovering fraud.[1]

Strategic communication fundamentally rests on informational differences and *perspective taking* about these differences. A receiver's interpretation of a sender's advice depends on what the receiver thinks the sender knows. In other words, it depends on the players' beliefs about the beliefs of others. Such *second-order beliefs*

[1] For example, SEC investigations of the failure to detect the Ponzi scheme operated by Bernie Madoff, concluded both that there was no evidence of corruption by investigators but also that it would have been fairly easy to uncover the fraudulent scheme many years before it collapsed. https://www.sec.gov/files/oig-509-exec-summary.pdf

are particularly important when there is a conflict of interest between the sender and the receiver and their incentives are not perfectly aligned.

Suppose Alice advises Bob. Even if Alice's message to Bob has a direct literal meaning, it asserts or denies the validity of a factual claim, or recommends a particular action, how much Bob should trust Alice's message depends on Bob's belief about what Alice knows. In the presence of misaligned preferences, in equilibrium, it is such second-order beliefs that simultaneously determine the actual informativeness of Alice's message (the extent to which Alice lies to Bob), and Bob's perception of the actual informativeness of Alice's message (the extent to which he thinks she lies to her).

In standard models of strategic communication, people have unbiased beliefs about the beliefs of others. A general feature of such Bayesian communication, given rational beliefs, e.g., Milgrom (1981), Crawford and Sobel (1982), is that in equilibrium the actual informativeness of the sender's message and the receiver's perception thereof are the same, *on average.* As a result, receivers benefit from advice and are never systematically fooled by it. Such optimal communication then requires the players to have unbiased beliefs about the beliefs of others. In the language of psychology, a receiver has to fully utilize his *theory-of-mind* capacity, e.g., Leslie (2001), and correctly simulate how the other party thinks.

Evidence from economics and psychology shows, however, that people form systematically biased beliefs about the beliefs of others, e.g., Fischhoff (1975), Wimmer and Perner (1983), Camerer et al. (1989), Gilovich et al. (1998), Birch and Bloom (2007), Danz et al. (2018). Specifically, they engage in limited informational perspective taking by projecting their private information onto others, Madarasz (2012, 2015), exaggerating the probability that others know what they know. This paper considers the implications of such information projection to strategic communication. Formally, I apply the notion of information projection equilibrium, Madarasz (2015), to a class of simple sender-receiver games where an unbiased and sophisticated advisor interacts with a receiver who engages in such projective thinking.

In an environment with a commonly-known conflict and distribution of the payoff state, information projection implies a systematic violation of the above-mentioned key features of Bayesian communication. A sender (advisor) provides free advice to a receiver (investor) about a statement being true or false. The sender has an incentive to claim that the statement is true, but the investor can verify the sender's recommendation at a cost *privately known* to the investor. If he verifies, he learns the truth, and if the sender lied, she suffers a loss.

An unbiased receiver fully adjusts his perspective. Each receiver type correctly understands the extent to which the sender is uncertain about the receiver's type. In effect, also that the sender's incentives are determined by the distribution of how *all* (counterfactual) types behave. Each receiver type correctly understands that the extent to which the sender is able to guess whether the receiver would check a recommendation and thus her incentive to be truthful, depends not on the receiver's privately realized type, but on the sender's uncertainty about the receiver's type. As a consequence, in equilibrium, the true informativeness of the sender's message and the receiver's perception thereof are the same. There is no deception and the receiver is never fooled on average.

A projecting receiver does not fully adjust his perspective. Instead, he exaggerates the chance that the sender knows his type thus the extent to which she can tailor the truthfulness of her message to it. The receiver then always interprets the sender's message in a manner that is too egocentric. If his realized cost of checking is relatively high, he is too incredulous following a positive recommendation because he overestimates the chance that the sender lies to him. If his realized cost is relatively low, he is persuaded too easily because he underestimates the chance that the sender lies to him. Nevertheless, I show that if the conflict between the parties is neither too low nor too high, and checking, on average, is neither too easy nor too difficult, such information projection endogenously predicts *uniform credulity.* Following the sender's advice, all receiver types are successfully deceived (some weakly some strictly) in that each type's ex ante expected posterior confidence in the statement being true is inflated on average. Persuasion is no longer neutral, but leads receivers to believe what the sender wants them to believe too much.

An exogenous assumption of receiver naiveté, in the presence of sophisticated senders, has been considered in the literature on signalling, e.g., Kartik et al. (2007) or Ottaviani and Indherst (2012) who consider naive clients who falsely believe that an advisor suffers a great utility loss whenever her recommendation hurts the client. This paper instead aims to understands systematically false inference from advice that arises endogenously as a function of the biased beliefs about the beliefs of others that result from *limited perspective taking.* In the presence of a transparent and commonly known conflict of interest between the parties, information projection predicts credulity, as well as its opposite, disbelief, endogenously. For a review of other behavioral factors in communication studying limited strategic sophistication, as opposed to limited informational perspective taking, e.g., Crawford (2003) or Ettinger and Jehiel (2010), see Sobel (2020) who also contains a summary of the current paper.

The analysis, however, not simply links such phenomena to a general and portable model of social cognition, but develops comparative static predictions on how the economic fundamentals — the conflict between the parties and the complexity of the recommendation / the receiver's financial literacy — determine the direction and the extent to which persuasion induces such false beliefs and then how policies that directly target these widely discussed, fundamentals may affect receiver welfare. [2]

Such comparative statics and the mechanisms leading to welfare damaging credulity matter. Given full perspective taking, a lower conflict, a greater punishment for false recommendations, shall improve the informativeness of the sender's message and the receiver's welfare. Similarly, lower verification costs, e.g., reduc-

[2] On such policy discussions see, e.g. the UK's Financial Conduct Authority's Consultation Paper 22/24 (2022).

ing the complexity of the statement to be evaluated, or increasing the receiver's financial literacy, shall also lead to the same. One might then suggests that when receivers appear credulous the same policy tools shall lead to the same effects a fortiori.

For example, in the wake of the Great Recession many have argued that financial education was a key policy tool in improving clients' financial choices, e.g., Lusardi (2013), Bernanke (2011). The 2010 Dodd-Frank Wall Street Reform and Consumer Protection Act, also mandated the establishment of the Office of Financial Education, which shall develop a strategy to improve the financial literacy of consumers.[3] Yet the overall evidence on such a channel being effective appears to be mixed, e.g., Hastings et al. (2013), Willis (2011). While many US households buy into fraudulent investment schemes each year, Willis (2011), as Hastings et al. (2013) point out in the review of the evidence "of the few studies that exploit randomization or natural experiments, there is at best mixed evidence that financial education improves financial outcomes." See also, e.g., Kaiser et al. (2021) for a survey of the empirical evidence on positive effects of financial education. Evidence further suggests the reverse relationship in some settings. The National Association of Securities Dealers (Investor Fraud Study Final Report, 2006) directly compared the characteristics of those who were successfully persuaded to buy into fraudulent investment schemes to a randomly selected pool of non-victims. They concluded that "a major hypothesis going into the survey was that investment fraud victims do not know as much about investing concepts as non-victims and would therefore score lower on financial literacy questions. In fact, the study found the exact opposite: investment fraud victims scored higher than non-victims on [all] eight financial literacy questions."

In the model valuable financial literacy and a positive conflict between the parties is necessary for advice to be deceptive. They allow for there to be a positive

[3]See Dodd-Frank Act, H.R. 4173, Title X Bureau of Consumer Financial Protection 2010, Section 1013.

gap between the receiver's perception and the actual informativeness of the sender's message. Indeed, both when there is no conflict between the parties and when the conflict between the parties is overwhelming, inference from advice is unbiased and the receiver is not deceived, on average, irrespective of projection. Limited perspective taking induces credulity and allows the sender to deceiver the receiver when there is small but moderate conflict and when there is valuable financial literacy. I then specify sufficient conditions under which decreasing the conflict or decreasing the distribution of checking costs (increasing financial literacy) will always *increase* credulity and *lower* receiver welfare.

Finally, I endogenize the conflict between the sender and the receiver, by invoking the producer of the asset who sets the sender's incentives. Here, I further show that if selling is sufficiently profitable, advice is always deceptive and valuable financial literacy always backfires. Specifically, in the unbiased case, the seller never finds it optimal to induce a conflict between the sender and the receiver. Since communication is neutral, a conflict only decreases the informativeness of the sender's advice at a pure cost to the seller. In contrast, whenever the receiver projects and has valuable financial literacy, the seller does introduce a conflict in a way that advice induces receiver credulity on average. Valuable financial literacy backfires. As projection becomes full, advice is always toxic in that the receiver is always better off without advice, and decreasing the cost of verification fuels credulity lowering the receiver's welfare and boosting the producer's profit.

The rest of the paper is organized as follows. Section 2 presents the setup. Section 3 provides the analysis of the model and presents comparative static results. Section 4 endogenizes the conflict of interest and Section 5 concludes.

2 Setup

Timing. Upon meeting the receiver (investor, patient), the sender (financial advisor, doctor) privately learns whether a statement is true, $\{\theta = 1\}$, or false, $\{\theta = 0\}$.

She sends a message about this to the receiver. The receiver then decides whether to verify the message at some cost c. If he verifies, he learns the truth. Finally, the receiver takes an action y. For notational simplicity, I assume the prior on θ to be symmetric. None of the qualitative results depend on this prior being symmetric.

Verification. The receiver's cost of verification is drawn from a commonly known cdf $F(c)$ with strictly positive density $f(c)$ over $[0, \infty)$. Its realization is the receiver's private information. This cost can be interpreted as one that is determined jointly by the complexity of the statement to be evaluated and the receiver's privately-known expertise of the specific problem.

A higher distribution of F, i.e., a first-order stochastic increase in F (an increase in F, henceforth), corresponds to a higher distribution of verification costs. In the above interpretation, it can correspond to a greater average complexity of the problem, or to a lower distribution of the receiver's expertise. In the context of financial advice, one can interpret a lower F as higher financial literacy and changes in F as changes in such financial literacy.

Investment. The receiver takes an action $y \in [0, 1]$ to maximize his expected utility. This may correspond to the fraction of resources allocated into buying or selling an asset, promoting or blocking a policy, or the extent of compliance with a doctor's recommendation. To keep the analysis transparent, I assume that the receiver's interim optimal action equals to the posterior probability that the proposition is true. This is captured by the standard assumption that the receiver's utility from investment, potentially only observed with noise, is: $u_r(y, \theta) = -(y - \theta)^2$.

Conflict of Interest. The sender gets a bonus $B > 0$ whenever she issues a positive recommendation of the statement being true. At the same time, if the receiver checks and finds out that the sender lied, the sender incurs a loss (of business, reputation, or regulatory fine) $S > 0$. I normalize $S = 1$ and interpret B in proportional terms. The same comparative statics hold when holding B constant and changing S^{-1}. I then also assume that $B < 1$, but will return to the relaxation of this assumption.

Welfare. When discussing receiver welfare (welfare, henceforth), I take the standard ex ante expected perspective: it equals the receiver's true expected utility, his expected utility from investment, given the *true* distribution of signals, minus the potential verification cost incurred in equilibrium. I refer to the receiver's expected utility when acting only on his prior as the receiver's welfare without advice.

2.1 Solution

I now incorporate limited perspective taking into this setup where the receiver projects her private information onto the sender. Formally, I adopt information projection equilibrium, Madarasz (2015), given a biased receiver ($\rho_r > 0$) and an unbiased, sophisticated, sender ($\rho_s = 0$) to this sequential move game with observable moves by the players. This adoption mimics the incorporation of Madarasz (2015) into sequential bargaining in Madarasz (2021).

Specifically, the receiver assigns probability $1 - \rho$ to the sender's real version and probability $\rho \in [0, 1)$ to the (fictional) projected version of the sender. The real sender knows θ but does not know c and her beliefs are given by $f(c)$. The projected sender instead knows both θ and c. The sender anticipates the receivers mistaken belief about her beliefs and such projection is then effectively common knowledge.

Formally, given the game described above, suppose that, before the players move, Nature first picks θ and c revealing the former only to the sender and the latter only to the receiver. She then plays a binary leakage lottery $\epsilon_t \in \{1, 0\}$ whose realization is observed only by the sender. This determines whether or not Nature leaks c to the sender. The sender initially believes that $\Pr(\epsilon_t = 1) = 0$, which is the case in reality. The receiver believes that $\Pr(\epsilon_t = 1) = \rho$. In solving the model, I adopt perfect Bayesian equilibrium with the players now having these heterogenous priors about Nature's moves regarding the distribution of information in the game.

Let σ_k be player k's strategy, I_k her collection of information sets, and μ_k her system of beliefs including those about the distribution of Nature's moves. There are information sets that the sender believes are never reached. Nevertheless, I require the sender's strategy to also be defined over such information sets and since these sets are singletons, the sender must have correct degenerate beliefs there. Finally, let $V_k(\sigma, \mu_k \mid \iota_k)$ denote k's expected utility of the lottery induced over terminal nodes at information set ι_k given a strategy profile σ.

Definition 1 *A pair $\mu = \{\mu_R, \mu_S\}$ and $\sigma^* = \{\sigma^*_R, \sigma^*_S\}$ forms a perfect equilibrium with information projection (equilibrium, henceforth) if*

1. $V_R(\sigma^*_R, \sigma^*_S, \mu_R \mid \iota_R) \geq V_R(\sigma_R, \sigma^*_S, \mu_R \mid \iota_R)$ *for any* σ_R *at any* $\iota_R \in I_R$*,*
2. $V_S(\sigma^*_R, \sigma^*_S, \mu_S \mid \iota_S) \geq V_S(\sigma^*_R, \sigma_S, \mu_S \mid \iota_S)$ *for any* σ_S *at any* $\iota_S \in I_S$*,*
3. μ_R *is derived from Bayes' rule (whenever possible) given* σ^* *and* $\Pr(\epsilon_t = 1) = \rho$*,*
4. μ_S *is derived from Bayes' rule (whenever possible) given* σ^* *and* $\Pr(\epsilon_t = 1) = 0$*.*

Remark. It is not important for the analysis that in reality the sender never observes the receiver's type. Instead, one can generalize the above informational setup and assume that the receiver's information may truly leak to the sender before she sends a message. Let α denote the true probability of such a leakage event and suppose that the realization pf this event is the sender's private information.[4] The seller now correctly believes that $\Pr(\epsilon_t = 1) = \alpha$ while a projecting receiver exaggerates this probability and believes that $\Pr(\epsilon_t = 1) = (1-\rho)\alpha + \rho$. The results presented below directly extent to the presence of such a true leakage probability.

[4] If the realization of the leakage event is public, then the problem is perfectly separable across the realizations of the leakage event and the continuation game after no leakage is isomorphic with the baseline analysis.

3 Analysis

I first describe the unbiased case, then turn to the case where information projection distorts second-order beliefs.

3.1 Bayesian Case

Consider the unbiased case ($\rho = 0$). If the statement is true, $\theta = 1$, the sender always reports this (makes a positive recommendation). If the proposition is false, $\theta = 0$, the sender lies about it being true with probability p^0 and otherwise makes a negative recommendation. The receiver checks a positive recommendation if and only if her cost is lower than some c^0 and never checks a negative one. Below, $E_\theta[y_c^{*,0}]$ denotes the true ex ante expected equilibrium investment (posterior) of type c. The receiver's prior confidence is $\overline{y}$.

Proposition 1 *If $\rho = 0$, the receiver checks iff $c \leq c^0(B, F)$. The sender lies with probability $p^0(B, F) > 0$. Both $c^0(B, F)$ and $p^0(B, F)$ are increasing in B and in F. Communication is neutral, $E_\theta[y_c^{*,0}] = \overline{y}$ for any c.*

Neutrality. Both a greater conflict B and greater distribution of verification costs, higher F, lead to more lying by the sender and more checking by the receiver in equilibrium. They both decrease the amount of information transmitted in equilibrium and lower the receiver's welfare. Communication, however, is always neutral. The ex ante expected posterior confidence of each receiver type is the same as the prior. This is a general consequence of the martingale property of Bayesian updating; persuasion is purely informative and never shifts ex ante expected posterior beliefs. In turn, it also follows that the receiver always (at least weakly) benefits from advice.

3.2 Persuasion under Projection

Consider, now, the biased case ($\rho > 0$). Recall that the receiver (investor) exaggerates the probability that the sender (advisor) knows her type, that is, how costly it is for him to verify a recommendation. The advisor is sophisticated and while she does not know the investor's type, she knows that he exaggerates his transparency.

Advice is no longer neutral in equilibrium. Instead, it systematically distorts posteriors. Specifically, it leads to two opposite mistakes: (i) *credulity*, whereby the receiver overinfers from a positive recommendation and becomes too optimistic, on average, and (ii) *disbelief,* or incredulity, whereby he underinfers from a positive recommendation and becomes too pessimistic, on average. The next proposition describes how the receiver's type determines the direction of the systematic bias in what he infers from advice.

Proposition 2 *For any $\rho > 0$, equilibrium is unique. There exist $0 < c_1^\rho < c_2^\rho \leq c_3^\rho$, with c_1^ρ decreasing and c_3^ρ increasing in ρ, such that*

for $c < c_1^\rho$ persuasion is neutral, $E_\theta[y_c^{*,\rho}] = \overline{y}$;

for $c \in [c_1^\rho, c_2^\rho)$ credulity holds, $E_\theta[y_c^{*,\rho}] > \overline{y}$;

for $c \in (c_2^\rho, c_3^\rho)$ disbelief holds, $E_\theta[y_c^{*,\rho}] < \overline{y}$;

for $c \geq c_3^\rho$ (weak) disbelief holds, $E_\theta[y_c^{*,\rho}] \leq \overline{y}$.

By projecting his information, the investor exaggerates the extent to which the sender can tailor the truthfulness of her recommendation to his actual type. The less costly it is for the investor to check the sender's message, the less he believes that the sender has an incentive to lie. In particular, while the investor always has a correct belief about the probability with which the real sender lies to him, he believes that, with probability ρ, the sender is the projected version who conditions her lying behavior on his type. Indeed, in equilibrium this fictional projected version of the sender is perceived to lie according to a monotone function $p^+(c)$ strictly increasing on $[c_1^\rho, c_3^\rho]$.

The investor's equilibrium checking strategy matches these type-dependent perceptions. Types for whom checking is sufficiently cheap always check and learn the truth. At the same time, since they overestimate the informativeness of the sender's message, fewer types check for sure than in the unbiased case, c_1^ρ decreases in ρ. Types in $[c_1^\rho, c_2^\rho)$ check too little relative to their true benefit of checking. They are too optimistic after a positive recommendation since they underestimate the probability with which the sender lies and thus the benefit of checking. Whenever they do not check, they are too optimistic following a positive recommendation, and are credulous and overinvest, on average. Types above c_2^ρ (at least weakly) underestimate the amount of information the sender transmits and may check too much relative to the benefit of checking. However, when they do not check, they are (weakly) too pessimistic. Hence, they are (weakly) incredulous and underinvest on average.[5]

While credulity is always present in equilibrium, disbelief is a limited phenomenon. Below, I refer to the case in which *all* types of the investor are at least weakly credulous, and a strictly positive measure of them are strictly credulous, as *uniform credulity.*

The next proposition shows that such uniform credulity holds: (i) when the conflict is not too small, provided $B < 1$ as assumed, and also when (ii) the recommendation is sufficiently costly to verify, provided F has full support as assumed.

Proposition 3 (Uniform Credulity) *Suppose that* $\rho > 0$*.*

1. *If* $B \geq \overline{B}(\rho, F)$*, uniform credulity holds. Furthermore,* $\overline{B}(\rho, F) < 1$*, is decreasing in* ρ*, and* $\lim_{\rho \to 1} \overline{B}(\rho, F) = 0$*.*
2. *If* $F \geq_{fosd} \overline{F}(\rho, B)$*, uniform credulity holds, and if it holds given* ρ*, it also holds given any* $\rho' > \rho$*.*[6]

[5] Note that the results rely on leakage being exaggerated due to projection. If the receiver's cost did true leak to the sender with some true probability α, but this probability was common knowledge, then communication would still be neutral.

[6] Since fosd is only a partial order, multiple such $\overline{F}(\rho, B)$ exist. Below, $\overline{F}(\rho, B)$ refers to any such distribution.

To provide intuition, note that as the conflict increases, the amount of information transmitted in equilibrium decreases both when the investor is biased and when he is unbiased, thus for any given $\rho \geq 0$. Eventually, the sender decides to always lie when $\theta = 0$, and her advice, in reality, conveys no information. In turn, types for whom checking is too costly, do not check, but since they are maximally skeptical and believe that the sender always lies when the news is bad, they now have correct beliefs.

At the same time, as long as $\rho > 0$, receiver types for whom checking is not a dominated choice, still think that the sender's message contains useful information. This wedge between the true and the perceived informativeness of a positive recommendation then implies both that such types check too little, and that they believe too much what the sender wants them to believe. Uniform credulity follows; advice leads to strict overinvestment from the ex ante expected perspective, $E_{\theta,c}[y_c^{*,\rho}] > \overline{y}$. An increase in the degree of projection increases the set of environments where persuasion is uniformly deceptive in this manner.

The same logic holds with respect to an increase in the distribution of verification costs, e.g., a statement that is sufficiently complex for most types to evaluate will induce uniform credulity since all types for whom it is rationalizable to check, will be credulous.

3.3 Comparative Statics and Welfare

I now turn to the comparative static predictions. In the unbiased case, $\rho = 0$, the ex ante expected impact of advice on the receiver's welfare is always positive. Furthermore, both a lower conflict and a lower distribution of verification costs (greater financial literacy) increases information transmission and how much the receiver benefits from advice on average. The next result describes sufficient conditions for the model to reverse these comparative static predictions given any $\rho > 0$.

Proposition 4 *If $\rho = 0$, welfare is decreasing in B and in F. Given any $\rho > 0$,* *there exists* $\overline{B}(\rho, F) < 1$ *and* $\overline{F}(\rho, B)$ *such that*

1. *if* $B \geq \overline{B}(\rho, F)$ *an increase in B decreases expected investment* $E_{\theta,c}[y_c^{*,\rho}]$ *and strictly increases welfare;*

2. *if* $F \geq_{fosd} \overline{F}(\rho, B)$ *and* $B < \frac{1}{2}$*, an increase in F, which does not change* $F(\frac{1-\rho}{(2-\rho)^2})$*, decreases expected investment* $E_{\theta,c}[y_c^{*,\rho}]$ *and strictly increases welfare.*

Comparative Static with B. An increase in the conflict, decreases information transmission and increases costly checking. In turn, it constitutes a negative welfare force. If $\rho > 0$, there is, however, a third force. A higher conflict cautions the receiver and reduces the overinference of, in equilibrium, credulous types. If uniform credulity holds, then all investor types are weakly credulous. Here an increase in B does not affect real information transmission, but decreases over-investment following an unchecked positive recommendation. In turn, it unambiguously *increases* welfare.

The presence of uniform credulity is sufficient, but not necessary for an increase in the conflict to increase welfare. A higher conflict always increases checking and effectively reduces the mistakes in the investment choices of both credulous types and those in disbelief. It will then often lead to higher welfare even if uniform credulity does not hold.[7]

Comparative static with F. Higher verification costs (higher complexity or lower literacy) also decreases information transmission and increases the cost of checking. If $\rho > 0$, there is, again, a countervailing force. Such a change decreases the scope for credulity because, all else equal, the easier it is for a receiver type to check, the more he believes the sender prima facie. In addition, if the conflict is not too great, $B < \frac{1}{2}$, there is always sufficiently little checking such that credulous

[7] To illustrate, suppose that F is such that $f(c) = 2$ if $c \leq \frac{1}{4}$ and f can be anything otherwise. Let $B = 0.24$. If $\rho = 0$, then $p^0 = 0.16$ and receiver welfare is -0.07. If $\rho = 0.9$, then $p^\rho = 0.94$, with $c_1^\rho = 0.07$ and $c_3^\rho = 0.24$, and receiver welfare is -0.23. In the unbiased case, as B increases, receiver welfare decreases. This is not so in the biased case. When $B' \geq 0.36$, welfare in the biased case is now always higher than -0.23; for example if $B \geq 1$, it is -0.18.

types lose less from a higher cost of checking than how much they gain from making less biased investments. Holding the measure of types who may always check constant, such types are bounded by $(1-\rho)/(2-\rho)^2$, under uniform credulity higher verification costs now only decrease the positive gap between the perceived and the real amount of information transmitted.

Again, uniform credulity is only sufficient but not necessary for higher verification costs to *increase* welfare. Since lower financial literacy reduces the scope for credulity, its positive impact can outweigh its negative impact of cheaper checking even if uniform credulity does not hold.[8]

Note finally, that these comparative statics are not global. When there is no conflict $B = 0$, or checking is for free for sure, $F(0) \to 1$, a projecting receiver always learns the truth and his welfare is maximal. While comparative statics are monotone when $\rho = 0$, the above results then imply non-monotonic comparative static predictions whenever $\rho > 0$.

4 Endogenous Conflict

So far, the conflict was a fixed parameter. I specified conditions where both higher financial literacy and a greater punishment for detected lying induced more credulity and lower receiver welfare. To refine these results, I now endogenize the conflict by bringing in the seller of the asset. One can think of a setting where the seller contracts with a financial advisor, or medical doctor, to provide recommendations to investors. Whether or not the seller prefers to employ an advisor whose preferences are aligned or are in conflict with that of the investor is now an endogenous choice of the seller and depends on which one allows the seller to achieve greater profit.

Suppose that, before the resolution of any uncertainty, the seller of the asset pledges to pay the advisor a bonus $B \geq 0$, where I now allow for any such B, when-

[8] To illustrate, suppose again that $B = 0.24$ and $\rho = 0.9$. Assume that $\widetilde{f} = 2$ if $c \leq \frac{10}{121}$, but $\widetilde{f} = 1$ if $c \in [\frac{10}{121}, \frac{1}{4}]$ and $\widetilde{F}$ can be anything otherwise. The receiver's welfare under any such $\widetilde{F}$ is -0.22.

ever the advisor issues a positive recommendation.[9] The seller wants to maximize her expected profit which is given by the receiver's expected investment times some markup $\gamma > 0$ minus the expected transfer to the sender:

$$R(\rho, B, F) - T(\rho, B, F) = \gamma E_{c,\theta}[y_c^{*,\rho}] - t^*(B)B,$$

where $t^*(B)$ is the probability with which the sender makes a positive recommendation in equilibrium. The chosen value of B is again public information.

What is the optimal bonus that a seller, who correctly understands the sender's and the receiver's true behavior in equilibrium, would want to set? Let $B^*(\rho, F, \gamma)$ denote the seller-optimal choice of B.

Proposition 5 *Consider any F.*

1. *If $\rho = 0$, then $B^*(0, F, \gamma) = 0$.*

2. *If $\rho > 0$, then $0 < B^*(\rho, F, \gamma) < 1$ and $E_{c,\theta}[y_c^{*,\rho}] > \overline{y}$ if and only if $\gamma \geq \overline{\gamma}(\rho, F)$, where $\lim_{\rho \to 1} \overline{\gamma}(\rho, F) = 0$.*

In the unbiased case, the seller-optimal conflict is always zero. Here persuasion is always neutral and a conflict only decreases equilibrium information transmission and does so at a pure cost both to the receiver and to the seller. Under unbiased beliefs, perfectly aligned preferences between the sender and the receiver are always optimal for the seller.

In the biased case, $\rho > 0$, the seller can instead always inflate investments by inducing a conflict. If the markup is not too low, the seller will always finds it profitable to do so. The extent of the seller-optimal conflict depends on the receiver's distribution of verification costs. Advice may contain limited but valuable information, and the overall value of advice can be both negative and positive for

[9] Allowing for any fixed unconditional transfer from the seller to the advisor does not change the analysis and the seller never wants to pay for negative recommendations.

the receiver's welfare. Nevertheless, advice always leads to overinvestment in the asset and it is always deceptive inducing credulity on average.[10]

Discussion. The analysis in Section 3 above emphasized the reversal of the unbiased comparative static results and the presence of non-monotone comparative statics given an exogenous conflict. A higher conflict and greater complexity of the problem, through leading to greater skepticism for the receiver, helped close the gap between the perceived and actual informativeness of advice. Indeed, while there was no credulity and no deception in the absence of conflict $B = 0$, or the absence of a verification cost, $F(0) = 1$, limited conflict and valuable expertise (literacy) of the receiver were necessary for credulity and deception to arise in equilibrium.

To provide more detail, consider again any exogenous B, but as above, let's relax the assumption that $B < 1$. If $B \geq 1$, the receiver correctly understands that the seller has a dominant strategy to lie. Hence, irrespective of projection, advice does not contain information, but the receiver understands this, and is never fooled on average. Since B is defined in relative terms, this entails all the cases where lying is 'free', i.e., $S = 0$.

Similarly, if the receiver lacks any valuable financial literacy, advice is again uninformative, but is also not deceptive; there is no credulity on average. In particular, recall that in our setup f had full support. If, instead, checking was always too costly for the receiver, mistaken second-order beliefs would not matter because it would again be common knowledge that the sender has a dominant strategy to lie. Here, however, considering an endogenous conflict qualifies the above point.

Formally, let $c_{\max}$ be the highest type for whom checking is rationalizable. Consider any F that is concentrated on values higher than $c_{\max}$, i.e., $F(c_{\max}) = 0$. It follows that for such an F, the seller-optimal conflict is zero: $B^*(\rho, F, \gamma) = 0$ given any γ and ρ. In turn, under endogenous conflict, advice is always fully revealing and the receiver's welfare is maximal. The fact that when the statement is

[10] Note also that the producer strictly prefers that the sender does not to commit to any disclosure policy about θ ex ante because such a commitment would no longer allow persuasion to be deceptive.

always too complex for the receiver to evaluate, the seller-optimal conflict is null, implies that under endogenous conflict valuable financial literacy can only hurt the receiver's welfare given any $\rho > 0$. If the receiver projects and the conflict is endogenously set, a positive chance that the receiver could gain from checking the sender's recommendation will always be exploited by the seller to the receiver's detriment.

Let me then conclude with a final observation. Returning to the assumption that f has full support, as the bias becomes full, the unbiased Bayesian result is completely reversed. The value of advice is always strictly negative and is decreasing in the receiver's financial literacy.

Corollary 1 *Consider any F and $\gamma > 0$ with $B = B^*(\rho, F, \gamma)$.*

1. *If $\rho = 0$, the receiver's welfare is strictly higher with advice than without advice.*
2. *If $\rho \to 1$, the receiver's welfare is strictly lower with advice than without advice. Furthermore, a decrease in F decreases the receiver's welfare and increases the seller's profit.*

In the unbiased case the value of advice is always positive. In the fully biased case it is always negative. Furthermore, here, the seller-optimal conflict vanishes and investment is decreasing in c globally. Hence, any decrease in F, increase in financial literacy, lowers the receiver's welfare and increases the seller's profit.

5 Conclusion

This paper adopts the model of limited informational perspective taking, Madarasz (2015), to a class of simple sender-receiver games. I show that biased beliefs about the beliefs of others resulting from information projection lead to both credulity and disbelief in strategic communication. In the presence of a commonly known conflict between the players, the model provides non-monotone comparative static predictions on how key policy instruments, e.g., limiting the conflict in financial advice by punishing lying more severely, or increasing financial literacy, impact such credulity and receiver welfare. The predictions may help rationalize puzzling empirical findings and future research can extend these results or test the mechanism directly. The results also point to novel implications for organizational design, e.g., the trade-off between communication versus delegation, e.g., Dessein (2006).

More generally, the paper highlights the role of limited perspective taking and miscalibrated second-order beliefs for strategic communication. Since communication rests on the presence of informational differences, such second-order beliefs are potentially key in this domain. Limited perspective taking leads to an egocentric distortion in such beliefs which may then be important for understanding communication beyond the setup considered in this paper. For example in contexts where meaning emerges endogenously in equilibrium, such beliefs about the beliefs of others could impact the possibility of meaningful communication and provide a structured approach towards an empirically-motivated joint theory of apparent lying aversion, deception, and miscommunication with implications to a variety of applications.

6 Appendix

Proof of Proposition 1. Let $\rho > 0$. First, $p^+(c)$ must be weakly increasing. Suppose that in contrast that for some $c'' > c'$, $p^+(c'') < p^+(c')$. Then type c'' has a strictly lower incentive to check than type c', hence, $p^+(c'') > p^+(c')$ must hold; a contradiction. Second, $p^+(c)$ must also be continuous. A jump in $p^+(c)$ at some $\widehat{c}$ implies the existence of a type above $\widehat{c}$ who checks strictly more than $\widehat{c}$, but checking cannot strictly increase in c if $p^+(c)$ is weakly increasing. Hence, the receiver always checks iff $c < c_1^\rho$, never checks iff $c > c_3^\rho$, and $p^+(c_1^\rho) = 0$ and $p^+(c_3^\rho) = 1$. Furthermore, $p^+(c)$ is strictly increasing on $[c_1^\rho, c_3^\rho]$. Hence, there exists c_2^ρ such that $p^+(c_2^\rho) = p^\rho$. If $c \in (c_1^\rho, c_2^\rho)$, then $E_\theta[y_c^{*,\rho}] > \overline{y}$; if $c > c_2^\rho$, then $E_\theta[y_c^{*,\rho}] \leq \overline{y}$. Finally,

c_3^ρ is increasing and c_1^ρ is decreasing in ρ. Suppose instead that $\rho' > \rho$ and $c_3^{\rho'} < c_3^\rho$. Then $p^{\rho'} < p^\rho$ since c_3^ρ must increase both in p^ρ and in ρ as $p^\rho \leq p^+(c_3^\rho)$. It then also follows that $c_1^{\rho'} < c_1^\rho$ since c_1^ρ is increasing in p^ρ and decreasing in ρ separately as $p^\rho > p^+(c_1^\rho)$. Hence, $p^{\rho'} \geq p^\rho$; a contradiction. The logic for c_1^ρ is analogous. .

Proof of Proposition 2. The sender's incentive condition, for any interior $p^\rho \in (0,1)$, is given by:

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

The LHS of Eq.(1) is increasing in B. Hence p^ρ, thus, also c_3^ρ and c_1^ρ, must increase in B. Since $c_3^\rho \leq c_{\max}$, if B is sufficiently high, Eq.(1) can no longer hold. Instead, $c_2^\rho = c_3^\rho$ binds, and $p^\rho = 1$. It also follows that the threshold above which this is true, $\overline{B}(\rho, F)$, is such that $\overline{B}(\rho, F) < 1$ because $F(c_{\max}) < 1$ by assumption. Finally, by Proposition 7, c_3^ρ increases and c_1^ρ decreases in ρ, hence, $\overline{B}(\rho, F)$ is decreasing in ρ.

To show the existence of $\overline{F}(\rho, B)$, let's rewrite Eq.(1) as:

$$B = F(c_3^\rho)B + F(c_1^\rho)(1 - B). \tag{2}$$

All else constant, an increase in F in the sense of fosd decreases the RHS of Eq.(2). Hence, c_3^ρ and p^ρ must increase in F. Since such an increase is again bounded, there must exist $\overline{F}(\rho, B)$ such that uniform credulity holds if $F >_{fosd} \overline{F}(\rho, B)$, where, naturally, multiple such $\overline{F}(\rho, B)$ exist. Given Proposition 7, if $B > F(c_3^\rho)B + F(c_1^\rho)(1 - B)$, the same holds given any $\rho' > \rho$.

Proof of Proposition 3. The case of $\rho = 0$ is immediate. If uniform credulity holds, $c_1^\rho = (1-\rho)/(2-\rho)^2$ and $c_3^\rho = c_{\max}$. Let $y_c^\rho(+)$ denote type c's investment conditional on receiving a positive recommendation and not checking. If $c \in (c_1^\rho, c_3^\rho)$, then $y_c^\rho(+) = (1 + \overline{p}^\rho(c))^{-1}$ where $\overline{p}^\rho(c)$ is:

$$\overline{p}^\rho(c) = \left(1 - 2c - \sqrt{1-4c}\right)/2c. \tag{3}$$

Let $E[u_r^\rho|c]$ be type c's welfare given ρ. For a $c \in [c_1^\rho, c_3^\rho]$, $E[u_r^\rho|c]$ is:

$$B(-c) - 0.5(1-B)((1-y_c^\rho(+))^2 + y_c^\rho(+)^2) = (2B-1)(c_{\max} - c) - c_{\max} \quad (4)$$

where I used Eq. (3) and some re-arrangements. The above expression is increasing in B. Consider now $c \notin [c_1^\rho, c_3^\rho]$. Here, $E[u_r^\rho|c]$ is independent of B. Hence, an increase in B decreases investment and increases welfare.

Note that if $B < \frac{1}{2}$, then, Eq.(4) implies that $E[u_r^\rho|c]$ is strictly increasing in c on $[c_1^\rho, c_3^\rho]$. Furthermore, for all $c > c_{\max}$, $E[u_r^\rho|c]$ is constant in c since those types never check and $y_c^\rho(+) = \overline{y}$. Hence, an increase in F which leaves $F((1-\rho)/(2-\rho)^2)$ unaffected shifts probability weight to types with lower investment and higher welfare

.

Proof of Proposition 4. Since uniform credulity holds if $B \geq \overline{B}(\rho, F)$, there exists $\widehat{\gamma}(\rho, F)$ such that $\widehat{\gamma}(\rho, F)[R(\rho, \overline{B}(\rho, F), F) - \overline{y}] = \overline{B}(\rho, F)$. Hence, $B^*(\rho, F, \gamma) > 0$ anytime that γ is larger than some $\overline{\gamma}(\rho, F)$. Furthermore, if $B^*(\rho, F, \gamma) > 0$, then $E_{c,\theta}[y_c^{\rho,*}] > \overline{y}$ must hold. Since $\lim_{\rho \to 1} \overline{B}(\rho, F) = 0$, then $\lim_{\rho \to 1} \overline{\gamma}(\rho, F) = 0$.

Proof of Corollary 1. The case of $\rho = 0$ is immediate. As $\rho \to 1$, $c_1^\rho \to 0$ and $\overline{B}(\rho, F) \to 0$, thus, uniform credulity always holds. Hence, $B^*(\rho, F, \gamma) < \frac{1}{2}$ also holds and $E[u_r^\rho|c]$ is now strictly increasing for all $c \leq c_{\max}$. Since $y_c^\rho(+)$ is now globally decreasing in c, a decrease in F decreases receiver welfare and increases the seller's expected profit .

References

[1] Bernanke, Ben. (2011). Statement by Chairman Bernanke on Financial Literacy. http://www.federalreserve.gov/newsevents/testimony/bernanke20110420a.htm.

[2] Bergstresser, Daniel, John Chalmers, and Peter Tufano. (2009). "Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry." *Review of Financial Studies*, 22(10): 4129-4156.

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

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

[5] 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.

[6] Crawford, Vincent, and Joel Sobel. (1982). "Strategic Information Transmission." *Econometrica*, 50: 1431-1451.

[7] Crawford, Vincent (2003). "Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions." *American Economic Review*, 93(1): 133–149.

[8] Danz, David, Kristóf Madarász, and Stephanie Wang. (2018). "The Biases of Others: Projection Equilibrium in an Agency Setting." mimeo LSE and U of Pittsburgh.

[9] Dessein, Wouter. (2002). "Authority and Communication in Organizations." *Review of Economic Studies,* 69(4): 811-838.

[10] Ettinger, David and Philippe Jehiel. (2010). "Theory of deception." *American Economic Journal: Microeconomics*, 2(1): 1-20.

[11] 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.

[12] Financial Conduct Authority CP22/24 (2022). Broadening access to financial advice for mainstream investments. https://www.fca.org.uk/publication/consultation/cp22-24.pdf

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

[14] Kaiser, Tim, Annamaria Lusardi, Lukas Menkhoff, and Carly Urban. (2022). "Financial education affects financial knowledge and downstream behaviors." *Journal of Financial Economics*, 145(2): 255-272.

[15] Gilovich, Thomas, Victoria Medvec, and 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.

[16] Hastings, Justine, Brigitte Madrian, and William Skimmyhorn. (2013). "Financial Literacy, Financial Education, and Economic Outcomes." *Annual Review of Economics*, 5: 347–73.

[17] Inderst, Roman and Marco Ottaviani. 2012. "How (not) to pay for advice: A framework for consumer financial protection." *Journal of Financial Economics,* 105(2): 393-411.

[18] Leslie, Alan. (2001). Theory of Mind. in International Encyclopedia of the Social & Behavioral Sciences, Editor(s): Neil J. Smelser, Paul B. Baltes, Pergamon.

[19] Lusardi, Annamaria. (2013). *Testimony Before the U.S. Senate Committee on Health, Education, Labor and Pension.* April 24, 2013. http://www.help.senate.gov/imo/media/doc/Lusardi.pdf.

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

[21] Madarász, Kristóf. (2015). “Projection Equilibrium: Definition and Applications to Social Investment and Persuasion.” mimeo LSE, CEPR D.P. 10636.

[22] Madarász, Kristóf. (2021). “Bargaining under the Illusion of Transparency.” *American Economic Review*, 111(11): 3500-3539.

[23] Madarász, Kristóf, David Danz, and Stephanie Wang (2023). Projective Thinking: Theory, Evidence, and Applications. SSRN 4480657at http://dx.doi.org/10.2139/ssrn.4480657

[24] Milgrom, Paul. (1981). “Good News and Bad News: Representation Theorems and Applications.” *The Bell Journal of Economics*, 12(2), 380–391.

[25] NASD Investor Education Foundation. (2006). Investor Fraud Study Final Report. https://www.sec.gov/news/press/extra/seniors/nasdfraudstudy051206.pdf

[26] Sobel, Joel (2020). “Lying and Deception in Games.“ *Journal of Political Economy*, 128(3): 907-947.

[27] Willis, Lauren. (2011). “The Financial Education Fallacy.” *American Economic Review P&P,* 101: 429-434.

[28] 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.